Advances in Science, Technology & Engineering Systems Journal

Editorial

Advances in Science, Technology and Engineering Systems Journal (ASTESJ) is an online-only journal dedicated to publishing significant advances covering all aspects of technology relevant to the physical science and engineering communities. The journal regularly publishes articles covering specific topics of interest.

Current Issue features key papers related to multidisciplinary domains involving complex system stemming from numerous disciplines; this is exactly how this journal differs from other interdisciplinary and multidisciplinary engineering journals. This issue contains 57 accepted papers in Computer Science and Electrical domains.

Editor-in-chief
Prof. Passerini Kazmersk

ADVANCES IN SCIENCE, TECHNOLOGY AND ENGINEERING SYSTEMS JOURNAL

CONTENTS

Evaluation of Methods for Sentence Similarity for Use in Intelligent Tutoring System

Emil Brajković[*], Daniel Vasić Tomislav Volarić

Faculty of Science and Education, Department of Informatics, University of Mostar, 88 000, Bosnia and Herzegovina

A R T I C L E I N F O

A B S T R A C T

Finding similarity of textual data is very important task in natural language processing. In this article we present approach to finding similarity of words, paragraphs, sentences and documents. Semantic similarity is one of the central tasks in many applications, including text summarization, Intelligent Tutoring Systems (ITS) etc. In ITS sentence similarity is used to compare the student's response with the correct answer. The result is used to gain information about student's level of knowledge. We propose three different methods that measure text to text semantic relatedness. There are multiple approaches to finding the right measure to determine the similarity of the sentences. Some measure the alignment of characters, and other measure semantic similarity between sentences. In this work we present and evaluate methods for finding not just similarity of sentences but even also similarity of whole paragraphs and documents. We have evaluated these methods using the data from the Yahoo Question and Answer of the Non-Factual Data Set.

1 Introduction

In this paper we extend our research from[1] where we presented and evaluated methods for finding similarity between textual data. These methods can be used inside modules of ITS to evaluate student performance [2]. ITS systems are computer-based instructional systems that are composed of four principal models: the expert model, the learner model, the tutor model and the interface. Expert model otherwise known as domain model is intended as a model that contains all concepts, rules and problem solving strategies. Expert model can be used in variety of contexts such as a source of expert knowledge, students evaluation and student performance error detection. In this model components for sentence similarity are very useful. Student model is an overlay on expert model and is one of the main components of ITS system. In this model special attention is directed to students affective and cognitive state and it's evaluation through the learning process [3]. Tutoring model receives the information from student and domain model. Tutoring model is an ITS component that makes choices about tutoring actions and strategies. One of the most important component of ITS system is communication

module [4]. Main objective of this module is to fill in the gap between the computer who understands machine language instructions and user of an ITS system who understands natural language. Often in communication module question answering systems are implemented. The student receives the question about domain knowledge and answers the question in natural language. It is very important to find robust method of evaluating the students answer. Students answer to given question can be full sentence, phrase or even one word, so this task is not trivial. In general sentence similarity methods can be divided into two categories true understanding and text to text methods [5]. Existing methods for finding sentence similarity work with very high dimensional data and are not adaptable for use in ITS systems. This paper focuses directly on computing the similarity between very short texts. The focus is on finding true understanding text similarity for use in student answer evaluation. The structure of this article includes related work section followed by Proposed algorithm for sentence similarity and Evaluation of proposed methods section. In Related work section we list all similar works in this field. We also show how our proposed method differs from other

---

[*]Emil Brajković, Matice hrvatske b.b., +38763181014 & emil.brajkovic@fpmoz.sum.ba

similar methods. The section Proposed algorithm for sentence similarity includes couple subsections where we discuss the implementation of algorithm evaluated in this article. In the section Evaluation of proposed methods, we expand our previous evaluation with the new method explained in this article. In this section we briefly comment our evaluation methodology and show our results. In the conclusion we list our conclusions and future work related to this article.

2 Related work

ITS systems that are used in many different fields[3] could greatly benefit from robust sentence similarity method. Recent advancements in natural language processing has contributed many valuable solutions in finding sentence similarity. This section reviews some of the solutions addresses the difference between them and methods proposed in this paper. Overall methods for finding sentence similarity could be knowledge based and statistical based. Some of the most popular statistical based sentence similarity methods is Latent Semantic Analysis (LSA)[6]. LSA works by applying Singular Value Decomposition to co-occurrence matrix that is generated from textual data. Resulting dense representations of words on which cosine distance can be used to determine its relatedness. LSA method can distinguish between highly similar and dissimilar words but it is insensitive to minor similarities or dissimilarities between word pairs[7]. More recent knowledge based methods for sentence similarity measurement rely on semantic databases to find distances between concepts from two compared sentences[8]. These methods relying on the existence of concepts from sentence inside the semantic database, word sense disambiguation tool and the algorithm for finding similarity of two concepts. Another approach finding sentence similarity are the methods that are based on word embeddings[9]. These methods are based on finding dense word representations that are extracted from hidden layers of neural network. The cosine distance can be used to calculate word similarity (word that appear in similar contexts are closer). This distance is very good approximation of similarity because Word2Vec might find different sorts of most-similar words. This makes Word2Vec model very scalable and usable in real world applications. Word embedding methods are more preferred over approaches using lexical databases and show better results on word similarity evaluation tasks. The state-of-the-art systems implement hybrid approach using word embeddings and lexical databases[10]. In this work we use hybrid approach for finding sentence similarity using word embeddings and non-projective tree matching algorithm.

3 Proposed algorithm for sentence similarity

In this section we provide overview of our algorithm for sentence similarity that is used for evaluation of students answers in ITS system. There are several natural language processing steps that include syntax and semantic text preprocessing. Such preprocessed text is used as a structure for our algorithm to find the list of phrases that are semantically similar. We preprocess the sentence based on its semantic constituents. Then we build semantic tree of phrases for the two sentences that are being compared. Global steps for this approach in sentence similarity finding are:

- Text preprocessing for two sentences for which we want to find the similarity

- Finding word similarity based on word embeddings

- Finding maximum common subgraph based on association graph

The details of the implementation will be explained in subsections.

3.1 Text preprocessing

Nodes are lists of words that can be connected to other words. Globally we can divide nodes into noun nodes, verb nodes, proposition nodes and adverbial node. Nodes with multiple words contain main word which is called head word. Visual representations of the nodes in sentences are shown in fig. 01.
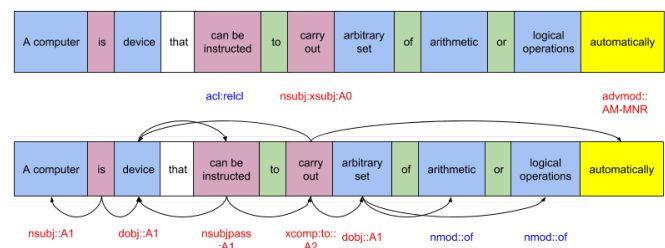
Figure 1: Parsed sentence using POS tagging algorithm for English language, node dependency relations and semantic roles

The words in sentence are segmented into nodes. These nodes are connected by syntactic and semantic relations that are parsed using CoreNLP[11] and SENNA[12] tools.

3.2 Finding similarity of words in sentences using word embeddings

To solve the word order problem the idea is to develop algorithms that will find similarity of two graphs. As shown in figures above the sentences are actually non projective trees, there are many algorithms that deal with tree similarity[13][14]. To find sentence similarity we need to compare the similarity of nodes in dependency tree of the two sentences. To find the word similarity, GloVe[15] word embeddings are used. The GloVe is actually a dictionary with word vector pairing

such as Word2Vec[16]. This dictionary is developed from training a shallow neural network on large text corpora using co occurrence matrix generated from all words in corpora, the network is optimized on custom loss function and 100 dimensional weights that are at start randomly initialized after the optimization with SGD algorithm are paired with corresponding words. Such vector representations are called word embeddings and are very popular because the structure and the way the vectors are obtained give great results for finding semantically similar words. The similarity of words can be calculated with cosine distance of two word vectors. Every word is paired with 100 dimension real valued vector that is extracted from shallow layer of neural network. Nodes that are most similar are paired to so called node pairs. Nodes are paired with nodes with maximal similarity and similarity is calculated with algorithm 01.

> **Data:** corresponding sentences G_1 and G_2
> **Result:** list of edges E and similarities S
> **while** v_1 *in* G_1 **do**
>   **while** v_2 *in* G_2 **do**
>   **if** *similarity(v_1, v_2) > border* **then**
>   put v_1, v_2 into list of edges E;
>   put *sim* into list of similarities S;
>   **else**
>   continue the iteration;
>   **end**
>   **end**
> **end**

Algorithm 1: Algorithm for pairing nodes inside two sentences

Similarity function is something that can be greatly improved, its very dependent on the dictionary and corpora that is trained on, the newer methods for sentence similarity include finding similarities of word phrases using approaches such as skip-thoughts[17]. For every node word sense disambiguation is applied in sentence preprocessing step so the concept based similarities can also be used, but our experiments showed better results for word embeddings approach mainly because not all phrases are contained in WordNet database[18].

Glove embeddings showed very good results for finding phrase similarity. Phrase similarity is similarity of multiple words, or in this case nodes. For calculation of phrase similarity we used Word Movers Distance[19] Finding similar node pairs is process where borders of similarity are set to exclude the nodes that have very small similarity or no similarity at all. Borders are defined based on the word type, our assumption is that if nodes are verbs or nouns the similarity should be larger, but if the node is apposition, proposition, pronoun or something else the border of similarity should be not that high. The borders are defined as follows: Verb nodes - 0.80, Noun nodes and all other node types (pronoun, aposition, proposition) - 0.60. The similar and paired nodepairs are shown in fig. 02.
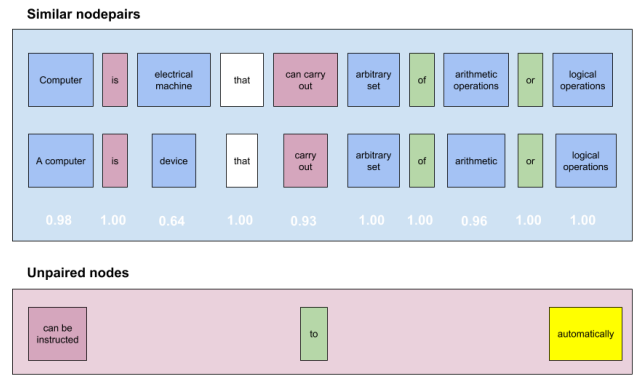
Figure 2: Paired node pairs from two sentences

3.3 Finding maximum common subgraph of whole sentences based on association graph

The next step is to find the maximum complete subgraph of two graphs, to do that we use nodes that are similar. From constructed node pairs we create so called association graph as defined in [20], all nodes that are similar are used to generate graph on which algorithm for finding maximum clique is applied.

Association graph AG The graph product $GP(V;E)$ of two graphs $G1(V1;E1)$ and $G2(V2;E2)$ is a new graph defined on the vertex set $V = V1 \otimes V2$ and the set of edges $E = E_1 \otimes E_2$. The association graph $AG(E,V)$ defined here is one of the graph products with the following adjacency conditions. Any $e = v_i v_s; v_j v_t$ considered to be adjacent:

- if $v_i \in V_{1is}$ adjacent to $v_j \in V_1$ in the original graph G_1 and $v_s \in V_2$ is adjacent to $v_t \in V_2$ in the original graph G_2 , or,

- if v_i is not adjacent to v_j and vs is not adjacent to v_t.

The association graph AG made by the previous definition should have all possibilities of vertex matches between two initial graphs $G1$ and $G2$; namely, a clique in AG corresponds to a common subgraph between $G1$ and $G2$; . Thus, the original problem of obtaining the largest match of sentences can be reduced to the computational task of searching for the largest clique in $AG, MCL(AG)$. For given two sentences the association graph is used for finding maximum clique.

A subgraph $subg_1$ of graph $G1$ is a new graph obtained from $G1$ by deleting some edges and vertices's, the same applies to $subg_2$.The nodes and vertices's that are deleted are those that do not belong to the maximum clique $MCL(AG)$.

If $(node_1, node_2) \notin AG$, where it is $node_1 \in G_1$ and $node_2 \in G_2$, then $node_1$ and $node_2$ are deleted from AG.

Subgraph $subg_1$ and subgraph $subg_2$ are maximal common subgraph MCS. For computing similarity we experimented with use of the Jaccard coefficient[23]

which can be calculated shown in equation (1).

$$JC = (G1, G2) = M\ddot{C}S(G1 + G2) - MCS \qquad (1)$$

This measure is the measure for the similarity of the sentences, so if we want to get the dissimilarity measure DM we use the JC and subtract it as follows $DM = 1 - JC$. We used this measure for evaluation and comparison with other algorithms.

Data: list of all edges E and sentences G_1 and
   G_2
Result: list of connected edges AG
while E_1 *in E* **do**
 while E_2 *in E* **do**
  if v_1 *of* E_1 **and** v_1 *of* E_2 *are adjacent in* G_1
  then
   if v_2 *of* E_1 **and** v_2 *of* E_2 *are adjacent in*
   G_2 **then**
    | put E_1, E_1 into list;
   else
    | continue the iteration;
   end
  else
   if v_2 *of* E_1 **and** v_2 *of* E_2 *are not*
   adjacent in G_2 **then**
    | put E_1, E_1 into list;
   else
    | continue the iteration;
   end
   continue the iteration;
  end
 end
end

Algorithm 2: Algorithm for finding association graph AG

4 Evaluation of proposed methods

Yahoo Non Factoid Question Dataset[24] was used in evaluation of good vs bad answers. The dataset contains 87,361 questions and appropriate answers, from which one of them was marked as the best answer for that question. Each question contains it's best answer along with other answers submitted by users. This dataset contains non-factoid QA, which means it covers topics beyond factoid question answering, such as Who is the father of artificial intelligence?.

These answers are somewhat complex and finding function for similarity is a complex task. The objective was to find the method that maximizes the function of dissimilarity of two sentences. Reason for this is that answers that are labeled as best must be different than the other answers, labeled as bad answers. The evaluation methodology was conducted in three phases using three algorithms. We evaluated parse tree, knowledge tree and word vector representations. In the evaluation phase we define dissimilarity function as the function where we subtract the similarity score from maximum score. The parse tree and knowledge tree was traversed using Eulers algorithm and

then Sorensen-Dice[21][22] algorithm was used to determine the similarity of trees. The results of our evaluation are shown in table I. We included the results of new proposed method described in this article. The resulting number shows the average dissimilarity measure that is generated by sum of all scores and divided by number of sentences.

Table 1: Results of two best algorithm evaluation on whole nf6l dataset

Algorithm	Number of Questions	Number of sentences	Result
Knowledge tree and Dice	5000	29723	74.078
Word2Vec sentence dissimilarity	5000	29723	68.740
New algorithm sentence dissimilarity	5000	29723	66.120
Knowledge tree and Dice	10000	61017	74.847
Word2Vec sentence dissimilarity	10000	61017	68.808
New algorithm sentence dissimilarity	10000	61017	70.234
Knowledge tree and Dice	87361	638847	76.302
Word2Vec sentence dissimilarity	87361	638847	67.778
New algorithm sentence dissimilarity	87361	638847	78.453

In our previous work[1] the method that was giving reasonable results was the knowledge tree approach. In this algorithm we used the constituency tree which was traversed using Euler's path algorithm on and nodes where compared using WordNet similarity measure. The new method in measuring sentence similarity has shown even better results for the whole corpus. Approach given in this work is insensitive to word order and semantically close terms such as synonyms, hypernyms and other. In table I it is shown that new algorithm proposed in this work gave the best results in finding dissimilarity measure between good and bad answers. In future work we plan to evaluate this method by standard means of evaluation using some of the standard tasks for sentence similarity, such as STS Benchmark[27].

References

[1] D. Vasic, E. Brajkovic, "Tree and word embedding based sentence similarity for evaluation of good answers in intelligent tutoring system", In proceedings SoftCom, 2017.

[2] T. Volarić, D. Vasić, E. Brajković Adaptive Tool for Teaching Programming Using Conceptual Maps. In: Hadźikadić M., Avdaković S. (eds) Advanced Technologies, Systems, and Applications. Lecture Notes in Networks and Systems, vol 3. Springer, 2017.

[3] S.R.D. Santos, J.L.M. Amaral, J.F.M. Amaral, ”Adaptive Intelligent Systems applied to two-wheeled robot and the effect of different terrains on performance”, Advances in Science, Technology and Engineering Systems Journal, vol. 2, no. 1, pp. 1-5 (2017)

[4] R. Nkambou, J Bourdeau, R. Mizoguchi, Advances in Intelligent Tutoring Systems, Springer-Link, 2010.

[5] V. Rus, M. Lintean, A. C. Graesser, D. S. McNamara, Text-to-text similarity of sentences, Applied Natural Language Processing: Identification, Investigation and Resolution, 2011.

[6] C. Papadimitriou, H. Tamaki, P. Raghavan, P. Raghavan, S. Vempala, ”Latent Semantic Indexing: A Probabilistic Analysis”, In proceedings of the Seventeenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, 1998, pp. 159–168.

[7] S. Simmons, Z. Estes, ”Using Latent Semantic Analysis to Estimate Similarity”, In proceedings The 28th Annual Conference of the Cognitive Science Society, 2006.

[8] A. Pawar, V. Mago, ”Calculating the similarity between words and sentences using a lexical database and corpus statistics”, IEEE transactions on knowledge and data engineering, 2018.

[9] M. Tomas, I. Sutskever, K. Chen, G. Corrado, J. Dean, ”Distributed Representations of Words and Phrases and their Compositionality”, 2013.

[10] R. Speer, J. Chin, C. Havasi, ”ConceptNet 5.5: An Open Multilingual Graph of General Knowledge”, In proceedings AAAI Conference on Artificial Intelligence, 2017,

[11] C. D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard, and D. McClosky, The Stanford CoreNLP Natural Language Processing Toolkit, in Association for Computational Linguistics (ACL) System Demonstrations, 2014, pp. 5560.

[12] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu and P. Kuksa. ”Natural Language Processing (Almost) from Scratch”, Journal of Machine Learning Research (JMLR), 2011.

[13] P. Bille. Asurvey on tree edit distance and related problems. In Theory of Computer Science, 2004., pp. 217–239.

[14] R. Yang, P. Kalnis, A. K. H. Tung, ”Similarity Evaluation on Tree-structured Data”, In proceedings of 2005 in SIGMOD Conference, 2005.

[15] J. Pennington, R. Socher, C. D. Manning, ”GloVe: Global Vectors for Word Representation”, Empirical Methods in Natural Language Processing (EMNLP), 2014, pp. 1532–1543.

[16] T. Mikolov, K. Chen, G. Corrado, and J. Dean, Efficient Estimation of Word Representations in Vector Space, Jan. 2013.

[17] R. Kiros, Y. Zhu, R. Salakhutdinov, R. S. Zemel, A. Torralba, R. Urtasun S. Fidler, ”Skip-Thought Vectors”, CoRR, 2015.

[18] C. Fellbaum, ”WordNet: An Electronic Lexical Database”, Cambridge, MA: MIT Press, 1998.

[19] M. J. Kusner, Y. Sun, N. I. Kolkin, K. Q. Weinberger, ”From Word Embeddings To Document Distances”,In Proceedings of the 32nd International Conference on International Conference on Machine Learning, 2015.

[20] M. Hattori,Y. Okuno ,S. Goto, M. Kanehisa, ”Genome informatics. International”, In proceedings of Conference on Genome Informatics, 2003.

[21] L. R. Dice, Measures of the Amount of Ecologic Association Between Species, Ecology, vol. 26, no. 3, 1945.

[22] T. Sorensen, A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons, Biol. Skr., vol. 5, 1948.

[23] P. Jaccard, ”Jaccard coefficient similarity”, New Phytologist vol. 11, 1912 pp. 37–50

[24] S. Harding, nfL6: Yahoo Non-Factoid Question Dataset. [Online]. Available: https://ciir.cs.umass.edu/downloads/nfL6/. [Accessed: 19-May-2017].

[25] D. Y. Chang, Applications of the extent analysis method on fuzzy AHP, Eur. J. Oper. Res., vol. 95, no. 3, pp. 649655, 1996.

[26] C.-L. Hwang and K. Yoon, The Technique for Order of Preference by Similarity to Ideal Solution. Berlin, Germany: Springer-Verlag Berlin Heidelberg, 1981.

[27] D. Cer, M. Diab, E. Agirre, I. Lopez-Gazpio, and L. Specia (2017) ”SemEval-2017 Task 1: Semantic Textual Similarity Multilingual and Cross-lingual Focused Evaluation Proceedings of the 10th International Workshop on Semantic Evaluation”, 2017.

[28] T. Landauer, P. Foltz, and D. Laham, An Introduction to Latent Semantic Analysis, Discourse Process., no. 25, 1998.

Harmonic Sensitivity Analysis of Power System with Wind Power

Jignesh Patel[*], Satish Joshi

Department of Electrical Engineering, The Maharaja Sayajirao University of Baroda, Vadodara - 390002, India

A R T I C L E I N F O	A B S T R A C T
	As the contribution from renewable energy sources increase, the power quality degradation has become a challenge for the power system operator. There are major two factors causing power quality deterioration. First is the current harmonic generation by the power electronic converters used in renewable energy sources, and second is the magnification of harmonic voltage due to harmonic resonance condition educed by power factor correction capacitors. The motivation behind this work is power quality related problems, observed recently in the field. To analyze the problem a comprehensive approach is adopted. A practical wind farm configuration is considered for study. Using an eigenvalue analysis method, critical resonance modes are determined for different value of capacitor connected to different buses. To check the effect of change in network parameter on eigenvalue, the sensitivity analysis is carried out and how resonant point is varied with the capacitor value and connection point is discussed by simulating different cases. The results are useful in determining the filter value and location for maximizing its effectiveness in curbing the harmonic problems.

1. Introduction

In the recent years, several problems of Wind Generator tripping due to over-voltage is observed. When the problems were investigated, it is found that the harmonic current, generated by the wind generator converter, produces high voltage due to harmonic resonance. Also, several incidences have been reported of converter tripping due to over-voltage caused by the harmonics. This is the prime reason of carrying out this work to know the exact problem and finally find the applicable solution. In the event of harmonic resonance, the harmonic impedance becomes very high four harmonics of frequencies near to resonance point and theoretically become infinite at the resonance frequency. Under the resonance condition, even a small amount of harmonic current is enough to produce high harmonic voltage drop, which is ultimately resulting in overall voltage distortion.

There are various methods to analyze the harmonic problem. Generally, the frequency scan method used to locate the harmonic resonance point. Though, it is simple and easy to use, but it does not give information about the participation of different network components and also how the resonant point is affected by changes in the value of the components.

To analyze the problem, several techniques were investigated. Finally, based on the ease of use and potential of in-depth analysis, the modal analysis is adopted for analyzing the problem in this work.

In conventional power generation, harmonics is virtually not generated by the generators. Also, the resonance condition does not prevail on the generator bus because of absence of shunt capacitors. But, both of these are present in the Wind Farm. It may impact significantly, if the resonance point coincides with the one of the frequencies of generating harmonic currents.

Harmonic Resonance can become complicated, because of complex network configuration and presence of multiple harmonic sources in the network. A power system network contains many inductive, capacitive and resistive elements. An interaction among inductive and capacitive elements leads to oscillations. The period of oscillation depends on the resistive component. Phenomena. Though, the resistive elements cause the loss of power, but they also act as a damping element, which reduces the harmonic magnitude exponentially and supports the harmonic stability. The exchange of energy between capacitive and inductive elements appears as an oscillations, the frequency of oscillation depends on the value of participating capacitor and inductor value. Such The

[*]Corresponding Author: Jignesh Patel, Email: jignesh.pravinbhai@gmail.com

The network is characterized by the critical frequency known as resonance frequency. If such a network is excited by the source having resonance frequency, will lead to a sustained oscillation.

Traditionally, the frequency scan analysis is a primary tool to find out the resonance in the system, but this technique lacks of follow information.

1. Which combination of L & C are involved in the resonance?
2. How many modes are participating in a given resonance?
3. What is the sensitivity of a resonance w.r.t different network element/component?
4. The effect of resonance on different buses of the network?
5. Which is the most effective way to avoid the harmonic resonance?

To answer these questions, a survey has been made and suitable analysis is selected, which is presented in this work.

All wind parks are unique in characteristics. They all have resonance frequencies that are dependent on various factors like grid topology, connected generators, and reactive power equipment used. Furthermore, the impedance and the resonance points of Wind Park changes with the change in number of turbine and capacitor or filter banks in operation.

The connecting cable has also a predominant effect on the resonance frequency. The internal cable connections in a Wind Park connect the turbine of the wind park, whereas the transmission cable connects the wind park to the nearby substation [1]. The total length of the cable varies from a few kilometers to few tens of kilometers. The length of cable, and hence the cable capacitance plays an important role in harmonic resonance.

Capacitor banks and reactors presence also increases the possibility of resonance in the wind power plant. Several capacitors are connected in the form of bank to improve the power factor and reduce the reactive power absorption from grid [2]. The capacitor bank at the PCC level is also used to support the voltage in abnormal conditions. There may be shunted reactors connected to the transmission cable terminations to compensate the high capacitance of a cable. All these components create conducive conditions for harmonic resonance.

In this work, a systematic approach is presented. Starting with the admittance matrix formation, followed by the finding of eigenvalues of discrete frequency points. For resonance or critical frequency, eigenvalue becomes very low or zero. These eigenvalues are then analyzed further to find out the effect of change in a network element on this eigenvalue. Because, the network configuration keeps on changing, depends on the operational requirement and so is the eigenvalue and critical frequency. So, it becomes imperative to find out, how firm or steady is the any resonance frequency.

The structure of this work is explained here. First, the contemporary literature is surveyed in section - 2. The modelling of component is given in section - 3, succeeded by the section - 4, explains the stepwise method for determining resonance frequency.

The sensitivity analysis is briefly presented in section - 5. The problem is defined in section - 6 and different cases are elaborated from section - 7 to 10. The filter design to reduce the effect of harmonic resonance is depicted in Section -11. Finally, the work is concluded in section - 12.

2. Review of Related Literature

Various literatures are investigated to find out the work done in the field wind power effect on power quality. There are quite a few numbers of literatures are available, which has focused primarily on the power quality aspect of the wind power. In reference [3], harmonic emission from four different wind park is measured and compared to the harmonic spectrum. Further, the harmonic emission from the wind power converter is small as compared to any harmonic emission load. And mostly inter-harmonics are generated by wind converters. The most important thing found in this work is the frequency - duration plot, depicts how the spectrum is over a period of time. The significant change in the spectrum is observed during the measurement period, not only in magnitude, but also the ratio between different components. It is interesting to know from this work that, some components of harmonics are co-related with the active power variation, but others are not. This is very confusing. Finally, the work is concluded with the comment that, a measurement of one location during a short period of time cannot be represented as the harmonic emission due to wind power installation.

In reference [4], measurement of voltage and current distortion, at two locations inside the wind farm, is compared. The variation of the spectrum with respect to time is presented to show the emission in a Wind Park.

Harmonic variation with the wind farm operating point and the random characteristics of their magnitude and phase angle is measured extensively and the deterministic and stochastic characterization is analyzed in reference [5]. Similar work is found in reference [6]. In this work, the author has done extensive field measurements on commercial available wind turbines and revealed several important observations. In the opinion of this work, the spectrum of harmonic from a wind turbine is extended to few kHz, but with the increase in size of WT, the spectrum limited to low frequency, around less than 1 kHz. Among the several other observations, the important observation is about the presence of lower frequency harmonics (5th, 7th etc…). The lower order harmonics are generally not expected from PWM converter. The lower order harmonics follows the Normal distribution, whereas Rayleigh distribution is found in higher order harmonics. Further, the low order harmonics synchronized with fundamental frequency, but high order harmonics varies randomly between 0 to 2π.

A structured framework is presented in reference [7] to analyse the harmonics emission and the resonance condition in the wind power interconnection with the grid. According to this work, the wideband harmonics from wind turbines are stochastic in nature and they are associated with the active power production. They may adversely interact with the grid impedance and cause unexpected harmonic resonance. This issue needs to be addressed

comprehensively at the planning stage, should become more critical as wind power contribution increases in the grid. Unlike conventional power generation, wind power is distributed in nature. So, it is very complex when it comes to analysing the harmonic generation from wind converters. To address this issue, an approach of aggregation of harmonics from individual turbine using a method, similar methodology of IEC 61400 - 21, is presented in reference [8]. In this work, the wind farm with different layout is simulated. Then this result is compared with the aggregated harmonics calculated using proposed methodology. The result presented in this work suggests the use of the proposed methodology for summation of harmonics from the wind farm. According to this work, for better result, both, the accurate modelling of wind farm topology and supply system is very much essential. Similar to this work is found in reference [9], where power quality assessment is done as per IEC 61400-21 procedure.

Another approach of aggregation of wind farm for harmonic propagation is reported in [10]. In this work, detailed model of DFIG is used and using this an aggregated DFIG model is developed to study the harmonic propagation. Frequency domain approach is taken to study the problem. The author emphasizes that all the elements of a system are important, because they are frequency dependent and the whole system can affect the resonance frequencies. The analysis part of this work showed that increase in number of aggregated DFIG cause a shift in the harmonic resonance order.

Reference [11] presented the network reduction method for reducing detailed low - voltage wind farm network to assess the effects of its connection onto a main interconnected transmission system. Application of this method reduces the number of network to be modelled and number of injection nodes which reduces the study time. To demonstrate and prove its validity and effectiveness, the results are compared with the IEC current aggregation method.

Harmonic problems become a concern with the growing number of Wind Power Plant [12], because along with a non-linear load, generation side also contributes to the harmonic emission. In Wind Farm, several conditions may give rise to the resonance phenomenon that will amplify the effect of the harmonic frequency. Add on to this, overview of international standards, grid codes, and basic modelling of Wind Power Plant is also given in this literature.

Reference [13] discusses how a wind power installation can affect the harmonic level in a number of ways. According to the author, Wind Turbines are an additional source of harmonics, particularly non-characteristics harmonics. The resonance point is also depends on strength of the grid. Stronger grid will have a higher order of harmonics, whereas weaker grid gives rise to a lower order of harmonics. Further, the author has extensively explained the sources of primary distortion and secondary distortion. Also, how increase in no of wind power installation weakens the grid is elaborated.

3. Modelling of Components

Electrical components are represented by mathematical equations for analysis purpose. The validity of mathematical

equation depends on many factors and it is applicable for certain operating range only. Also, the method of modelling affects the outcome of the analysis. Here, by taking the case of cable modelling, an effort is made to explain the associated difficulties.

HV Cables and Transmission lines are represented either by equivalent PI or T circuit. How the resonant frequency of two different models of the same component is differing, is explained here by example.

A cable with following parameters is modelled.

R_{SYS}=0.04 Ohm
R_{CABLE} = 0.835 Ohm (0.13 Ohm/km)
L_{CABLE} =12.73 mH (0.112 Ohm/km)
C_{CABLE}=4.2 uF (12.73 Ohm/km)
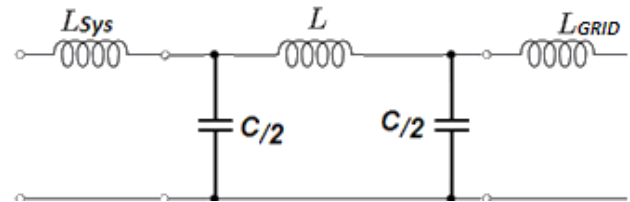
Figure 1: 'PI' model of Cable
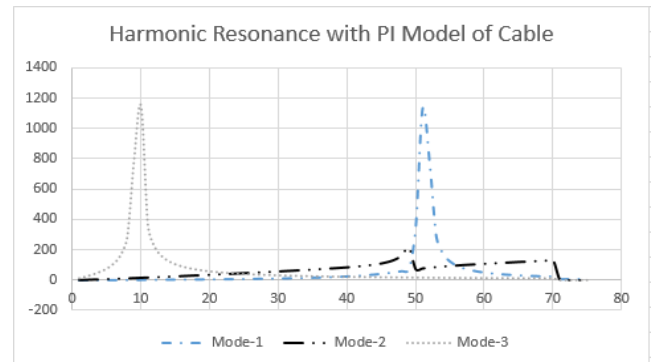
Figure 2: Harmonic Resonance with 'PI' model of Cable
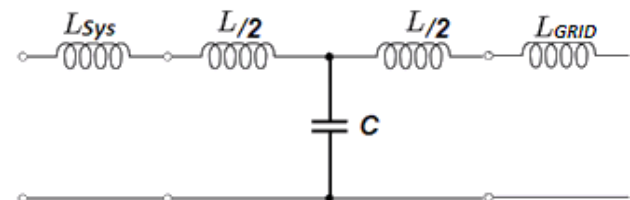
Figure 3: 'T' model of Cable
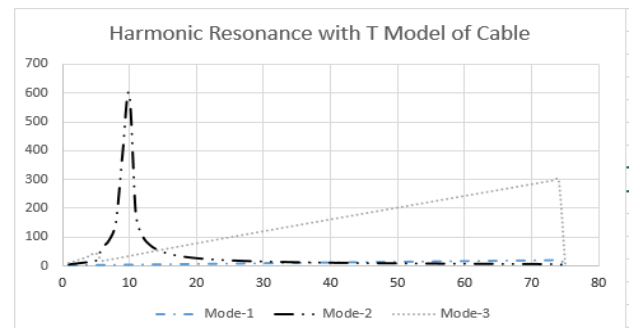
Figure 4: Harmonic Resonance with 'T' model of Cable

Comparison of figure 2 and 4 reveals the effect of modelling on the resonance mode. Mode 1 and Mode 3 are predominant with 'PI' model, whereas Mode 2 is significant and visible in 'T' model of cable. It is important to note that the Mode - 2 with PI model having resonance frequency near to the Mode - 1 of 'T' network. Thus, it is inferred from the result that, the mode number depends on the number of capacitive elements in the system.

In 'PI' model, the cable capacitance is split into two halves and represented in the form π configuration, so there are two modes one at around 10th order and second is around 50th order. Whereas, in the case of 'T' model, only one mode around 10th order is significant, because of lumping of cable capacitance and represented as one element in the form of 'T'.

The nodal impedance or self-impedance of a network is given by

$$Z_{11} = \lambda_1^{-1}PF_{11} + \lambda_2^{-1}PF_{12} + \lambda_3^{-1}PF_{13}$$

$$Z_{22} = \lambda_1^{-1}PF_{21} + \lambda_2^{-1}PF_{22} + \lambda_3^{-1}PF_{23}$$

$$Z_{33} = \lambda_1^{-1}PF_{31} + \lambda_2^{-1}PF_{32} + \lambda_3^{-1}PF_{33}$$

The effective impedance at any bus is depends on participation factor and modal impedance. Mode with smaller values or poor participation factor, may be omitted for calculation of bus impedance. So, when the impedance at any bus is calculated using any method, the result will not differ much. This is because of change of bus participation factor for a given mode. This can be shown here by calculation, but it is avoided to maintain the privity.

4. Determination of Resonant Frequency Based on Critical Mode

Generally, frequency scan is used for finding harmonic resonance. Though, it is an effective method, but lacking in some aspects. For e.g. frequency scan doesn't find the involvement of passive elements for a particular resonance frequency. To overcome, this difficulties, modal analysis is adopted in this work. The modal analysis overcomes the points, in which, frequency scan lacks. Here the modal analysis is explained in detailed.

The first step of analysis is the formation of 'Y' matrix.

$$[Y_h] = \begin{bmatrix} Y_{11}^h & Y_{12}^h & Y_{13}^h \\ Y_{21}^h & Y_{22}^h & Y_{23}^h \\ Y_{31}^h & Y_{32}^h & Y_{33}^h \end{bmatrix} \tag{1}$$

The voltage at different buses is given by

$$[V_h] = [Y_h]^{-1}[I_h] \tag{2}$$

$[Y_h]$ is the network admittance matrix at harmonic order h. $[V_h]$ is the voltage at different nodes, and $[I_h]$ is the harmonic current injected at a different nodes. The admittance matrix is a function of network configuration and the frequency. Harmonic sources having frequencies near to the resonance frequency, the admittance matrix offers very high impedance and at resonance frequency the admittance matrix becomes singular. Some nodal voltages rises sharply, when $[Y_h]$ matric approaches singularity. The transition of admittance matrix to singularity offers an effective way to

analyze the harmonic problem. The eigenvalue analysis has been used for more than a decade to analyze such problems. The eigenvalue analysis technique is applied to system of linear differential equations. In fact they are applicable to any system involving matrices. In power system, it has been widely used to analyze the dynamic stability.

According to [3], the matrices can be decomposed in to following form

$$[A] = [L][\lambda][T] \tag{3}$$

Where, $[\lambda]$ is the diagonal matrix contains eigenvalues, and $[L]$ and $[T]$ are the left and right eigenvectors respectively. The product of left and right eigenvector is identity matrix, applying this to admittance matrix,

So,

$$[L][T] = [I_n]$$

$$[V] = [L][\lambda]^{-1}[T][I] \tag{4}$$

Multiplying both side by $[T]$,

$$[T][V] = [T][L][\lambda]^{-1}[T][I]$$

$$[T][V] = [\lambda]^{-1}[T][I] \tag{5}$$

Defining, $[U_m] = [T][V]$ as the modal voltage vector and $[J_m] = [T][I]$ as the modal current vector respectively, the above equation can be written as

$$[U_m] = [\lambda]^{-1}[J_m] \tag{6}$$

The inverse of the eigenvalue, $[\lambda]^{-1}$ has the unit of impedance and is known as the modal impedance $[Z_m]$. From (4), it can be seen that, if $\lambda_i=0$ or is very small, even a small injection of modal current J_i will lead to a large modal voltage V_i.

$$[T][I] = [J_m] \tag{7}$$

Nodal voltage,

$$[T][V] = [U_m] \tag{8}$$

And nodal current,

$$[T][I] = [J_m] \tag{9}$$

Determination of system resonant frequencies using finding of critical mode is possible by finding out the eigenvalue of 'Y'

matrix at various discrete frequency. System resonant frequencies can be located from Admittance-frequency curves of the modal Admittance, where admittance approaches vary low value, near to zero. The procedure of resonance determination by using modal analysis is illustrated in flow chart given in figure 1. The maximum magnitude of modal impedances at a resonance can be interpreted as the degree of resonance severity.
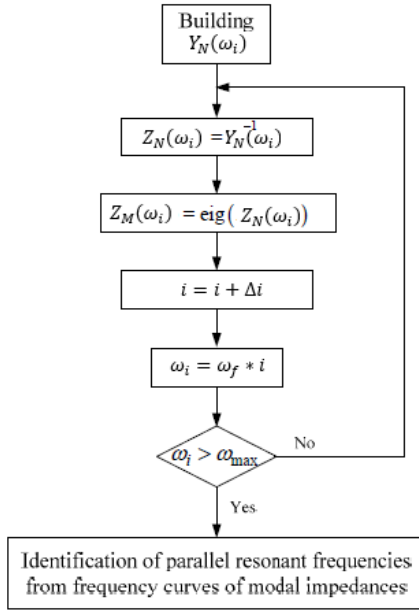
Figure 5: Flow Chart for Critical Eigenvalue Finding

Plot impedance-frequency curve and identify the peaks from curve. Now reduce the Δi, and run the scan again between two frequency value, where the peaks are observed. Stop the scan when $i < i_{min}$.

5. Modal Sensitivity Analysis

The modes or eigenvalues varies with the passive elements of the circuit. So, as the element value changes, mode value also change. The sensitivity of any mode with respect to the any element of matrix is a function of left and right eigenvector [3-5].

According the reference [6], the modal sensitivity is given by equation 10. First order eigenvalue sensitivity of λ_k w.r.t parameter a_{ij} is given by

$$\frac{\partial \lambda_k}{\partial a_{ij}} = l_i^k r_j^k \ (i, j, k = 1,2,3 \dots. n)$$

(10)

Where,
l_i^k = Left eigenvector
r_j^k = right eigenvector

Integration of Wind Farms to the power system has some challenges, one of them is the harmonic resonance, which result in to magnification of harmonics. Several tripping has been observed

in the recent days in Wind Farms. The problem analysis pointed out the harmonic resonance as a root cause. Before inverters are connected to the grid, a detailed assessment is required to find out network impedance and also the behavior of network impedance with the change in frequency and also with the change in passive elements. Direct connection of inverters to the grid may give rise to power quality problems. One such problem is analyzed here.

The harmonic emission of WT converters are normally low, and hence voltage distortion usually remain below standard limits. However, the presence of parallel resonance in the Wind Park grid may increase the voltage distortion above the limits and also affect WP harmonic emission to the main grid. Several work analyze the resonance problem at WT terminal by frequency scan method. But as mentioned in section – 4, it is lacking in some crucial aspects and does not provide information related to origin of harmonic magnification. It is used widely because it requires less computation efforts and give quick overview of harmonic resonance. So, the frequency scan method can be used as a compliment of modal analysis.

The wind farm under study consists of 60 WT of 2 MW each. The aggregate capacity of the wind farm is 120 MW. Each WT is connected to the 0.440 /1.1 KV Transformer. The collector level, which is known as pooling substation, is at 33 KV. At the Pooling substation, the voltage is raised through 1.1/33 KV, 12 MVA transformers. The polling station is connected to the 132 KV transmission substation through 36 KV, 485 Sq.mm, and Armoured XLPE cable of 10 km in length. The power is then evacuated from 132 KV substation through transmission line with Moose Conductor.
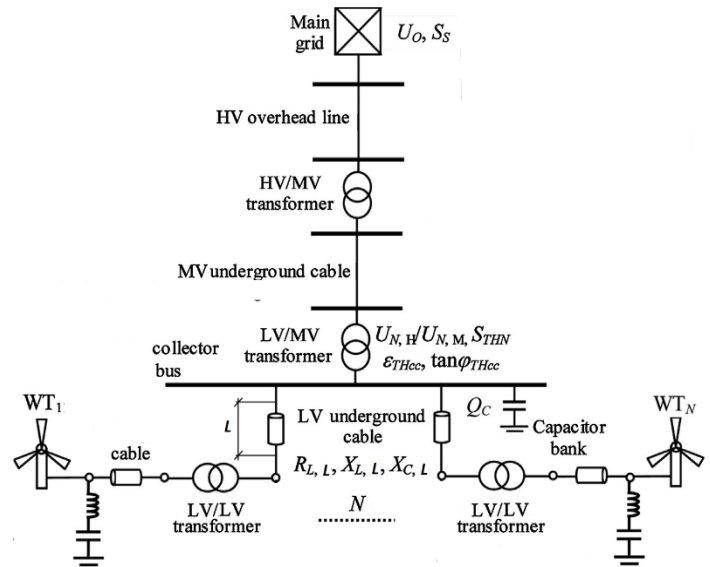
Figure 6: Network Diagram

While running in parallel to the grid, each inverter is able to produce up to 1.95 MW of power. During moderate wind speed condition, each WT generates around 1.2 – 1.3 MW of Power. There was intermittent inverter tripping observed from the site. The inverter tripping history recorded the "Line Overvoltage Fast". This implies that the line voltage has exceeded 120% for more than

8 cycles (i.e. 120 ms). Also, there was an unusual hissing sound before the tripping.

7. Case – I: Single Wind Turbine with 200 kVAr Capacitor at the Terminal of Wind Turbine

In this case, only one turbine is considered in the Wind Park and 200 kVAr capacitor is connected to its terminal for PF Correction.

Figure 7: Harmonic Mode – 3, 9th Order

The network is further analyzed to find out the effect on the concerned bus. Using analytical method, the resonant frequency is found out. The first dominant frequency found is 435 Hz, which is near to 9th order. The PF correction capacitor along with combined reactance of LV transformer (0.440 / 1.1 kV), reactance of LV cable and reactance of LV/MV (1.1/33 kV) transformer, forming a parallel resonance at 11th order. PF Capacitor value is 0.595 uF and transformer reactance is 43.56 Ohm.

$$h_{res} = \frac{1}{2\pi f} \sqrt[2]{\frac{1}{(0.595e^{-6})(0.2236)}} = 8.72$$

(12)
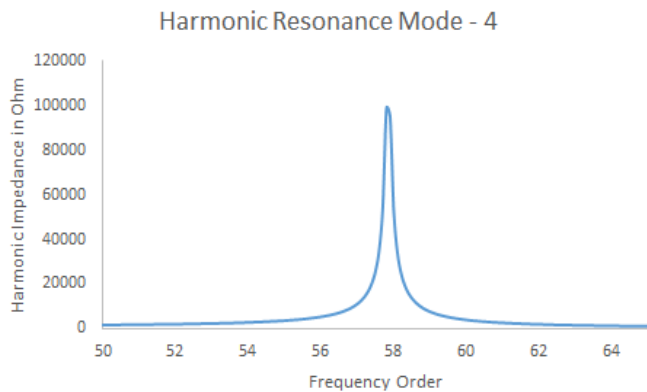
Figure 8: Harmonic Mode – 4, 59th Order

Harmonic Resonance Mode – 4 is contributed by 33 kV Cable capacitance of 0.36 uF. Cable capacitance is acting along with 33 kV Cable reactance, 132 kV transformer reactance, 132 kV line

reactance and Grid reactance, having combined value of 2.40 Ohm. According to analytical expression the harmonic frequency order is 60.7th order as given in equation 13. This is very near to the mode-4 as shown in figure -7.

$$h_{res} = \frac{1}{2\pi f} \sqrt[2]{\frac{1}{(0.36e^{-6})(7.63e^{-3})}} = 60.7$$

(13)
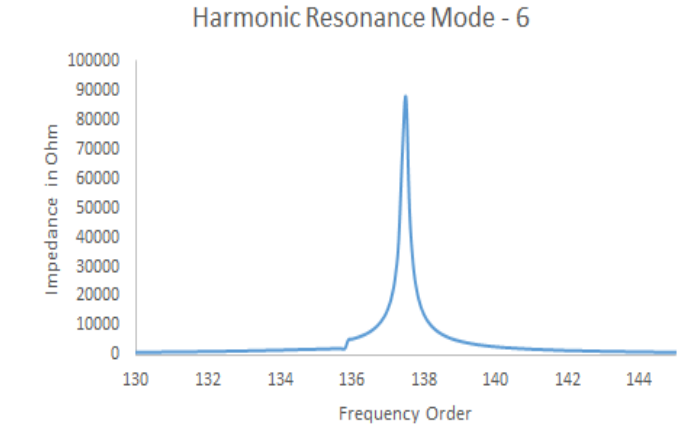
The Mode – 6 is also contributed by 33 kV Cable capacitance. As, the cable is represented by PI equivalent circuit, the total capacitance is divided in to two halves and each capacitor is lumped on two sides of the cable impedance. So, according to analytical expression, the cable capacitance of 0.36 uF is reacting with effective reactance of 33kV Cable, 132 kV transformer, 132 kV line and Grid, having aggregate value of 0.533 Ohm (1.6 ‖ 0.8). Resonance frequency as per the analytical equation (14) is 128.8th order, which is near to the Mode – 6.

$$h_{res} = \frac{1}{2\pi f} \sqrt[2]{\frac{1}{(0.36e^{-6})(1.7e^{-3})}} = 128.8$$

(14)

Table 2: Eigenvalue Sensitivity for Case-I

Modal Sensitivity	Bus-1	Bus-2	Bus-3	Bus-4	Bus-5	Bus-6	Bus-7
Mode-3	0	0	0.0197	0.2411	0.6227	0.1128	0.0039
Mode-4	0	0	0.0113	0.4139	0.1736	0.1556	0.2456
Mode-6	0.5	0	0.0103	0.0045	0.0124	0.1503	0.3224

In this case, number of wind turbine is increased to two. It can be seen from result plot in figure 8. Two resonance points is seen, one is at 8.1th and second is at 9.5th order. The reason for two resonance peak is multiple capacitors, one is due to the PF capacitor and second is due to the LV cable capacitance.
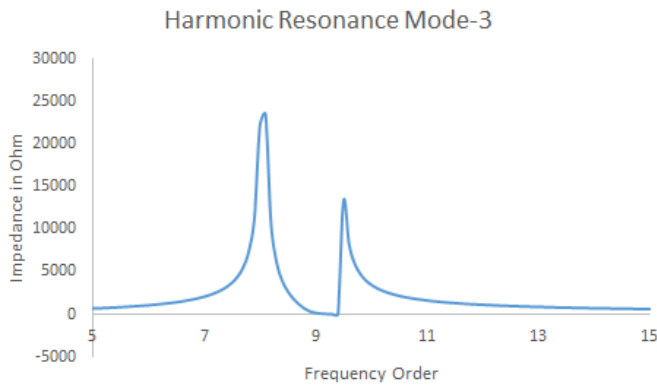
Figure 10: Harmonic Mode – 3, 8.1th and 9.5th Order
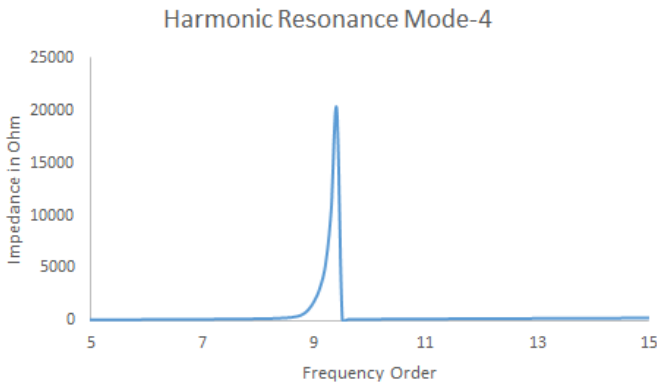
Figure 11: Harmonic Mode – 4, 9.4th Order

Mode – 4 is near to the second peak of Mode – 3. The frequency span between two resonances depends on the impedance between two capacitors. Here, the 1.1kV cable impedance parting cable in to two halves, in PI form. Due to this, two peaks are observed. As the cable impedance is very low, two peaks are observed nearby.
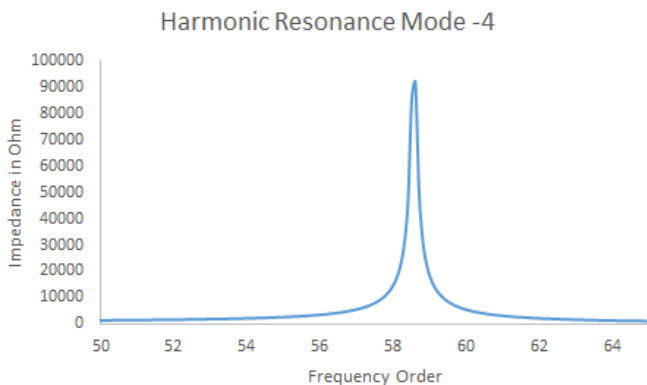
Figure 12: Harmonic Mode – 4, 58.6th Order

Mode – 4 has second peak at 58.6th order, which is the same as Mode – 4 observed in Case – I. This remains unaffected, as the addition of second wind turbine is not affecting the network above 1.1 kV collector bus to Grid. The same is observed in Mode – 6,

as shown in figure – 11. This is also not affected by addition of additional wind turbine.
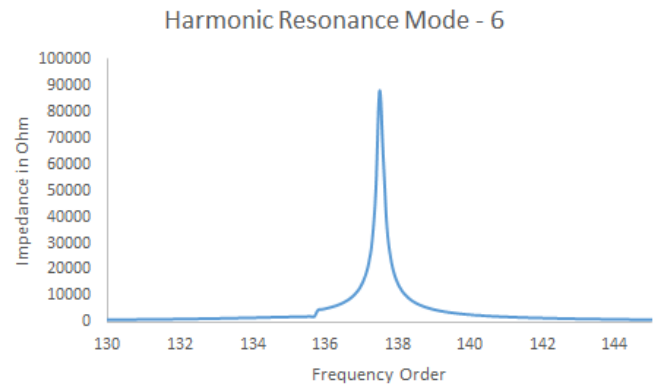
Figure 13: Harmonic Mode – 6, 137.5th Order

Table 3: Eigenvalue Sensitivity for Case-II

Modal Sensitivity	Bus-1	Bus-2	Bus-3	Bus-4	Bus-5	Bus-6	Bus-7
Mode-3 (8.1)	0	0	0.0563	0.3751	0.4794	0.0859	0.0036
Mode-3 (9.5)	0	0	0	0.0059	1	0.0382	0.01
Mode-4	0	0.0131	0	0	0.0284	0.6769	0.333
Mode-6	0.5	0	0.0103	0.0042	0.0105	0.1526	0.3224

9. Case – III: Six Wind Turbine with 200 kVAr Capacitor at the Terminal of Wind Turbine

Harmonic Resonance Mode -3
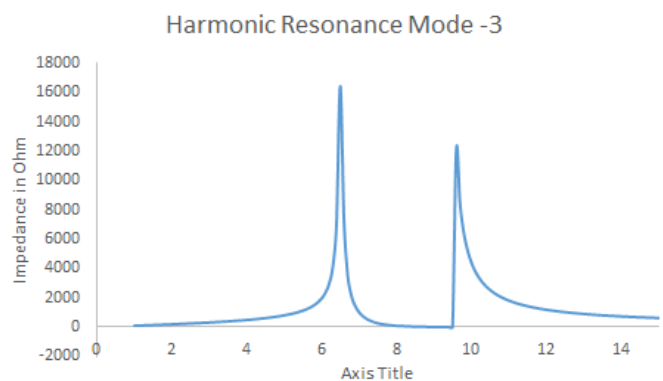
In this case, Mode – 4 and Mode – 6 remains unchanged, as this are network dependent modes and as the network configuration is not changed, the modes is also remaining unaffected. The resonance sensitivity analysis is given in Table – 4.

Table 4: Eigenvalue Sensitivity for Case-III

Modal Sensitivity	Bus-1	Bus-2	Bus-3	Bus-4	Bus-5	Bus-6	Bus-7
Mode-3 (6.5)	0	0	0.1548	0.3946	0.3747	0.0731	0.0034
Mode-3 (9.6)	0	0	0	0	0.9966	0.0031	0.0046

10. Case – IV: Single Wind Turbine with 200 kVAr Capacitor at the 1.1 kV Collector Bus (Bus – 3)

This case is same as Case – I, except one change i.e. PF Capacitor is connected at Bus – 3 instead of Bus – 1. This case is analyzed to study the effect of location of PF Capacitor. Figure – 13. It is observed that, the resonance frequency is relocated to 21.7th order.
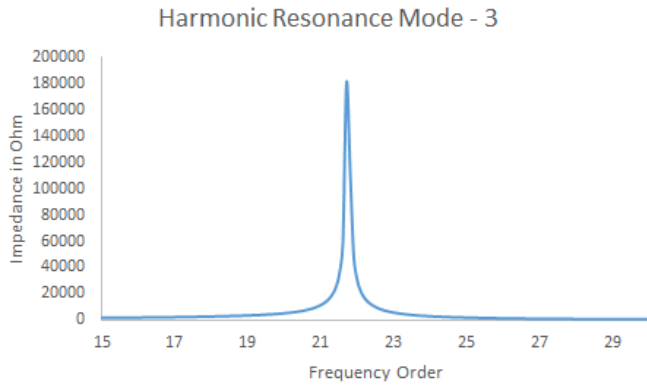
Figure 15: Harmonic Mode – 3, 21.7th Order
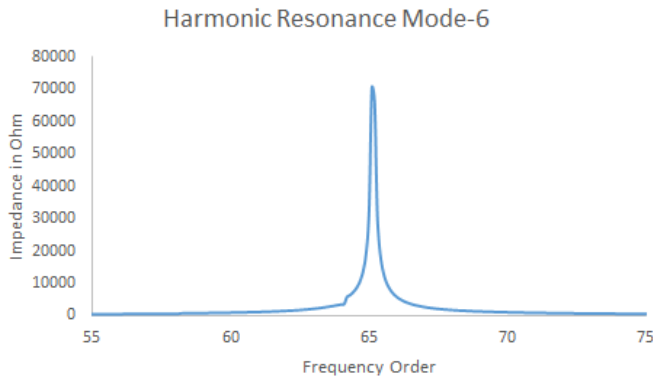
Figure 16: Harmonic Mode – 6, 65.2nd Order
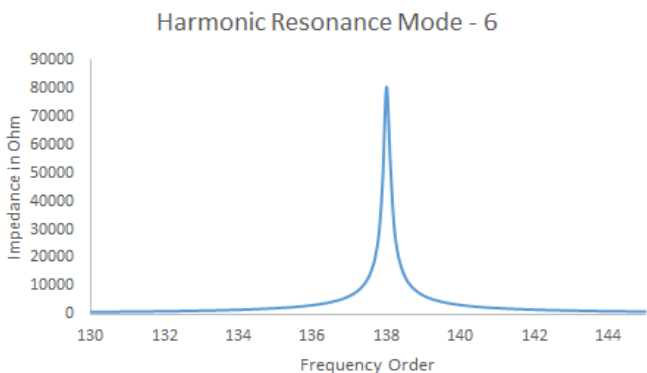
Figure 17: Harmonic Mode – 6, 138th Order

It is observed in figure – 14 and 15, Mode – 6 has now two resonance peak one is at 65.2nd order and second is at 138th order. The second peak is less affected by the relocation of PF capacitor but first resonance is not introduced, which was absent earlier. As the PF capacitor is now connected at 1.1 kV collector bus, it is affecting the network and it is the reason for reshaping of Mode – 6. Sensitivity of Modes is for Case – IV is given in Table – 5.

Table 5: Eigenvalue Sensitivity for Case-IV

Modal Sensitivity	Bus-1	Bus-2	Bus-3	Bus-4	Bus-5	Bus-6	Bus-7
Mode-3 (21.7)	0	0	0.3099	0.2915	0.3077	0.0877	0.0033
Mode-6 (65.2)	0.5	0.218	0	0	0.0014	0.0423	0.2375
Mode-6 (138)	0.5	0	0.0104	0	0	0.1661	0.3233

11. Filter Design
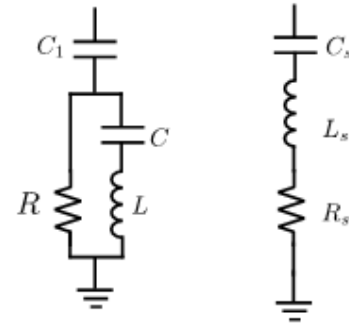
Figure 18: C-Filter and Equivalent Tuned Filter

$$(15)$$

$$L = \frac{m^2 - m\sqrt{m^2 - 1}}{h_0^2 - 1} \frac{2U_1^2}{Q_1 \omega_F}$$

The detail analysis of derivation of equation of 15 and 16 is given in reference [14]. The inclusion of C-Filter reduces the harmonic impedance to low level. For several value of m, the resonance point is checked for the stability and it is observed quite stable.

Table 6: Filter Parameter for Different Value of Cost Optimization Parameter

Sr. No	m	L (mH)	C(uF)
1	2	232.2	43.63
2	3	223	45.4
3	4	220	46
4	5	218.8	46.3
5	10	217.2	46.65
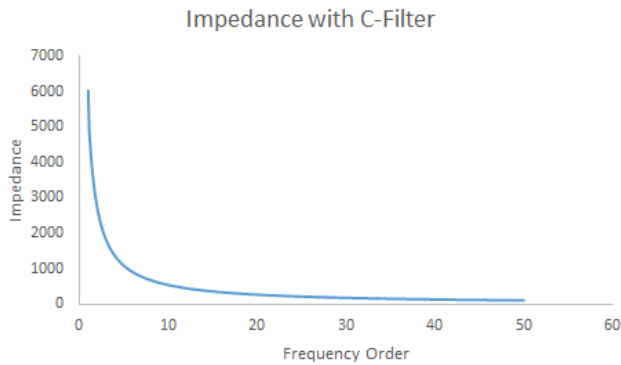
Figure 19: Impedance at 33 KV with R=450 Ohm

The impedance at 33 KV bus after inclusion of C-type filter is given in figure 15 for 450 Ohm of damping resistor. It reveals the important characteristics of C-type topology. The impedance is high around fundamental frequency and is reducing with frequency, so it offers lower impedance path for higher order harmonics and thus reduces the effect of resonance. The measurement data confirms the absence of such a high order harmonic in the inverter output, so the high impedance around 50th order will not affect the wind farm operation. The C-type topology mitigates the harmonic resonance effectively without any resurgence of new resonance point. The application of different topologies and their effectiveness is planned for the future work.

12. Conclusion

To recapitulate the core idea of this work, wind farms cause harmonic distortion in the power system. Also, the resonance point varies depends on the system parameters and system configuration. From the result of different cases, following points are concluded

1. As the wind power varies, the resonance point also varies, due to the change of power factor capacitor switching to maintain the power factor and also to support the system voltage.
2. The location of PF Capacitor will have significant effect on the resonance frequency
3. The resonance frequency reduces with increase in the size of Wind Farm.
4. The sensitivity of resonant point with respect to any component of power system is also demonstrated.
5. Sensitivity analysis helps us to plan and also design the harmonics filter to avoid critical conditions.

The optimized filter design using sensitivity analysis itself is a dedicated topic and future work on it is already planned. It is the reason why, it is not included in this work.

13. System Data

1. 1.1 kV Transformer: X=0.484 Ohm,
2. PF Capacitor: Xc=0.968 Ohm,
3. 1.1 kV Cable: R=0.01 Ohm, X=0.02 Ohm, C=0.36 uF,
4. 33 kV Transformer: X=8.712 Ohm,
5. 33 kV Cable: R=0.47 Ohm, X=1.6 Ohm, C=0.36 uF,
6. 132 kV Transformer: X =5.808 Ohm,
7. 132 kV Line: R=0.00080475 Ohm, X=0.01447 Ohm, R/X=18,
8. Grid: Fault Level=2500 MVA, R=0.3868 Ohm, X=0.9556 Ohm, R/X=18.

Conflict of Interest

We hereby declare that this work is based on the authors own idea and there is no conflict of interest.

References

[1] Sergio L. Varricchlo, Nelson Martins, Leonardo T. G. Ulma, "A Newton-Raphson Method Based On Elgenvalue Sensitivities to Improve Harmonic Voltage Performance", submitted for publication in the IEEE Transactions on Power Delivery.

[2] Sergio L. Varrichlo and Nelson Martins, "Filter Design Using a Newton-Raphson Method Based on Elgenvafue Sensitivity", IEEE Proceedings of the Surer Power Meeting, July 16-20, 2000, Seattle, Washington, USA.

[3] Kai Yang,Math H.J. Bollen, Fellow, IEEE, and Mats Wahlberg, "A comparison study of Harmonic Emission Measurements in Four Windparks", Paper to 2011 IEEE Power & Energy Society General Meeting. 978-1-4577-1002-5/11/$26. 00 ©2011 IEEE

[4] Kai Yang, *Student Member, IEEE,* Math H.J. Bollen, *Fellow, IEEE,* and Mats Wahlberg, "Comparison of Harmonic Emission at Two Nodes in a Wind Park", https://www.researchgate.net/publication/261055566, Conference Paper, DOI: 10.1109/ICHQP.2012.6381163

[5] Luis Sainz, Juan Jose Mesas, Remus Teodorescu, *Senior Member, IEEE*, and Pedro Rodriguez, *Member, IEEE, "*Deterministic and Stochastic Study of Wind Farm Harmonic Currents", IEEE Transactions On Energy Conversion, Vol. 25, No. 4, Pg 1071-1080, December 2010

[6] Paulo F. Ribeiro, *Fellow, IEEE,"*An Overview of Probabilistic Aspects of Harmonics:State of the Art and New Developments", IEEE Transaction on Power System, 2005.

[7] Fabio Medeiros, Dalton C. Brasil, Member, IEEE, Paulo F. Ribeiro, Fellow, IEEE, Cristiano A. G. Marques and Carlos A. Duque, Member, IEEE, "A New Approach for Harmonic Summation Using the Methodology of IEC 61400-21", DOI: 978-1-4244-7245-1/10/$26.00 ©2010 IEEE

[8] Morales, X. Robe and J.C. Maun, "Assessment of Wind Power Quality: Implementation of IEC61400-21 Procedures", RE&PQJ, Vol. 1, No.3, Pg 470 - 476, March 2005.

[9] Silas Y. Liu, Caio M. Pimenta, Heverton A. Pereira, Victor F. Mendes, Gabriel A. Mendonça And Selênio R. Silva, "Aggregated Dfig Wind Farm Harmonic Propagation Analysis", Anais do XIX Congresso Brasileiro de Automática, CBA 2012. ISBN: 978-85-8001-069-5

[10] Foroozan Ghassemi and Kah-Leong Koo, "Equivalent Network for Wind Farm Harmonic Assessments", IEEE Transactions On Power Delivery, Vol. 25, No. 3, Pg 1808-1815, July 2010.

[11] K.N Md Hasan, Kalle Rauma2, P. Rodriguez, J. Ignacio Candela, Raul S. Muñoz-Aguilar, Alvaro Luna, "An Overview of Harmonic Analysis and Resonances of Large Wind Power Plant", Pg 2467-2474,, 978-1-61284-972-0/11/$26.00 ©2011 IEEE

[12] Math H. J. Bollen And Kai Yang, "Harmonic aspects of wind power integration", J. Mod. Power Syst. Clean Energy (2013) 1(1):14–21, DOI 10.1007/s40565-013-0001-7

[13] Jignesh Patel and Satish K. Joshi, "Harmonic Stability of Grid Connected Renewable Energy Source- A Case Study", UPEC-2017, Greece, October 2017.

[14] Yao XIAO, Jie ZHAO and Shijie MAO "Theory of the Design of C-Type Filter", 11th Conference on Harmonic and Quality of Power (2014).

[15] S. L. Varricchio, N. Martins, L. T. G. Lima and S. Cameiro Jr."Studying Harmonic Problems Using a Descriptor System Approach", Proceedings of the IPST'99-International Conference on Power System Transients, Budapest, Hungary, June, 1999.

[16] P. M. Anderson, B. L. Agrawal and J. E. Van Ness, "Subsyncronous Resonance in Power System", IEEE Press, New York, USA, 1990.

[17] IEEE Harmonics Model and Simulation Task Force, "Test systems for harmonics modeling and simulation," *IEEE Trans. Power Del.*, vol. 14, no. 2, pp. 579–587, Apr. 1999.

[18] P. Kundur, Power System Stability and Control. New York: McGraw- Hill, 1994.

[19] G. T. Heydt, *Electric Power Quality*. West Lafayette, IN: Star in a Circle Publications, 1991

[20] R. C. Dugan, M. F. McGranaghan, and H. W. Beaty, *Electrical Power Systems Quality*. New York: McGraw-Hill, 1996.

[21] T. H. Ortmeyer and K. Zehar, "Distribution system harmonic design," *IEEE Trans. Power Delivery*, vol. 6, pp. 289–294, Jan. 1991.

[22] J. Martinon, P. Fauquembergue, and J. Lachaume, "A state variable approach to harmonic disturbances in distribution networks," in 7th Int. Conf. on Harmonics and Quality of Power, Las Vegas, NV, Oct. 16th–18th, 1996, pp. 293–299.

[23] Jignesh Patel, and Satish Joshi, "Harmonic Stability of Grid Connected Renewable Energy Source – A Case Study", UPEC 2017, Egypt.

[24] Verghese, G.C.; Perez-Arriaga, I.J.; Schweppe, F.C. "Selective modal analysis with applications to electric power systems—II. The dynamic stability problem," *IEEE Trans. Power App.Syst.*, vol. PAS-101, no. 9, pp. 3126–3134, Sep. 1982.

Improvement of Transmission Characteristics in Multilayer Dual Band Filter

Ryosuke Nakano*, Yuta Takeuchi, Mikio Tsuji, Hiroyuki Deguchi

Department of Electronics, Doshisha University 1-3, Miyakodani, Tatara, Kyotanabe-shi, Kyoto, 610-0321 Japan

A R T I C L E I N F O	A B S T R A C T
	This paper first simplifies the GA-optimized filter structure developed by us. Then the filter structure is modified by introducing slot and conductor stubs, based on electric fields and magnetic currents at resonant and transmission zero points. As a result, return losses in two passband regions are suppressed more than 28 dB, and also insertion loss in stopband dividing two passbands is suppressed more than 20 dB. Finally, two kinds of the proposed filters are fabricated, and effectiveness of them is confirmed from comparison between the measured and the numerical results.

1. Introduction

This paper is an extension of work originally presented in 2017 Asia Pacific Microwave Conference, (APMC2017) [1]. Based on our work in [2], the goal is to develop high-performance UWB bandpass filter.

In recent years, wireless communication systems such as wireless local area network (LAN) and mobile terminals have spread rapidly. High-frequency filters are used to extract a signal of a desired frequency band for the purpose of prevention of unnecessary radiation and interference. Increasing in complexity and volume of wireless communication, a filter having a large attenuation amount close to pass band and an improved insertion loss is indispensable. Conventional resonant elements that have been developed to achieve high performance such as improved multi-band sharing and attenuation characteristics outside a pass band required for the wireless communication device are not sufficient. Therefore, development of a new resonant element desired. Under such circumstances, we have proposed the ultra-wideband (UWB) filter, of which conductor patches and slot shapes were optimized under the genetic algorithm (GA) technique [2]. As for UWB filters, many filter structures have been developed [3-17]. They are essentially based on the multi-mode resonators (MMR) which were first developed in [3]. Then, various stub-loaded resonators were often used to achieve good skirt selectivity and compact size [4-7]. To perform further good filter performance, the resonator shapes including a stepped-impedance technique were investigated [8-17]. The planar circuit filter designed by GA shows the desired characteristics according to the

evaluation function, but the return loss there was not so high (about 10 dB). In addition to the problem of calculation capacity and computation time, the circuit element shape is complicated and the degree of freedom is small, so that it is difficult to improve the filter characteristics by slightly modifying the GA-optimized filter shape.

In this work, we clarify resonant mechanism and transmission zero mechanism of the UWB filter optimized by the GA. Then, the simplified filter structure based on the their mechanisms is proposed. Furthermore, to improve loss characteristics between two passbands of the dual-band filter, the conductor and slot stubs are newly loaded. Validity of the proposed filter is also verified experimentally.

The rest of this paper is structured/organized as follows: Section 2 discusses resonant mechanism of the UWB filter; Section 3 presents improvement of filter transmission characteristics; Section 4 verifies effectiveness of the proposed filter; and finally, in Section 5, we summarize the conclusions derived from this work and indicate possible future works.

2. Simplified filter generated by GA

New planar circuit filters have been developed by GA evaluated characteristics of both passband and stopband [18-19]. In this section, we briefly describe the structure of the multilayer planar circuit filter and the outline of the optimized design by the GA of the planar circuit filter [18-19]. After the circuit shape and filter characteristics are described, we simplify the filter based on resonant principle.

16

2.1. Structure of the GA-optimized UWB filter

Figure 1 shows the UWB-filter structure developed by us. This filter optimized by GA consists of slot elements on the ground plane and patch elements on the conductor-backed dielectric substrate ($\varepsilon_r = 2.2$, thickness $t_2 = 0.51$ [mm]). Both elements have arbitrary shapes and the filter is covered by the shielding box ($a = b = 15.2$ [mm], $c = 20.0$ [mm]). Input and output of the filter are conected through a microstrip line. A lot of resonances are mainly caused by slot elements.
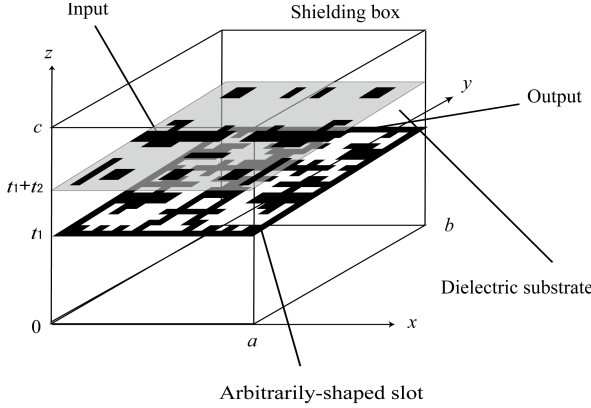
Figure 1. Structure of the GA-optimized UWB filter.
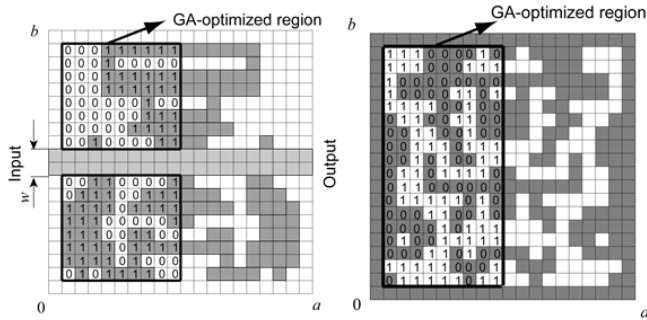
Figure 2. An example of an arbitrarily-shaped element constructed by uniform mesh.

2.2. Shape of the GA-optimized UWB filter

As shown in Fig. 2, the planar circuit shape is represented by a set of conductor patch cells and handled by binary codes of "0" and "1" in the optimized design method by GA [3]. The analysis is performed by the moment method. In the design, the symmetrical structure on the yz plane of $x = a/2$ is assumed and the shape of the input/output line is fixed. The parameters for optimization by GA are set as follows. The xy planes at $z = t_1$ and $t_1 + t_2$ are divided into 20×20 cells, and the size of one cell is 0.76 mm. The evaluation frequency bands are 3.4–4.8 GHz, and 7.25−10.25 GHz in the passband, while 1.0–3.0 GHz, 5.0−7.0 GHz, and 10.35−14.0 GHz in the stopband, and the target of the return loss in the passband and the insertion loss in the stopband are set as 10dB. The mutation rate is 0.02 and the crossover rate is 0.8. The generation number is 400, the population size of one generation is 50, and the elitist strategy is applied.

Figure 3 (a) and (b) show the optimized patch and slot shapes, respectively and Fig. 4 shows its transmission characteristics. It is found that the passbands in the low and high frequency ranges are

constituted by five resonances. Since the circuit shape obtained from GA is very complicated, it is very difficult to modify its shape for improving the filter characteristic. Therefore, the resonant principle is clarified from the electric-field distribution, and the shape is simplified based on this.

(a) Patch shape   (b) Slot shape
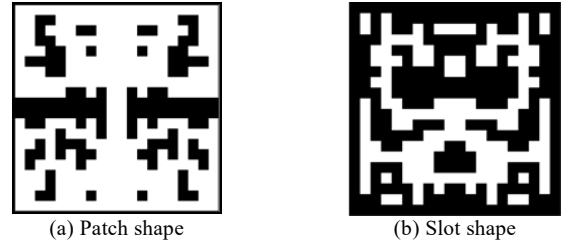
Figure 3. Shape of the GA-optimized UWB filter.
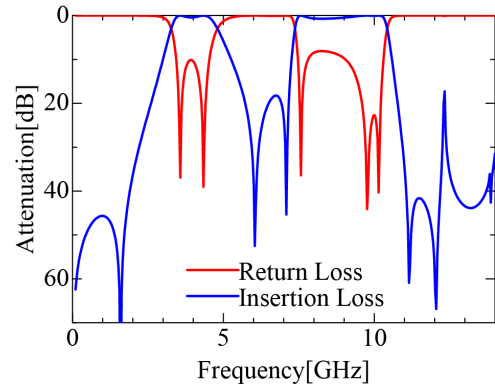
Figure 4. Frequency characteristics of the UWB filter optimized by GA.
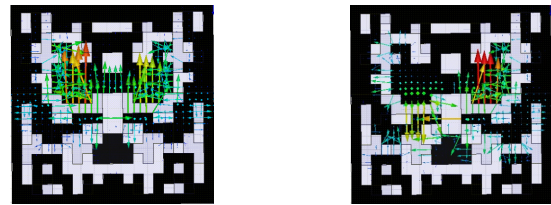
(a) 3.6 GHz   (b) 4.5 GHz
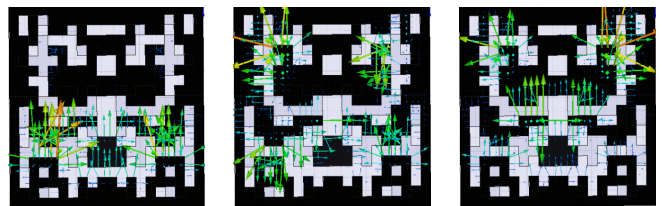
(c) 7.6 GHz   (d) 9.6 GHz   (e) 10.1 GHz

Figure 5. Electric-field distribution at each resonance.

2.3. Resonant mechanism

Resonances are mainly caused on the slot face, so we observe the electric-field distributions on the slots. Figure 5 shows the electric field on the slot at each resonant point. Resonance at 3.6 GHz shown in Fig. 5 (a), is the half-wavelength even resonant mode of the upper-part slot. Resonance at 4.5 GHz in Fig. 5 (b) is the half-wavelength odd resonant mode of the two slots arranged symmetrically for the center line. Resonance at 7.6 GH in Fig. 5 (c) is the half-wavelength even resonant mode of the lower-part slot. Resonance at 9.6 GHz in Fig. 5 (d) is the odd resonant mode that is the next higher-order one of the resonance at 4.5 GHz. Resonance at 10.1 GHz in Fig. 5 (e) is the 3/2-wavelength even

resonant mode that is the higher-order one of the resonance at 3.6 GHz. From the above-mentioned guess of the resonant mechanism, we first calculate the frequency characteristics for the simplified structure shown in Fig. 6 (a) and (b). Its frequency characteristic is shown in Fig. 7, where four

(a) Patch shape   (b) Slot shape
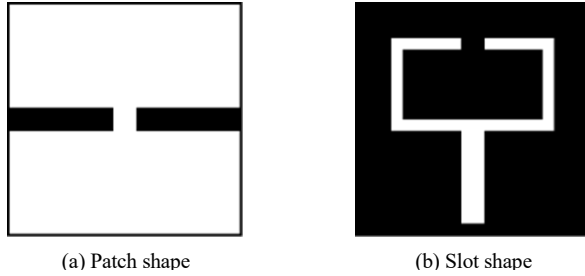
Figure 6. Basic structure of the upper-part slot.
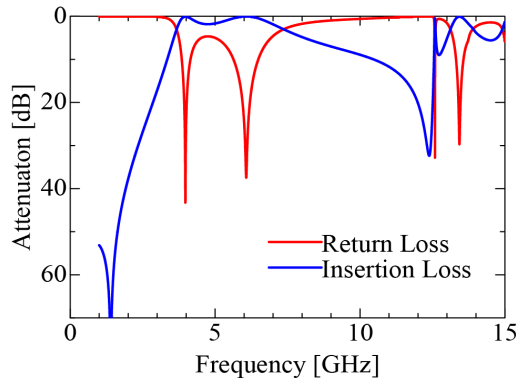
Figure 7. Frequency characteristic of the simplified slot structure shown in Fig.6.

(a) Patch shape   (b) Slot shape
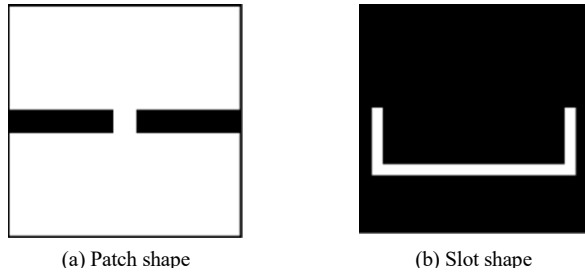
Figure 8. Basic structure of the lower-part slot.
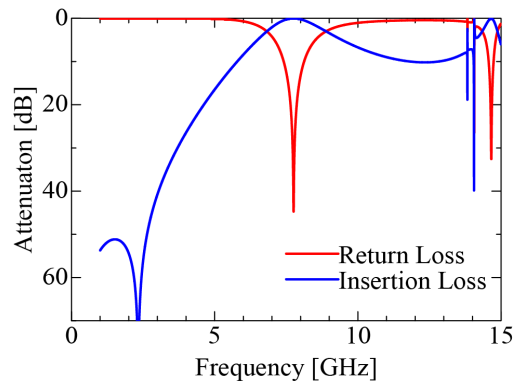
Figure 9. Frequency characteristic of the simplified slot structure shown in Fig.8.

resonant points are produced corresponding to the first, the second , the forth and the fifth resonant points in Fig. 4. Figures 8 (a) and (b) also show another simplified structure and its frequency characteristic is given in Fig. 9, where one resonant point is produced corresponding to the third resonance in Fig. 4. Then we

combine these two basic simplified structures and adjust the filter characteristics to the UWB filter by adding slot stubs and microstrip patches and also by modifying the slot junction part. As a result, Fig. 10 (a) and (b) are obtained as a simplified filter structure and the enlarged part of the slot junction is also shown in Fig.10 (c) The dimensions are listed in Table 1. The slot width is 0.76 mm. Figure 11 shows the frequency characteristics of the simplified UWB filter. Five resonances are reproduced near the original resonant frequencies shown in Fig. 4. Figure 12 shows the magnetic currents at the resonant point to confirm the resonant mechanism. It is clear from this figure that the simplified filter structure works by the same resonant mechanism with that of the GA optimized filter. However, the insertion loss in the stopband is not so high less than 20 dB. So we improve it by considering mechanism of the transmission zeros in the next session.

Table 1. Dimensions of the patch and the slot parameters (unit: mm).

$p_1 = 1.52$	$p_2 = 0.76$	$p_3 = 1.78$	$p_4 = 2.24$	$p_5 = 3.61$
$p_6 = 1.04$	$p_7 = 1.52$	$p_8 = 2.78$	$p_9 = 2.38$	$p_{10} = 4.90$
$p_{11} = 1.42$	$p_{12} = 0.74$	$p_{13} = 2.58$	$g = 0.74$	$s_1 = 2.58$
$s_1 = 3.04$	$s_2 = 6.84$	$s_3 = 10.64$	$s_4 = 3.80$	$s_5 = 4.56$
$s_6 = 13.68$	$s_7 = 5.80$	$s_8 = 1.06$	$s_9 = 1.47$	$s_{10} = 1.4$
$s_{11} = 3.06$	$s_{12} = 0.46$	The slot width is 0.76 mm		

(a) Patch shape   (b) Slot shape

(c) Slot shape of the slot-junction point
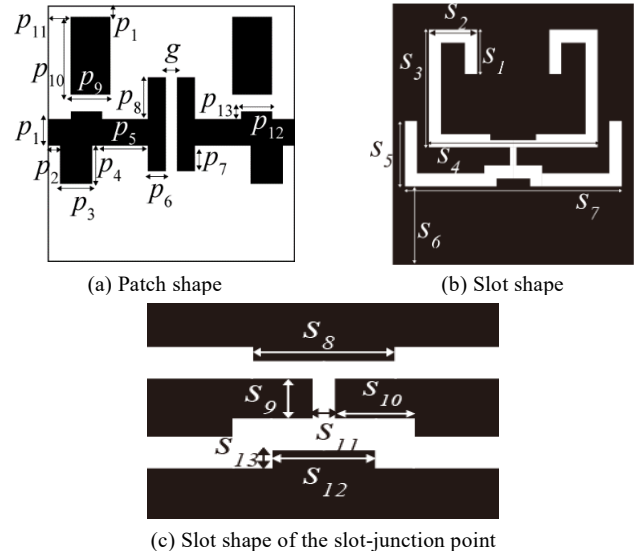
Figure 10. Shape of the simplified UWB filter.
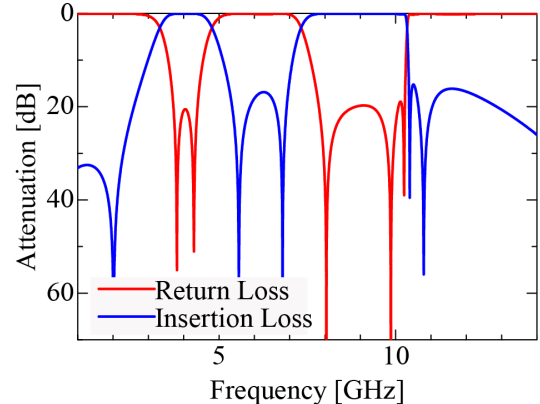
Figure 11. Frequency characteristics of the simplified filter.

(a) 3.6 GHz   (b) 4.5 GHz

(c) 7.6 GHz   (d) 9.6 GHz   (e) 10.1 GHz
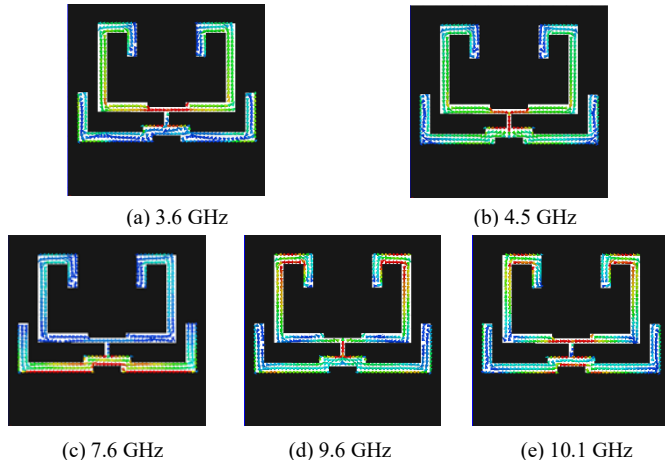
Figure 12. Magnetic-current distribution at each resonance.

(a) patch shape

(b) Insertion Loss
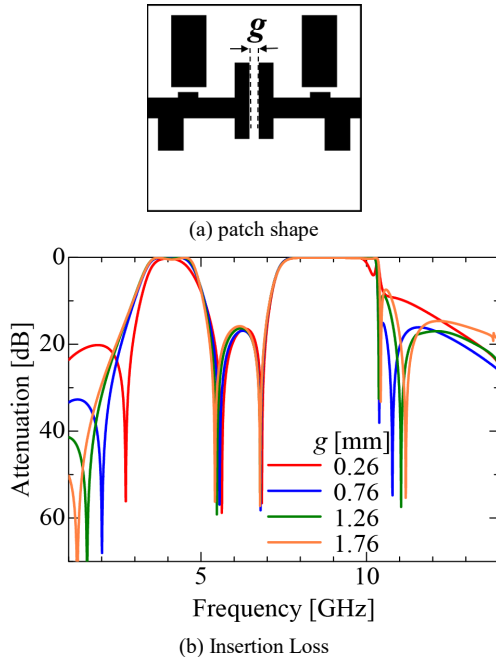
Figure 13. Frequency for various gap *g*.
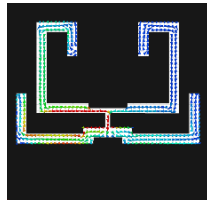
Figure 14. Magnetic-current distribution at 5.5 GHz.

3. Improvement of stopband characteristics

We can see five transmission zeros within 12 GHz in Fig. 4. Among these transmission zeros, the first and second zeros can control by slight modifications of the filter shape, which have little influence for the passband characteristics. We first take notice of the transmission zero at 2.0 GHz. This transmission zero may be caused by the cancellation due to the direct coupling between the input and the output microstrip line through the gap *g* shown in Fig. 13 (a). Figure 13 (b) shows the transmission characteristics for varying the value of the gap *g*. As expected, only the first transmission zero greatly moves, depending on *g*. It is clear that the gap *g* is a useful parameter for controlling the frequency point
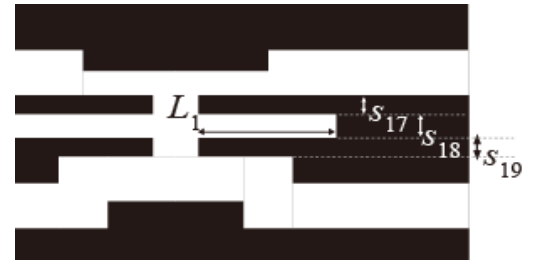
Figure 15. Parameters of the slot stub.
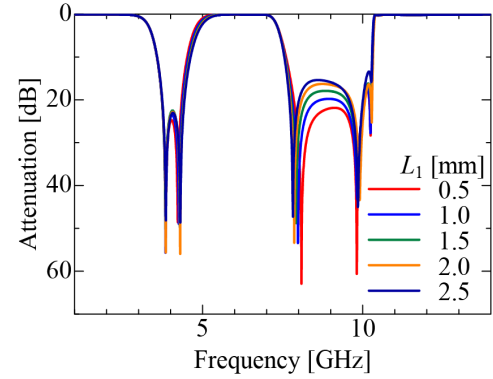
(a) Return loss
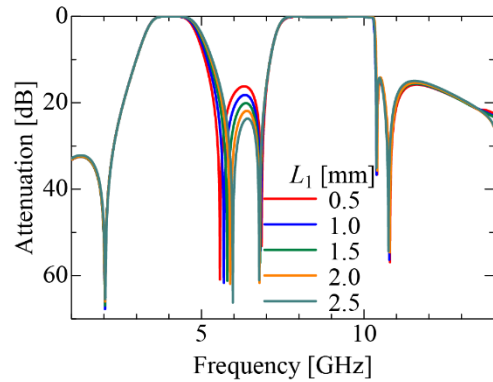
(b) Insertion loss

Figure 16. Frequency characteristics for various slot stub length L_1.

of the first transmission zero. Next, we notice of the second transmission zero of 5.5 GHz. Figure 14 shows the magnetic-current distribution at 5.5 GHz. We can see from this figure that the magnetic current strongly flows in the center slot connecting between the lower-part and higher-part slots. Therefore, if the slot stub is loaded into the center slot, the transmission zero at 5.5 GHz can move independently.

Figure 15 shows the slot stub of the length L_1 and the other parameters are $s_{17} = 0.33$ [mm], $s_{18} = 0.30$ [mm], $s_{19} = 0.33$ [mm]. Figure 16 shows the frequency characteristics for varying L_1. Although only the second zero point shifts to the higher frequency region with increasing L_1 and the stopband characteristic is improved as expected, the third resonant point also moves to the higher region, and the passband characteristic slightly deteriorates. The third resonance at 7.6 GHz is generated by the bottom slot as mentioned in Sec. 2, but if the length of the resonator is made longer, the second and the forth resonances also a little bit move, so such a deformation of the bottom slot for characteristic improvement is difficult. Therefore, we install a small projection of the length L_2 in the upper slot resonator shown in Fig. 17. The projection is coupled with the lower slot resonator so that only the

(a) Patch shape   (b) Slot shape
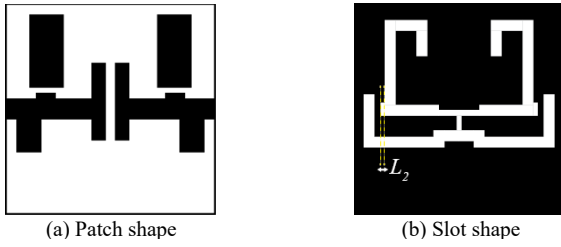
Figure 17. Shape of the filter with an additional projection of the length L_2
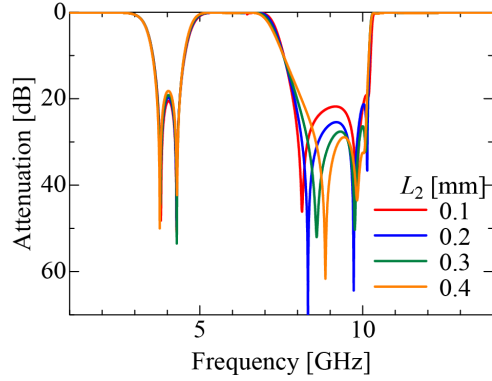
(a) Return loss
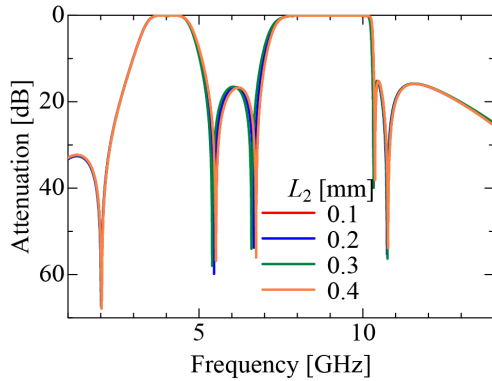
(b) Insertion loss

Figure 18. Filter characteristics for various projection length L_2.

(a) Patch shape   (b) Slot shape
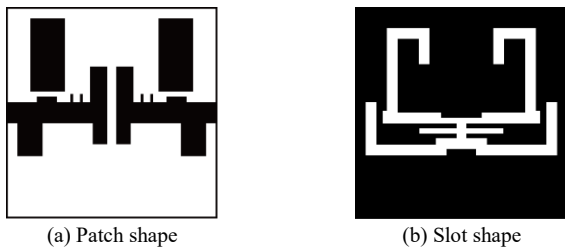
Figure 19. Proposed filter structure loading two conductor stubs.
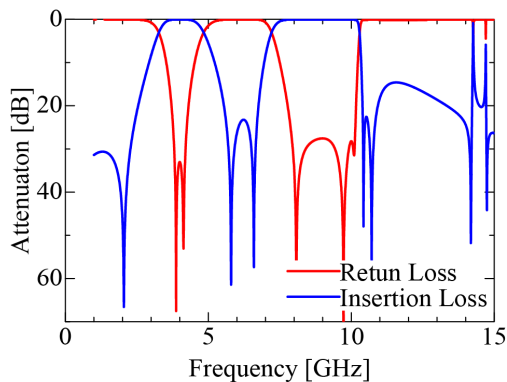
Figure 20. Filter characteristics of the proposed filter.

third resonance can move. Figure 18 shows the transmission characteristics for varying value of L_2. Clearly only the third resonance moves and other resonances, and the transmission zero points don't move. Furthermore, two small conductor stubs are loaded on the input/output line in order to improve the transmission characteristics at around 10 GHz as shown in Fig. 19. Figure 20 shows the transmission characteristics of the filter proposed under the above-mentioned investigation and, the return losses in the two passbands are suppressed to 28 dB, and the insertion loss in the stopband is suppressed to 20 dB or more. Most of UWB filters do not reach to these-losses levels in three pass and stop bands.
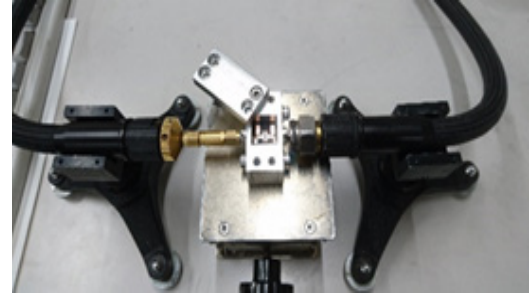
Figure 21. Measurement system.

(a) Patch shape   (b) Slot shape
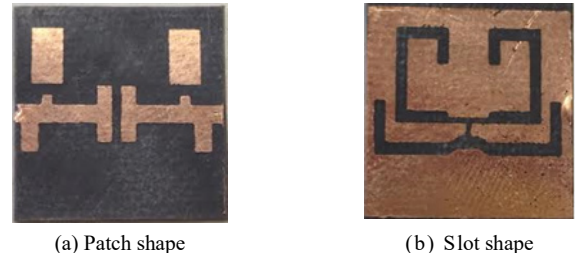
Figure 22. Photograph of the simplified filter.
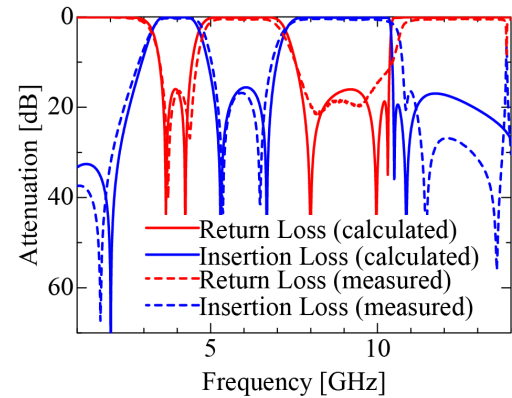
Figure 23. Measurement frequency characteristics of the simplified filter.
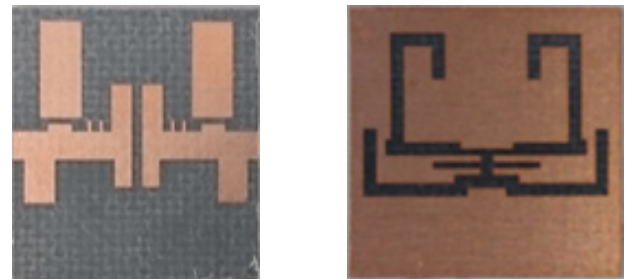
(a) Patch shape   (b) Slot shape

Figure 24. Photograph of the proposed filter.
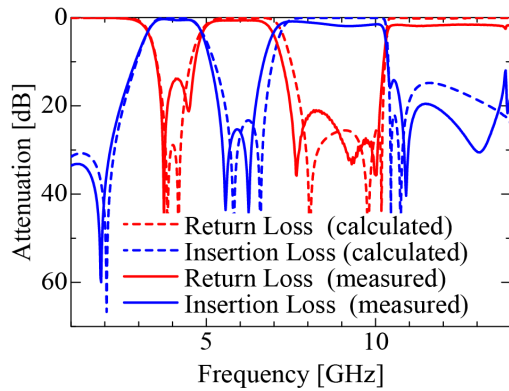
Figure 25. Measured frequency characteristics of the proposed filter.

4. Experimental results

To confirm effectiveness of the proposed filters, we fabricate two kinds of them by the photo-etching technique. As a dielectric substrate, we use Rogers RT/duroid 5880 with dielectric constant $\varepsilon_r = 2.2$ and the thickness $t_2 = 0.508$ [mm], and also measure the filter characteristics in the shielding box (see the photograph of the measurement system in Fig. 21) by the vector network analyzer (Agilent Technologies Inc. E8361C). The first fabricated filter has the simplified structure shown in Fig. 10 and its photograph is shown in Fig. 22. Comparison of the transmission characteristics between the measured and the calculated results are performed in Fig. 23, where the solid lines indicate the measured results and the dashed lines are the calculated ones. Both results are in good agreement with each other in the insertion and the return losses. The return losses in the two passbands are suppressed to 15 dB, and the insertion loss in stopband is also suppressed to 15 dB. On the other hand, the second fabricated filter has the structure with further improvement of the filter characteristic shown in Fig. 20 and its photograph is shown in Fig. 24. The measured results of the transmission characteristics indicated by the solid lines are compared with the calculated ones of the dashed lines in Fig. 25. The measured return loss in the low-frequency passband is about 15 dB lower that the calculated one, whereas that in the high-frequency passband is more than 20 dB and also the measured insertion loss in the stopband dividing two passbands is more than 25 dB. Although there is some deviation between both results due to the fabrication error for the small patch stubs and the slot stub added to the simplified structure. We can confirm the obvious improvement of the filter characteristics experimentally.

5. Conclusion

In this paper, we have clarified from the electric fields that two passbands in the GA-optimized filter developed by us are constructed by five resonant modes and have proposed the simplified structure based on the resonant mechanism. For further improvement of the simplified filter, the mechanism of the transmission zero is investigated from distributions of the magnetic currents. So that by newly loading both the slot stub and the conductor stub, the return loss more than 28 dB in two passbands and also the insertion loss more than 20 dB in the stopband inbetween passbands have been achieved. Finally, effectiveness of the proposed filter has been verified experimentally. To suppress spurious responses in the frequency region more than 14 GHz is the future work, although we don't investigate it in this paper.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

References

Conference Papers:

[1] Y. Takeuchi, R. Nakano, H. Deguchi, M. Tsuji "Improvement of transmission characteristics in multilayer dual band filter," 2017 IEEE Asia Pacific Microwave Conference (APMC) 13-16, Nov. 2017 https://doi.org/10.1109/APMC.2017.8251528.

[2] T. Kido, H.Deguchi, M.Tsuji "Compact planar bandpass filters with arbitrarily-shaped conductor ptches and slots," IEICE Trans. Electron., vol. E94-C, no. 6 pp.1091-1097, June 2011 https://doi.org/10.1587/transele.E94.C.1091.

[3] L. Zhu, S. Sun, W. Menzel "Ultra-wideband (UWB) bandpass filters using multiple-mode esonator," IEEE Microw. Wireless Compon. Lett., vol. 15, no. 11, pp. 796-798, Nov. 2005 https://doi.org/10.1109/LMWC.2005.859011.

[4] R. Li, L. Zhu, "Compact UWB bandpass filter using stub-loaded multiple-mode resonator," IEEE Microw. Wireless Compon. Lett., vol. 17, no. 1, pp. 40-42, Jan. 2007 https://doi.org/10.1109/LMWC.2006.887251.

[5] L. Han, K. Wu, X.P. Chen, "Compact ultra-wideband bandpass filter using stub-loaded resonator," Electron. Lett., vol. 45, no. 10, pp. 504-506, May 2009 https://doi.org/10.1049/el.2009.0510.

[6] Y. Zhou, B. Yao, Q. Cao, H. Deng, X. He, "Compact UWB bandpass filter using ring open stub loaded multiple-mode resonator," Electron. Lett., vol. 45, no. 11, pp. 554-556, May 2009 https://doi.org/ 10.1049/el.2009.3724.

[7] H. W. Deng Y. J. Zhao L. Zhang X. S. Zhang S. P. Gao "Compact quintuple-mode stub-loaded resonator and UWB filter," Progress In Electromagnetics Research Letters., vol14, no. 8, pp. 111-117, 2010 https://doi.org/10.2528/PIERL10030912.

[8] K. Song, Q. Xue, "Novel broadband bandpass filters using Y-shaped dual-mode microstrip resonators," IEEE Microw. Wireless Compon. Lett., vol. 19, no. 9, pp. 548-590, Sept. 2009 https://doi.org/10.1109/LMWC.2009.2027058.

[9] L. Han, K. Wu, X. Zhang, "Development of packaged ultra-wideband bandpass filters" IEEE Trans. Microw. Theory Tech., vol. 58, no. 1, pp. 220-228, Jan. 2010 https://doi.org/10.1109/TMTT.2009.2036399.

[10] Q.-X. Chu, X.-K. Tian, "Design of UWB bandpass filter using stepped-impedance stub-loaded resonator," IEEE Microw. Wireless Compon. Lett., vol. 20, no. 9, pp. 501-503, Sept. 2010 https://doi.org/10.1109/LMWC.2010.2053024.

[11] J. Xu, W. Wu, W. Kang, C. Miao, "Compact UWB bandpass filter with a notched band using radial stub loaded resonator," IEEE Microw. Wirel. Compon. Lett., vol. 22, no. 7, pp. 351-353, July 2012 https://doi.org/10.1109/LMWC.2012.2201930.

[12] Z.X. Zhang, F. Xiao, "An UWB bandpass filter based on a novel type of multi-mode resonator," IEEE Microw. Wirel. Compon. Lett., vol. 22, no. 10, pp. 506-508, Oct. 2012 https://doi.org/10.1109/LMWC.2012.2218229.

[13] H. Zhu, Q.-X. Chu, "Compact ultra-wideband (UWB) bandpass filter using dual-stub-loaded resonator (DSLR)," IEEE Microw. Wireless Compon. Lett., vol. 23, no. 10, pp. 527-529, Oct. 2013 https://doi.org/10.1109/LMWC.2013.2278278.

[14] L.C. Lin, S. Yang, S.J. Sun, C.H. Liang, "Ultra-wideband bandpass filter using multi-stub-loaded ring resonator," Electron. Lett., vol. 50, no. 17, pp. 1218-1220, Aug. 2014 https://doi.org/10.1049/el.2014.1256.

[15] X. Li, X. Ji, "Novel compact UWB bandpass filters design with cross-coupling between short circuited stubs," IEEE Microw. Wirel. Compon. Lett., vol. 24, no. 1, pp. 23-25, Jan. 2014 https://doi.org/10.1109/LMWC.2013.2287231.

[16] A. Taibi, M. Trabelsi, A. Slimane, M. T. Belaroussi, J. P. Raskin, "A novel design method for compact UWB bandpass filters," IEEE Microw. Wireless

Compon. Lett., vol. 25, no. 1, pp. 4-6, Jan. 2015 https://doi.org/10.1109/LMWC.2014.2363016.

[17] S. Kumar, R. D. Gupta, M. S. Parihar, "Multiple band notched filter using C-shaped and E-shaped resonator for UWB applications," IEEE Microw. Wireless Compon. Lett., vol. 26, no. 5, pp. 340-342, May 2016 https://doi.org/10.1109/LMWC.2016.2549700.

[18] M. Tsuji, H. Deguchi, A. Kido, M. Ohira "Multi-resonator generation by genetic optimization for application to panar-circuit bandpass filters," Special issue on recent progress in computational electro magnetics and its applications. Volume 129 (2009) Issue 10 Pages 681-686 https://doi.org/10.1541/ieejfms.129.681

[19] A. Kido, H. Deguchi, M. Tsuji, M. Ohira "Multi-resonator generation in arbitrarily-shaped planar-circuit filters by genetic optimization," 2007 European Microwave Integrated Circuit Conference 8-10 Oct. 2007 https://doi.org/10.1109/EMICC.2007.4412744

An Optical-based Fingertip Force Sensor

Emanuele Lindo Secco[*], Taye F. Agidew, Atulya Kumar Nagar

Robotics Laboratory, Department of Mathematics and Computer Science, Liverpool Hope University

A R T I C L E I N F O	A B S T R A C T
	This work introduces the design and prototype development of a novel optical-based fingertip force sensor, which is integrated in a bio-mimetic finger for robotic and prosthetic manipulation. Given the optical nature of the sensor, the proposed device is free of any electrical and metal components. Accordingly, the design improves manipulation of objects with high electromagnetic compatibility performance, reducing sensor's susceptibility in the presence of these disturbances. The sensor has inherently high SNR and low-power consumption. The concept of using simplified and low cost optical based fingertip force sensor for bio-mimetic anthropomorphic fingers has not been widely implemented so far. The sensor is based on a cantilever design combined with a couple of fiber optics and a reflective surface, it is integrated on the distal phalange of a novel bio-mimetic and anthropomorphic robotic finger. Sensor design was performed and optimized throughout a 3D printing process and Finite Element Analysis. The sensor's sensitivity (0.098V/N), resolution (0.01N), accuracy (3% of full scale output (FS)) and hysteresis (9.24% of FS) were characterized through a calibration processes. Various applications like surgical manipulation or handling in harsh environment (i.e. high voltage, high temperature and explosive environments) will benefit from the proposed sensor's performance, reliability, simplicity and bio-compatibility.

1. Introduction

Many multi-fingered robotic hands with multi-axial force and torque sensors have been developed. These multi axial force sensors usually employ strain gauges driven by electrical circuitry and are made of metal components [1]. The use of sensors with electrical circuit and metallic components increases the sensor's vulnerability to electrical or magnetic fields [2]. In contrast with multi-axial force sensors based on strain gauge technology, sensors designed based on the use of fibre-optics has several benefits including simplified design, immune to electromagnetic fields, miniaturization, low-level noise, has low power consumption and no need for any electronic noise filtering [3]. In this context, previous research has already proposed interesting applications on the development of force and tactile sensors with the use of metal free components and biologically friendly solutions which, moreover, will also guarantee compatibility in medical scenarios [2-3].

This work has been proposed due to the following limitation observed from current fingertip force measuring systems:

- Complex tactile system design
- High electromagnetic vulnerability
- Unavailability of simplified and low cost optical based fingertip force sensor design for biomimetic anthropomorphic fingers.

The concept of using simplified and low cost optical based fingertip force sensor for biomimetic anthropomorphic fingers has not been widely implemented so far. The proposed fingertip force sensor takes advantage of the use of fibre-optics and is integrated in a biomimetic finger in order to get tactile information during grasping and manipulation of objects. The objective of this work is designing optical based fingertip force sensor. The design involves only the distal phalange of the finger. The phalange has the same shape, arrangement, length and proportion as its human counterpart [4].

2. Sensor Design

To exactly imitate the size and shape of the human index finger bones, the laser-scan model of human finger was used. The Stereo Lithography (STL) format of the model was first imported into

Autodesk123 (Autodesk, Inc.). In this stage the original file was further modified by introducing a cantilever beam design, holes for placement of optical fibre and grove between sensing head and reflective surface (Figure 1).The design involves optical fibre, mirror and built-in cantilever beam [2]. After modifying the file, it was further analysed in Autodesk Mesh mixer to reduce the number of meshes [5] in the model. The aim of mesh reduction was to easily simulate Finite Element Analysis (FEA) of the model in Solid Works (Dassault Systèmes Solidworks Corp) [6]. Finally the model with reduced mesh (Figure 2) was exported to Solid Works and the FEM analysis was performed. The standard parameters of phalanx of index finger [7] are shown in Table 1.
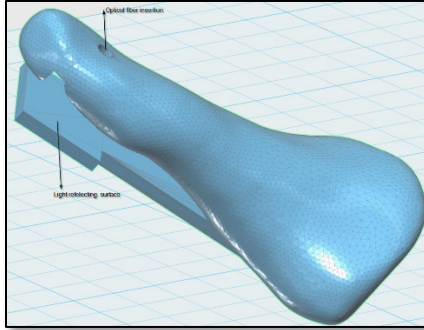
Figure 1: Built-in cantilever beam and optical fibre holes: the wired fibre will be inserted through the hole in order to design the tactile sensor (see also Figures 4 and 5)

Table 1: Parameters of the index bio-mimetic finger: length and weight of the index finger phalanges

Index finger phalanx	Length(mm)	weight(g)
MCP to PIP	53.4	5.5
PIP to DIP	32.0	2.0
Distal phalanx	23.7	1.2

2.1. FEA Simulation

The purpose of FEA simulation is to ensure how much amount of force the fingertip can resist without undergoing plastic deformation. First geometry or shape of finger phalanx was defined. Next, an appropriate material (i.e. Acrylonitrile Butadiene Styrene, ABS) for the model has been selected. Mechanical properties [8] of ABS have been used as input for the FEA simulation (Table 2).
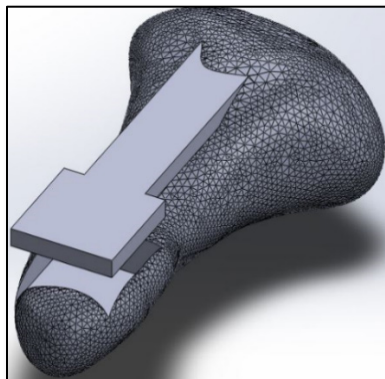
Figure 2: Reduced mesh of the distal phalanx of the index finger: the mesh was distributed along the fingertip in order to perform the FEM

Table 2: Mechanical Properties of the ABS material

Property	Value	Units
Elastic Modulus	2000	N/mm^2
Poisson's Ratio	0.35	
Shear Modulus	318.9	N/mm^2
Mass Density	1020	Kg/m^3
Tensile Strength	40	N/mm^2
Compressive Strength	42	N/mm^2

The mean maximum external finger force (i.e. 5[N]) for static prehension is applied to the phalanx of index finger [9]. Finally FEA analysis of the 3D model been performed. The description of how the applied forces spread throughout the phalanx is illustrated in Figure 3.
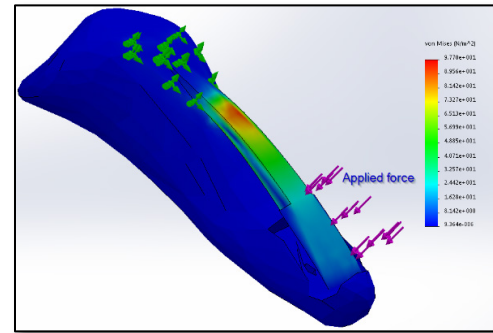
Figure 3: FEM simulation of the cantilever (ABS material, load of 5[N]): force is applied to the cantilever in order to estimate Von Mises stress

The red portion of the simulation indicates high stress concentration areas and the blue part shows unloaded or stress free region of the phalanx Figure 3. According to the finite element simulation result the maximum stress applied on the beam due to the applied force is much lower than, the yield strength of ABS. This means the applied force doesn't produce mechanical failure or damage to the built in cantilever beam of the phalanx and it shows the design is safe.
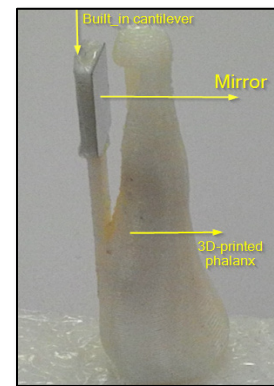
Figure 4: 3D printed distal phalanx of index finger: the mirror allows measuring the cantilever flexion according to the applied load and therefore to transduce such a flexion into a modulation of the intensity light provided through the fibre (see also Figures 1 and 5)

2.2. 3D Printing and sensor component integration

After the FEM simulation, the real distal phalanx of index finger was fabricated using HP 3D printer. The phalanx has built-

in cantilever beam used for positioning light reflective surface. The movement of this beam moves the mirror and modulates the intensity of reflected light back to the receiving terminal of the optical fibre. The 3D printed phalanx is shown in the Figure 4.

2.3. Working principle of the sensor

The light beam is projected on a moving cantilever beam reflective surface by fibre optic cable (i.e. the emitter), a fraction of the reflected beam is received by the second optic fibre (i.e. the receiver) that are closely located [1] (Figure 5). The reflected light intensity is a function of the amount of force applied at the fingertip. The applied force changes the distance between the reflective surface and the sensing head of optical fibre cables. This change modulates the intensity of reflected light and produces different voltage outputs. The amount of output voltage from the sensor is directly proportional with the force applied at tip of the phalanx.

3. Sensor Calibration

To use the designed sensor, the sensor has to be calibrated. The calibration should convert force applied to the fingertip into corresponding voltage [2].

3.1. Calibration set-up

The experimental consists of a Keyence FS-N11MN Digital Fibre Optic Amplifier, IIT-FT17 force (IIT-FT17 force & torque sensor), which is combined with a torque (F/T) sensor: sensor data are acquired via an NI USB-6000 Data AcQuisition (DAQ) board (National Instruments Inc.) while the sensor is mounted on a mechanical vice as it is reported in Figure 6. The phalanx of the finger is fixed on the vice, while the force sensor is physically connected with the moving element of the vice. This solution allows to manually modulate the effective applied force to the phalanx, i.e. to the sensor. During the experiments, the electrical output of the FS-N11MN amplifier is sent to an analogue pin of the DAQ. Then, a LabView (National Instrument, NI) graphical interface captures the output signal in real-time, while another program – developed in C++ Programming Language – acquires the force sensor outputs. This further measurements are processed via another LabView mask. These data (i.e. output of the two force sensors) are saved and processed.
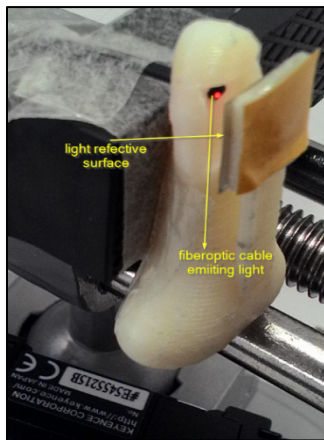
Figure 5: Fingertip fibre optic cable and mirror arrangement (see also Figures 1 and 4)

3.2. Calibration procedure

Different trials were performed while applying force to the sensor and releasing the load, thanks to the linear guide which was provided by the vice in Figure6. The IIT-FT17 force and torque (F/T) sensor was used to compare the reading vs. the fibre-optic sensor and therefore calibrate this latter one. Different values of force were applied. Precisely, a sequence of 15 values of forces were applied during the 'loading' and 'unloading' stages. The NI DAQ board was used to acquire and save the value of the signals (i.e. the voltage) from the KEYENCE amplifier and the reference F/T sensors. Signals were acquired every 2 ms, namely at a sampling frequency of 500 Hz.
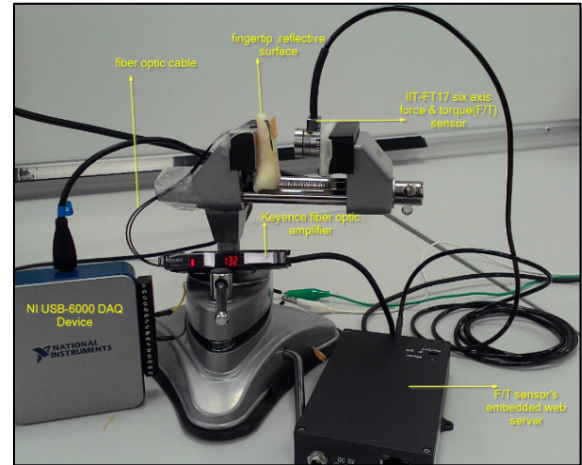
Figure 6: Force sensor calibration setup: load is applied to the fingertip and sensor output is measured accordingly (see the details on Paragraph 3)

4. Sensor Performance

4.1. Regression analysis

Following the experiment and data acquisition, all information were re-shaped into a data base format to be analysed and to identify a proper calibration between the applied load and the voltage output of the sensor. According to a preliminary qualitative analysis of the data, linear regression method was adopted in order to find a proper fitting of the data [10]. The electrical signal was correlated with the applied force on the finger, finally, a fitting curve was obtained (Figure 7). Here, the linear regression equations are reported:

$$V = 0.098 \cdot F + 1.10 \tag{1}$$

$$F = 10.22 \cdot V - 11.29 \tag{2}$$

where V and F are representing the sensor signal and applied force, respectively.

Thanks to the Equation (2), the effective force can be calculated, provided that the output voltage of the sensor is measured. Moreover, these equations define the sensor sensitivity, namely 10.22 N/V [8]. A regression analysis shows that the relationship between the force and the output of the sensor is quite linear with a coefficient of determination ($R^2 = 0.9739$): more than 97.39 % of the electric voltage is due to the load which is applied to the fingertip. An analysis on the standard deviation of the

prediction of the value of the force provides a result of 1.86%: a small deviation occurs between the data vs. the actual values.
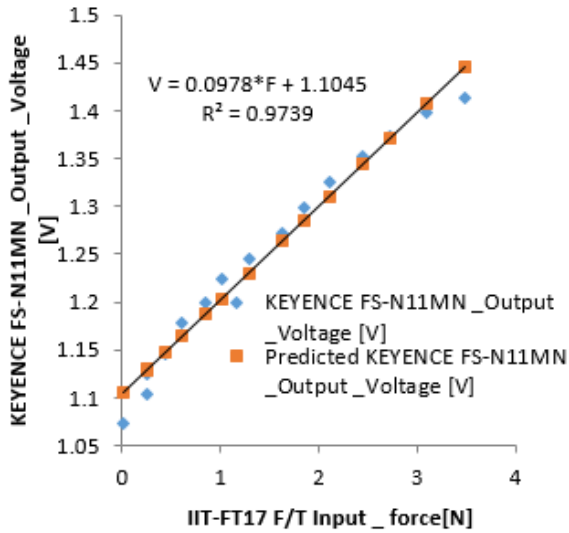
Figure 7: Calibration curve of force sensor performed with a regression analysis (see details in Paragraph 4)

4.2. Sensor accuracy

In this context accuracy is measure of closeness between actual sensor readings and its corresponding predicted value. The sensors accuracy is measured by its relative and absolute errors. The sensor relative error is the difference between the actual voltage reading obtained from KEYENCE FS-N11MN Fibre Optic Sensor and the corresponding predicted voltage value obtained from fitted linear equation during the calibration procedure. Table 4 shows the actual and predicted voltage values. Based on the data given on Table 4 percentage of sensor's maximum relative error is 3.1097%. This means the sensor introduces maximum of ±0.03109 error from each of its readings and is indication of sensors better accuracy.

Table 3: Results of the Regression Analysis which brought to the linear calibration expressions (see details in Paragraph 4)

Parameter	Value
Multiple R	0.9868
R^2	0.9739
Adjusted R^2	0.9719
Standard Error	0.0184
Observations	15

4.3. Sensor range & resolution

The designed sensor works over an input range between 0[N] and 4[N]. If exceeded the specified design range ,it result in permanent damage or destruction of sensor. The corresponding output range of the sensor is between 1.1045 [V]and 1.6[V].

The smallest input force value the senor can detect [11] is 0.01[N]. In other words the sensor can discriminate input force values above 0.01[N].

4.4. Sensor sensitivity & saturation

It is change in output voltage of the designed sensor per unit change in input force. The slope of fitted linear regression equation

(i.e. Equation 1 is sensitivity [11] of the sensor (i.e. 0.0978 V/N or 10.2249 N/V).

Table 4: Actual and predicted reading of the voltage outputs

Sensor's actual reading	Predicted value	Residual
1.070	1.110	-0.033
1.100	1.120	-0.250
1.120	1.130	-0.005
1.145	1.147	-0.002
1.170	1.160	0.013
1.198	1.186	0.012
1.224	1.203	0.021
1.245	1.230	0.015
1.272	1.2638	0.009
1.298	1.285	0.013
1.325	1.310	0.014
1.351	1.343	0.007
1.373	1.370	0.002
1.3973	1.406	-0.009
1.413	1.446	-0.033

As the input force applied to the fingertip increases the distance between optical fibre and reflective sensor decreases and sensor's output voltage increases. Beyond its input range (i.e. 4[N]) the sensor reaches its saturation limit. This means, any further increment of input force will no longer increase [11] output voltage. Sensor's saturation emerges from its physical construction.

Table 5: Voltage outputs of the 'loading' and 'unloading' stages

Input Force[N]	Loading[V]	unloading[V]	Residual
0.00	1.02	1.05	0.030
0.01	1.07	1.09	0.022
0.25	1.10	1.12	0.024
0.26	1.12	1.50	0.024
0.44	1.14	1.163	0.017
0.61	1.17	1.18	0.009
0.84	1.19	1.24	0.040
1.012	1.22	1.25	0.031
1.29	1.24	1.27	0.024
1.62	1.27	1.29	0.020
1.85	1.29	1.31	0.019
2.109	1.32	1.33	0.012
2.45	1.35	1.36	0.01
2.73	1.37	1.38	0.01
3.09	1.34	1.41	0.01
3.5	1.41	1.41	0.00

4.5. Zero offset & mechanical hysteresis

The zero offset is the amount of deviation in sensor's output voltage from the exact value at the lowest point of the measurement range. Zero offset can be expressed in signal units (i.e. Volts), or as a percentage of the full scale range. As it is shown in Equation (1) from the linear regression, the zero offset value for the designed sensor is 1.1045[V]. since the input range is from 0 to 4[N] and output range is 1.1045 [V] to 1.6[V], the zero offset tolerance of the sensor is 1.1045[V] ± 0.1238[V].

Mechanical hysteresis can affect the sensor performance, due to the elastic-plastic deflection of the cantilever beam affecting not only the instantaneous application of the load, but also the previous

application or history. Table 5 shows the voltage outputs during 'loading' and 'unloading'.
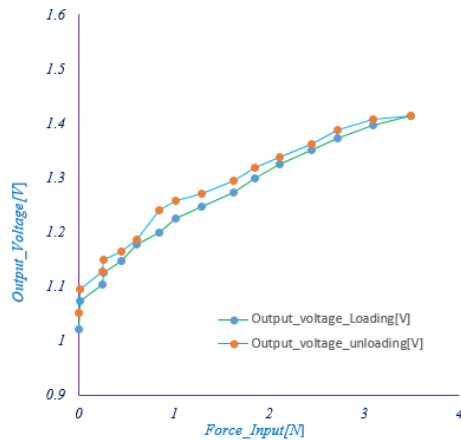
Figure 8: Sensor hysteresis curve

The data of Table 5 were collected during the calibration processes. These values are also plotted within Figure 8. According to these output, the 'loading' and ''unloading' curves provide different values of the sensor output values vs. same value of the applied force (i.e. the input). Therefore, the percentage of the hysteresis van be calculated as the ratio of the maximum difference between the output values curve and sensor range. The effective percentage of the maximum hysteresis error is 9.24%, namely each sensor measurements during unloading will be 9.24% greater from each loading measurements [11]. A summary of the overall sensor performance parameters is reported in Table 6.

Table 6: Summary of the performance parameters of the sensor

Sensor parameter	Value	Unit
Maximum percentage relative error	±0.03109	Full scale output(FS)
Span	[0 - 4]	N
Output range	[1.1045 - 1.6]	V
Resolution	0.01	N
Zero offset	1.1045	V
Zero offset tolerance	±0.123	V
Sensitivity	0.0978	V/N
Maximum Hysteresis error	±9.24%	FS

5. Conclusion and Future work

This works presented simplified design and integration of an optical force sensor to measure external applied force to the biomimetic fingertip during physical interaction. The FEA analysis of the phalange has been conducted to ensure that how much amount of force the fingertip can resist without undergoing plastic deformation. Then the actual finger has been fabricated using HP 3D printing machine. The structural components of the sensor (i.e. cantilever beam, reflective surface and the two optical fibre) were integrated at the fingertip. Sensor calibration and linear regression analysis has been conducted. The regression analysis shows high correlation between applied force and corresponding voltage output from the digital fibre optic sensor. The sensors performance parameters have been briefly addressed. Various applications like surgical operation, manipulations in area of high electromagnetic noise can be benefited from the proposed sensor's performance (i.e. sensitivity, accuracy, linearity and resolution) and its EMC design. As a future work the structural components

and arrangement of the sensor can be further optimized to get a better performance and improved sensor characteristics. The deployment of optical fibre, shape of the reflective surface and structure of built-in cantilever beam can be further modified and improved.

Acknowledgment

References

[1] Konstantinova, J, Stilli, A, Althoefer, K, Force and proximity fingertip sensor to enhance grasping perception. Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on. IEEE, 2015, 2118–2123.

[2] Noh, Y, Sareh, S, Back, J, Würdemann, HA, Ranzani, T, Secco, EL, Faragasso, A, Liu, H, Althoefer, K. A three-axial body force sensor for flexible manipulators. Robotics and Automation (ICRA), 2014 IEEE International Conference on. IEEE, 2014, 6388–6393.

[3] Noh, Y, Secco, EL, Sareh, S, Würdemann, HA, Faragasso, A, Back, J, Liu, H, Sklar, E, Althoefer, K, A continuum body force sensor designed for flexible surgical robotics devices. Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE. IEEE, 2014, 3711–3714.

[4] Xu, Z, Todorov, E, Design of a highly biomimetic anthropomorphic robotic hand towards artificial limb regeneration. Robotics and Automation (ICRA), 2016 IEEE International Conference on. IEEE, 2016, 3485–3492.

[5] Salinas, R, 3D Printing with RepRap Cookbook; Packt Publishing Ltd, 2014.

[6] Akin, JE, Finite element analysis concepts: via SolidWorks; World Scientific Publishing Co Inc, 2010.

[7] Xu, Z, Kumar, V, Matsuoka, Y, Todorov, E, Design of an anthropomorphic robotic finger system with biomimetic artificial joints. Biomedical Robotics and Biomechatronics (BioRob), 2012 4th IEEE RAS & EMBS International Conference on. IEEE, 2012, 568–574.

[8] Callister, WD, Rethwisch, D.G. Materials science and engineering; Vol. 5, John Wiley & Sons NY, 2011.

[9] Radwin, RG, Oh, S, Jensen, TR, Webster, JG, External finger forces in submaximal five-finger static pinch prehension. Ergonomics 1992, 35, 275–288.

[10] Draper, NR, Smith, H, Pownell, E, Applied regression analysis; Vol. 3, Wiley New York, 1966.

[11] Fraden, J, Handbook of modern sensors; Vol. 3, Springer, 2010.

[12] Agidew TF, Secco EL, Maereg AT, Reid D, Nagar AK, Fibre Optic-Based Force Sensor for Bio-Mimetic Robotic Finger, IEEE Sensors 2017, Glasgow, UK.

Acquiring Wi-Fi Energy to Charge a Mobile Phone

Mohamed Zied Chaari*,[1], Mohamed Al-Kuwari[1], Rashid Al-Rahimi[1], Mongi Lahiani[2], Hamadi Ghariani[2]

[1]*FAB LAB, Qatar Scientific Club, 9769, Doha- Qatar*

[2]*National School of Engineers of Sfax, ENIS, 3000, Sfax-Tunisia*

A R T I C L E I N F O	A B S T R A C T
	As a result, outdoor travelers, tourists and businessman who utilize portable electronic devices often experience battery depletion of their devices without the means of charging their device. The number of Wi-Fi access point continues to increase. There are many occasions where it is desirable to acquire energy from many radio frequency stations, while charging any electronics device through RF-DC circuit with high efficiency, for example, Wi-Fi hotspot access points. We focus on ambient radio frequency signals and power available in the ambient. This research explain, how can we charge portable devices without electrical source but means of propagating signal generated by Wireless-Fidelity (Wi-Fi) router. The research will show all the activities addressed to design a system to recover energy from electromagnetic sources presents around us like Wi-Fi routers. An energy acquiring circuit includes one loop antenna to detect RF signals and electronic board for converting RF energy to DC current. We can use to charge the battery of our portable devices by acquiring the electromagnetic energy around us.

1 Introduction

Particularly Wi-Fi routers are increasingly used in many places such as coffeehouse, airport, school, mall, etc. So, Wi-Fi access point is available and easy to find in any public place, at home and almost anywhere. There are many high powered Wi-Fi stations whose transmitted signals that create a Wi-Fi hotspot to be implemented are projected to be rather strong signals [1]–[6] in the future. Currently, when we experience our portable devices has no battery charge, we lost an important communication with people around us. We usually encounter this situation when we travel or vacation, at the airport, coffee shop, business trip, etc. Many of us dont know or neglect that energy from high-frequency radio (RF) signal can be acquire and used as a power supply to our devices. But in this research show how useful it is. The main objective of this work is to acquire the Wi-Fi signals and converting the electromagnetic energy to a DC voltage [7]–[9]. Providing to charge the batteries of an external electronic device connected thereto, or to continuously charge an internal battery within the energy acquiring device until its storage capacity is full, and that battery energy can be reused to power our electronic devices, such as smartphone. From Wi-Fi hotspot and convert the received signal into DC voltage [10]–[15]. RF energy acquiring is acquiring the microwave energy around us and converting it into usable electrical power for electronic devices as shown in Figure 1.
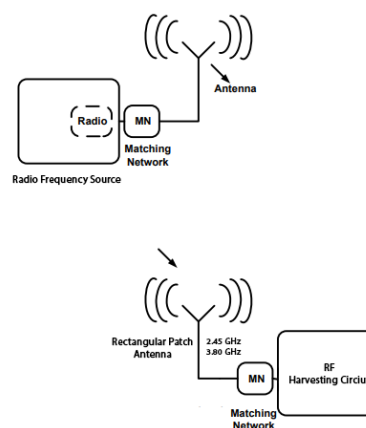
Figure 1: Conceptual view of radio-frequency (RF) energy harvesting

This high technology is also named as power scavenging and power harvesting. The evident summons

*Corresponding Author: Mohamed Zied Chaari, Doha-Qatar, chaari_zied@ieee.org

of acquiring ambient RF energy is that it is essentially free energy. The purpose of this research is to explain and to implement one band of the center operating frequency ($f_0 = 2.45 GHz$). One square loop antenna can collect maximum microwave energy from Wi-Fi router with high efficiency and high directivity. The reflected microwave energy converted from microwave signal to DC power. The results are satisfactory and show the advantage of using a square antenna for acquiring microwave energy. The efficiency of the antenna is related to the shape and impedance of the antenna. In this research, we focus on the possibility of charging the mobile phone.

The purpose of this work is related to energy conservation methods, and more particularly related to charging the battery of mobile phone by RF acquiring energy as illustrated in Figure 2.
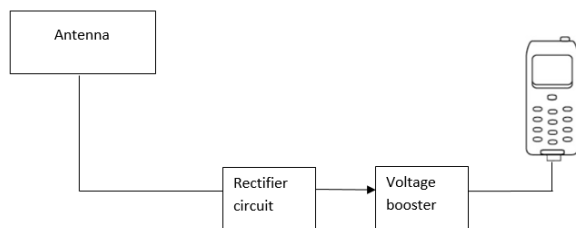
Figure 2: Conceptual view of WIFI harvesting system

This research is organized as follows: in Section 2: We present the measurements of radio frequency exposure from Wi-Fi router. Section 3: The antenna designed method. Section 4: Presents the converter of Wi-Fi signal to DC power. Section 5: We present simulation work and experimental results of fabricated prototype with analyses of data. We describe the simulation and results. Finally, Section 6: Conclusions.

2 Measurements of Radio Frequency Exposure from Wi-Fi router

First of all we should be measures the Wi-Fi signal that people are exposed. The measurements of radio frequency exposure generated by Wi-Fi router using the spectrum and real-time spectrum analyzer SRM3006 [16,17,18,19]. It measures radiating sources at frequencies from 300 MHz to 3 GHz, but it does not display the frequency. We can measure the Wi-Fi field strength in Volts per meter (V/m), and the time-average RF power in uW/m2 as shown in figure 3.

We are measure the quantity of propagated energy around us and we observe that the quantity is very high, it is around 600 mW/m2 at operating frequency ($f_0 = 2.45 GHz$). The results obtained after 10 iterations are shown in Figure 4. The main objective of the research is to supply a device that converts the acquired energy (600 mW/m2) from Wi-Fi router into DC (direct current) with high efficiency. The DC energy from which is stored in a super-capacitor (15F/5.6V),

and has the capability to connect into electrical power source for restoring power [20]–[25].

Figure 3: Measurements of radio frequency exposure from Wi-Fi access point
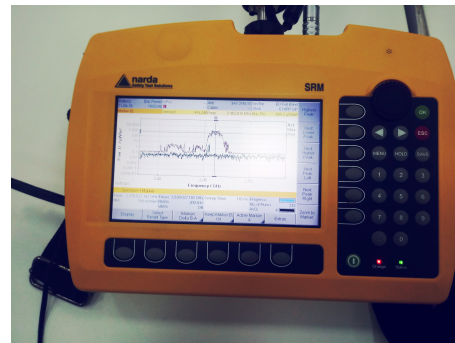
Figure 4: Experimental setup (Wi-Fi Spectrum analysis)

In this research, a provide device for acquiring energy from Wi-Fi access point to power an external electronic device connected thereto smartphone. It is yet a further objective of the present research to supply device for acquiring energy from WIFI power received and convert into a DC (direct current) with high efficiency. The power from which is stored in a supercapacitor (15F/5.6V) device within the device, and which further has the capable to connect into a source of electrical power for storing power. The main objective of the present work is to provide a device which acquires energy from the Wi-Fi router. The bandwidth from 1.45 GHz to 3.6 GHz and the center frequency is ($f_0 = 2.45 GHz$). Wi-Fi spectrum analysis allows identifying the power of the signals sources then we can conclude.

3 Wi-Fi harvesting energy antenna

The antennas generating RF electrical signals in response to the transmitted RF signals received that way and only rectifier circuits [26,27,28,29,30]. The loop antenna collects all energy radio frequency around us and generating a direct current (DC). The radiation

field of the loop antenna could be determination using an electric current model. In the electric current model, the current is used directly to find the far-field radiation pattern. The sketches of the square loop antenna is shown below in Figure 5.
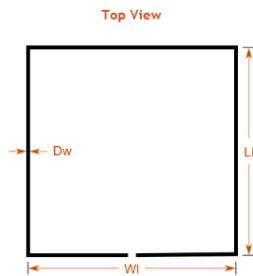
Figure 5: Geometry of the proposed square loop antenna

The bandwidth of the loop antenna is from 1.45 GHz to 3.6 GHz and the center frequency is (f_0 = 2.45GHz).

3.1 Equivalent Circuit of a Loop Antenna

Frequently, it is assumed that the loss resistance of loosely harvesting antenna equals the high-frequency loss resistance of a right copper wire of the same length as the loop and of the same current distribution [31,32,33,34,35,36] as shown in Figure 6. In the case of a uniform current distribution, the high-frequency resistance is calculated as

$$R_h = (l/P) \times R_s \qquad (1)$$

where :
l : length of the wire,
p : perimeter of the wires cross-section.
C_r : resonance capacitor,
R_l : loss resistance of the loop antenna,
R_r: radiation resistance,
L_a: inductance of the loop,
L_i: inductance of the loop conductor (wire).
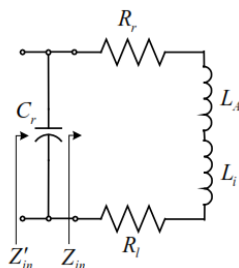
Figure 6: Equivalent Circuit of a Loop Antenna

3.2 Design Parameters

The parameters selected for design a square loop antenna are shown in Table 1.

Table 1: Dimensions of loop antenna

Name	Parameters	Value
Ll	Loop length	36.56 mm
WI	Loop width	36.56 mm
DW	Wire diameter	2 mm
FW	Feed section width	0.66 mm

3.3 Simulated and Analysis

In order to construction loop antenna, we use copper wire diameter of 2 mm and loop length equal approximately to the width 36.56mm. The antenna is designed and simulated using Antenna Magus and CST studio microwave software which uses FEM and tetrahedron method to solve and simulate complex 3D electromagnetic (EM) efficiency and uses a feature of the graphical user interface. After simulated our antenna on Antenna Magus and CST software, we calculated parameters like Gain, VSWR, SMITH, Radiation pattern 3D etc.

3.3.1 Gain PLOT

The reflection coefficient (dB) of the proposed loop antenna showing the gain at 2.419 GHz is -26.95 dBi. The gain plot can be seen in Figure 7.
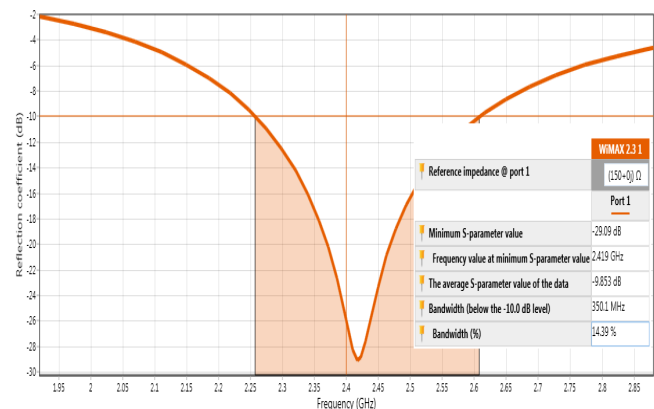
Figure 7: Reflection coefficient Vs frequency (GHz)

3.3.2 VSWR PLOT

VSWR is a measure of how well matched the loop antenna to the SMA connector impedance 50Ω. A perfect result would have a VSWR of 1:1. VSWR obtained from the simulation on antenna magus and CST software is 1.073 at 2.419 GHz which approximately equals to 1:1 as shown in Figure 8. This present the perfect impedance matching of the loop antenna with the port at 50Ω.
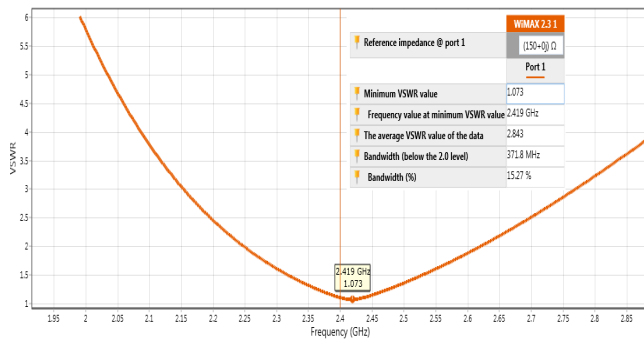
Figure 8: VSWR of proposed antenna

3.3.3 SMITH PLOT

The Smith Chart is very important to estimate the performance of the harvesting antenna [43]. Smith Chart of proposed antenna shows at radiation frequencies 2.419 GHz, inductance is very low. So, Impedance is purely resistive circuit were shown in Figure 9.
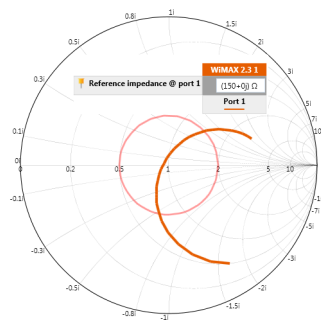
Figure 9: Smith chart of proposed antenna

3.3.4 Radiation pattern 3D

Figure 10 shows the radiation pattern simulated result of gain for the proposed harvesting antenna (maximum E-plane 3.53 dB and minimum E-plane -36.5 dB).
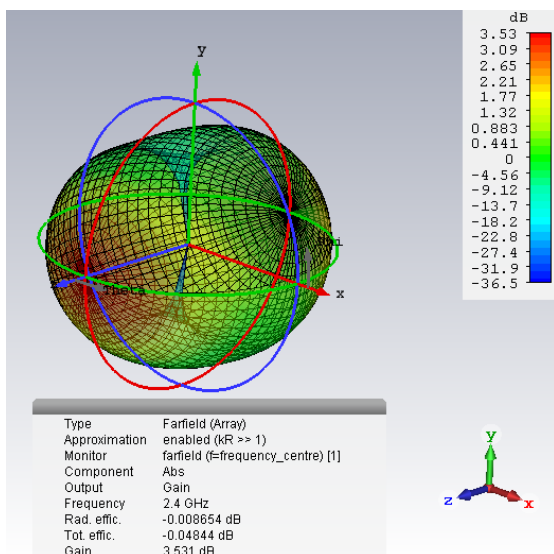
Figure 10: Radiation pattern 3D

3.3.5 Current distribution

Figure 11 shows the simulated current distribution of the proposed antenna for different polarization at different phases..
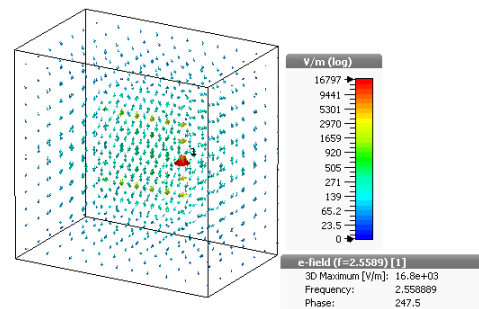
Figure 11: Current distribution of the harvesting antenna

4 RF-DC converter

Figure 12 is a schematic diagram of a first preferred form of an electrical circuit constructed in accordance with present work for acquiring energy from Wi-Fi bands. The energy acquiring circuit, in accordance with this first embodiment, includes an antenna which receives Wi-Fi source, and converts the microwave signals to an electrical signal an RF-to-DC rectifier circuit [44,45]. The rectifier circuit converts the RF signal to a DC (direct current) signal on its output. This DC signal is provided to a first storage device. The first storage device (super-capacitor) stores and accumulates charge thereon from the rectified (DC) signal outputted by the rectifier circuit.
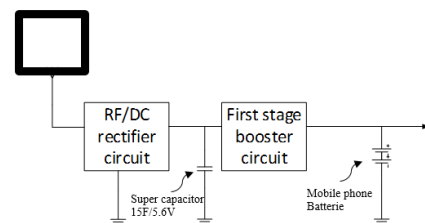
Figure 12: Block diagram for Wi-Fi acquiring energy

4.1 Matching circuit using Advanced design system (ADS)

The maximum microwave power distribution from proposed square loop antenna to the full wave rectifier circuit is ensured by the matching circuit. We are sure that the VSWR obtained from the simulation on CST is 1.073. From the maximum microwave source, the transfer occurs when the circuit is matched with the loop antenna, the impedance matching is mostly performed at the particular input signal [19]. The matching circuit design performs impedance transformation to guarantee maximum power between the antenna and RF-DC converter circuit. A matching circuit as show in Figure 13 that operates at 2.450 GHz the input impedance of 50Ω and the load resistance of 50Ω were made.
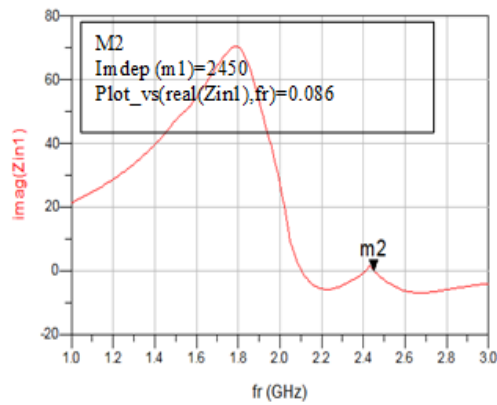
Figure 13: Input impedance versus frequency

4.2 A full bridge rectifier rectenna circuit

After designing a matching circuit, all components are optimized by setting two goals at the same time ; minimizing return loss between 1.4 GHz and 3.6 GHz and maximizing the DC output voltage. To design the full wave bridge rectifier using HSMS 2820 diode for microwave energy circuit, the following components in figure 14 were used [46]. The output voltages as a function of time for an input power of 2.450 GHz and the simulated value. The P_{in} versus load are presented in figure 15 and figure 16, respectively. In addition, figure 17 shows the output current (A) in the load circuit. The maximum output voltage shows 0.0025A at 6 dBm.
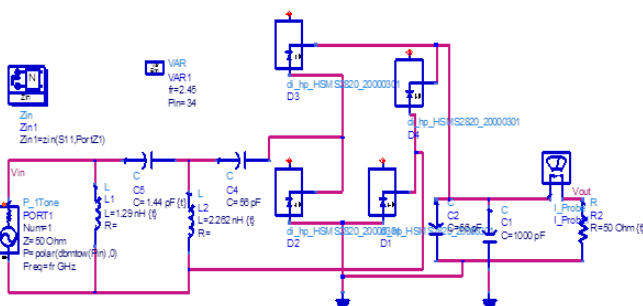
Figure 14: Schematic diagram of microwave harvesting circuit of full wave bridge rectifier with HSMS 2820 diode (Agilent ADS)
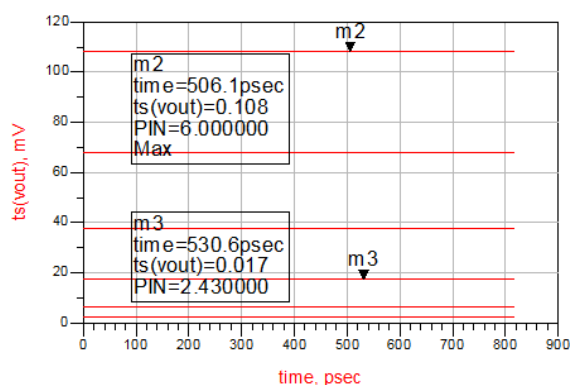
Figure 15: Simulation of the output voltages

The simulation shows the output voltage (V) of the circuit using the full wave bridge rectifier circuit. It can be observed that using an HSMS 2820 diode enables to generate a power around 102 mW, as shown in figure 15.
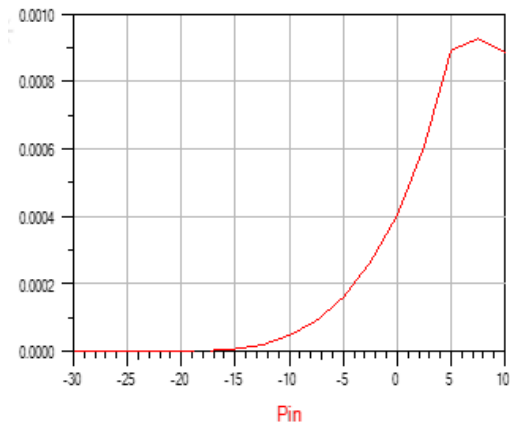
Figure 16: The simulation output power (w) of one stage circuit

The Figure 17 shows the output current of the circuit using the full wave bridge rectifier circuit. It can be observed that using an HSMS 2820 diode enables to generate a DC current around 0.0025 A.
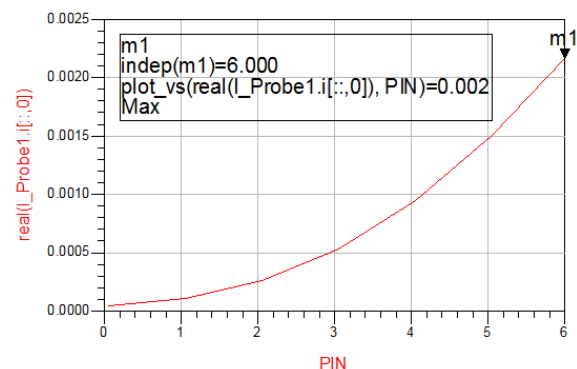
Figure 17: The simulation output current (I) of the full wave bridge rectifier circuit
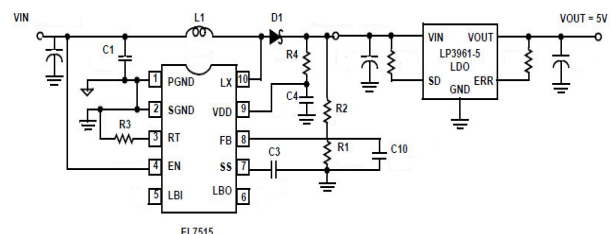
Figure 18: A voltage booster circuit

4.3 A voltage booster circuit

The voltage on the super capacitor (15F/5.6V) is also provided to the input of DC-to-DC converter circuit,
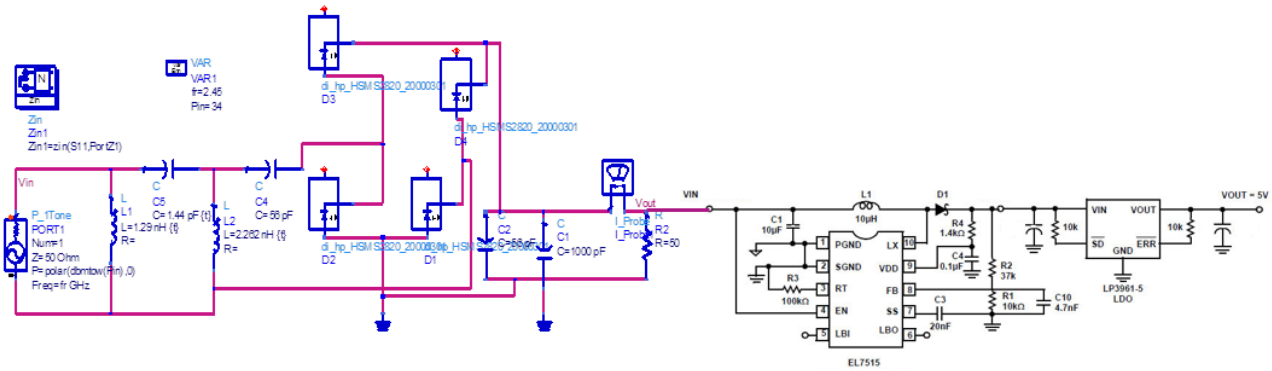
Figure 19: RF harvesting electronic circuit and power changing

which acts as a charge pump. Polarized energy storage capacitor with high capacity. A voltage booster integrated circuit was used to try to boost the voltage outputted by the RF/DC circuit to 5 V, can charge the battery of mobile phone. This integrated circuit can be used to boost a low-voltage variable input (1.8 V - 9 V) to a constant 5 V can charge the battery of mobile phone. The integrated circuit LP3961EMP-5.0 it is linear regulators circuit can provide +5V/ 0.8A. The circuit is shown in Figure 18.

5 Simulation results

More specifically, the antenna is coupled to the rectifier circuit, which converts the Wi-Fi radiating source into a DC voltage multiplied. The energy acquiring device of the present work is capable of receiving microwave power from 1.440 GHz to 3.6 GHz, and convert the received signals into electrical energy, as illustrated in Figure 19.

purpose and implementation of the energy acquiring method and apparatus of the present research. The results obtained in the different simulation are remembered in Table 2.

Table 2: Simulated results

Performance parameters	Value
Maximum gain value (dBi)	4.013
Minimum gain value (dBi)	2.711
Bandwidth	40 %
VSWR	1.073

This work present propose method assembly of PCB circuit, the assembly method and the topology to put the PCB circuit inside the mobile phone. The circuit are presented in Figure 20 and Figure 21, respectively.

Figure 21: PCB circuit
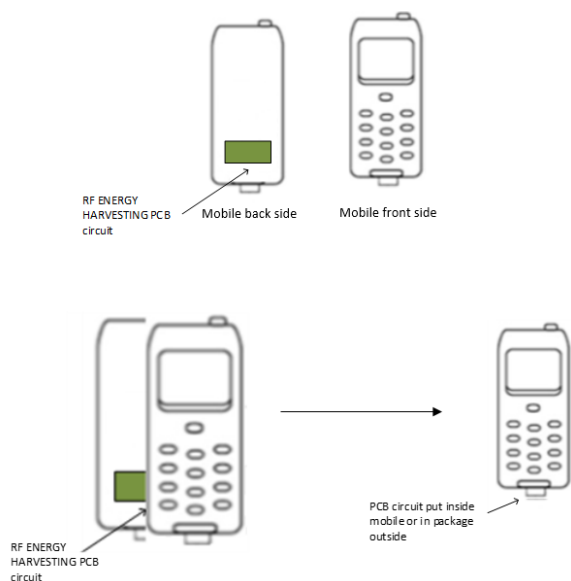
Figure 20: Method assembly of PCB circuit

Provided above is a general description of the
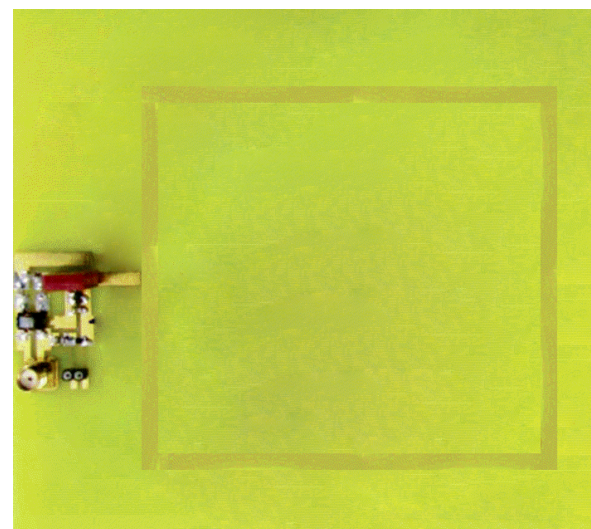
6 Conclusion and Outlook

In this research, a new method for wireless charging the mobile phone by acquiring Wi-Fi energy. The idea is to acquire a maximum energy while other is charging

or using their electronic devices. The present research paper we are tested and proven that we can generate voltage from Wi-Fi access radio frequency to power and charge a battery of any electronic devices. Future goals of this work is to acquire energy from any signal around us like Wi-Fi, GSM, VHF, UHF, etc. We can change the topology of rectifier circuit and the design of loop antenna and check if we can acquire more energy or not. The most important that we have one prototype for this work and try to improvement in the future work.

References

[1] John Ross, A Painless Guide to Wi-Fi and Broadband Wireless 2nd Edition, 2008.

[2] Steve Pease, WI - FI: HOW TO BOOST YOUR WI - FI SIGNAL: Get the wi - fi and internet access you need all over your house Kindle Edition, 2016.

[3] Icon Group International , The 2019-2024 World Outlook for Gigabit Wi-Fi Access Points, 2018.

[4] Icon Group International ,The 2019-2024 World Outlook for Carrier Wi-Fi Access Points and Controllers, 2018.

[5] Alliance for Telecommunication Industry Solutions , ATIS ATIS-0600015.13.2017 Energy Efficiency for Telecommunication Equipment: Methodology for Measurement and Reporting of 802.11xx Wi-Fi Access Points, 2017.

[6] Joseph Epstein, Scalable VoIP Mobility: Integration and Deployment 1st Edition, Elsevier, 2009.

[7] M. Hadi Habaebi, N. Izzati Nabilah Bt Azizan, "Harvesting WiFi Received Signal Strength Indicator (RSSI) for Control/Automation System in SOHO Indoor Environment with ESP8266", 2016 International Conference on Computer and Communication Engineering (ICCCE), 416 - 421, 2016.

[8] Xuelin Chen, Lianfen Huang, Jianli Xing, Zhiyuan Shi, Zuosheng Xie,"Energy harvesting system and circuits for ambient WiFi energy harvesting",12th International Conference on Computer Science and Education (ICCSE), 769 - 772, 2017.

[9] Jingjing Wang, Chunxiao Jiang, Zhu Han, Yong Ren, Lajos Hanzo," Network Association Strategies for an Energy Harvesting Aided Super-WiFi Network Relying on Measured Solar Activity",IEEE Journal on Selected Areas in Communications, 3785 - 3797, 2016.

[10] Lidiya Mary Lawrence, Jobin K Antony, "Design and performance analysis of RF to DC converter for wireless sensors" , 2nd IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT), 1660 - 1665, 2017.

[11] A. Almansouri, M. Ouda, K. Salama,"A CMOS RF-to-DC Power Converter With 86% Efficiency and -19.2-dBm Sensitivity"IEEE Transactions on Microwave Theory and Techniques, 1-7, 2018.

[12] R. Correia, N. Carvalho,"HEMT based RF to DC converter efficiency enhancement using special designed waveforms", 2017 IEEE MTT-S International Microwave Symposium (IMS), 609-612, 2017.

[13] Sanjeev K, Manjunath Machnoor, K. J. Vinoy, T. V. Prabhakar,"A high efficiency 2.4GHz RF to DC converter using 130nm CMOS Cross-Coupled Rectifier",2016 Twenty Second National Conference on Communication (NCC), 1-4, 2016.

[14] H. Ouda, W.khalil, N. Salama,"Wide-Range Adaptive RF-to-DC Power Converter for UHF RFIDs",IEEE Microwave and Wireless Components Letters, 634-636, 2016.

[15] CHAN CHUN YEW,"DESIGN OF RECTIFYING CIRCUIT WITH IMPROVED RF-DC CONVERSION FOR WIRELESS POWER TRANSFER",Ph.D Thesis, UNIVERSITI TEKNIKAL MALAYSIA MELAKA , 2015.

[16] Zaln Horvth, Ferenc Csikor,"Personal Electromagnetic Exposure Measurements in the Radiofrequency and Microwave Range",Ph.D Thesis, Etvs Lornd University, 2012.

[17] Finta V, Vradi L, Juhsz P, Kiss , Thurczy Gy," A Method for Personal RF Exposure Assessment of Children", EMF Health Risk Research: Lessons Learned and Recommendations for the Future, workshop poster, 2012.

[18] G. F. Pedersen and J. B. Andersen, "RF and ELF exposure from cellular phone handsets: TDMA and CDMA systems," Radiation Protection Dosimetry, 131-138, 1999.

[19] S. Iskra, B. W. Thomas, R. McKenzie, and J. Rowley, "Evaluation of potential GPRS 900/1800-MHz and WCDMA 1900-MHz interference to consumer electronics," Ieee Transactions on Electromagnetic Compatibility, 951-962, 2005.

[20] G. Wang, L. Zhang and J. Zhang, "A review of electrode materials for electrochemical supercapacitors", Chem. Soc, 797-828, 2012.

[21] H. Wu, G. Yu, L. Pan, N. Liu, M. T. McDowell, Z. Bao and Y. Cui, "Stable Li-ion battery anodes by in-situ polymerization of conducting hydrogel to conformally coat silicon nanoparticles", Nature Communications, 2013.

[22] M. S. Halper, J. C. Ellenbogen, Supercapacitors: A Brief Overview, MITRE, 2006.

[23] C. Liu, Z. Yu, D. Neff , A. Zhamu, and B. Z. Jang, "Graphene-based supercapacitor with an ultrahigh energy density", Nano Lett., 4863-4868, 2010.

[24] B. E. Conway and W. G. Pell, Double-layer and pseudocapacitance types of electrochemical capacitors and their applications to the development of hybrid devices, J.Solid State Electrochem., 637644, 2003.

[25] C. Mid-Eum and S. Seung-Woo, Robust energy management of a battery/supercapacitor Hybrid Energy Storage System in an electric vehicle, presented at the Electric Vehicle Conference (IEVC), 2012 IEEE International, 15,2012.

[26] Muhammad S Khan, Hai Deng,"Design and implementation of a highly efficient UHF energy harvesting antenna", IEEE International Symposium on Antennas and Propagation (APSURSI),611-612, 2016.

[27] Hirokazu Kamoda, Shoichi Kitazawa, Naoya Kukutsu, Kiyoshi Kobayashi,"Loop Antenna Over Artificial Magnetic Conductor Surface and Its Application to Dual-Band RF Energy Harvesting",IEEE Transactions on Antennas and Propagation, 4408 - 4417, 2015.

[28] Youssef Tawk, Joseph Costantine, Firas Ayoub, Christos G. Christodoulou,"A Communicating Antenna Array with a Dual-Energy Harvesting Functionality",IEEE Antennas and Propagation Magazine, 132-144, 2018.

[29] Mamta Kurvey; Ashwini Kunte,"Design and optimization of stepped rectangular antenna for RF energy harvesting", International Conference on Communication information and Computing Technology (ICCICT),1-4, 2018.

[30] Jingna Mao, Jian Zhao, Huazhong Yang, Bo Zhao,"Using human body as a monopole antenna for energy harvesting from ambient electromagnetic energy", IEEE Biomedical Circuits and Systems Conference (BioCAS),1-4, 2017.

[31] S. Hayashida, T. Tanaka, H. Morishita, Y. Koyanagi, and K. Fujimoto, Built-in folded monopole antenna for handsets, Electron. Lett, pp. 15141516, 2004.

[32] B. Jung, H. Rhyu, Y. J. Lee, F. J. Harackiewicz, M. J. Park, and B.Lee, Internal folded loop antenna with tuning notches for GSM/GPS/DCS/PCS mobile handset applications, Microw. Opt. Technol. Lett.,vol. 48, pp. 15011504, 2006.

[33] C. Y. D. Sim, P.-C. Cheng, and C.-H. Lee, Multiband loop antenna design for mobile devices, Microw. Opt. Technol. Lett., vol. 51, pp. 22422248, 2009.

[34] B.-P. Su, M.-K. Sung, and G.-Y. Woon, Broadband internal antenna by combination of a loop type antenna and a shorted monopole, Microw. Opt. Technol. Lett., vol. 50, pp.28102812, 2008.

[35] Chia-Hao Ku, Hsien-Wen Liu, Sheng-Yu Lin, Folded Dual-Loop Antenna for GSM/DCS/PCS/UMTS Mobile Handset Applications, IEEE Antennas and Wireless Propagation Lett., vol. 9, pp. 998-1001, Oct. 2010.

[36] H. Jairam, "Novel blade antenna using microstrip patch elements," in Microwave Conf., 19th European, London, UK, 1989, pp. 1118-1122.

[37] Z. Novacek, "Radiation of a whip antenna on the car body," Radioelectronika, 17th Int. Conf., Brno, June 2007, 1-4.

[38] A. Walbeoff and R. Langley, "Multiband PCB antenna," Proceedings of IEE Microwaves, Antennas and Propag., Centre, Sittingbourne, UK, Dec. 2005, 471-475

[39] F. Mariottini, M. Albani, E. Toniolo, D. Amatori, and S. Maci, "Design of a compact GPS and SDARS integrated antenna for automotive applications," Antennas and Wireless Propag. Lett., IEEE, 405-408, 2010.

[40] S. Choi, H. Lee, and K. Kwak, Circularly Polarized H-Shaped Microstrip-Array Antenna with a T-Slot for DSRC System Roadside Equipment, Microwave and Optical Technology Letters, pp. 1545-1548, 2009.

[41] G. Whyte, Antennas for Wireless Sensor Network Applications, Ph. D. Thesis, University of Glasgow, Glasgow, United Kingdom, 2008.

[42] C. Balanis, Antenna Theory: Analysis and Design, 3rd ed, Wiley, Hoboken, New Jersey, 2005.

[43] A. Heidari, M. Heyrani, M. Nakhkash, A Dual-Band Circularly Polarized Stub Loaded Microstrip Patch Antenna for GPS Applications, Progress in Electromagnetics Research, PIER 92,195-208, 2009.

[44] Luis Filipe Ribeiro Dias, Alrio Boaventua, Nuno Borges de Carvalho,"RF-DC converter efficiency optimization using source-pull techniques", International Workshop on Integrated Nonlinear Microwave and Millimetre-wave Circuits (INMMiC), 1-3, 2014.

[45] M. Roscia, D. Zaninelli, G. C. Lazaroiu; H. Shadmehr,"Using design optimization in radio frequency energy harvesting systems", International Conference on Clean Electrical Power (ICCEP), 740-744, 2013.

[46] Faruk Erkmen; Thamer S. Almoneef; Omar M. Ramahi," Electromagnetic Energy Harvesting Using Full-Wave Rectification", IEEE Transactions on Microwave Theory and Techniques, 1843-1851, 2017.

On The Development of a Reliable Gate Stack for Future Technology Nodes Based on III-V Materials

Abhitosh Vais[1,*], Sonja Sioncke[1], Jacopo Franco[1], Vamsi Putcha[1,2], Laura Nyns[1], Arturo Sibaja-Hernandez[1], Rita Rooyackers[1], Sergio Calderon Ardila[2], Valentina Spampinato[1], Alexis Franquet[1], Jan Maes[3], Qi Xie[3], Michael Givens[4], Fu Tang[4], Xiang Jiang[4], Marc Heyns[1,2], Dimitri Linten[1], Jerome Mitard[1], Aaron Thean[5], Dan Mocuta[1] and Nadine Collaert[1]

[1]*IMEC, Kapeldreef 75, 3001 Leuven, Belgium*

[2]*ESAT department, KU Leuven, 3000 Leuven, Belgium*

[3]*ASM Belgium, Kapeldreef 75, 3001 Leuven, Belgium*

[4]*ASM America, Phoenix, AZ 85034, USA*

[5]*Department of Electrical and Computer Engineering, NUS, 117546 Singapore*

A R T I C L E   I N F O

A B S T R A C T

In this work, we discuss how the insertion of a $LaSiO_x$ layer in between an in-house IL passivation layer and the high-k has moved the III-V gate stack into the target window for future technology nodes. The insertion of this $LaSiO_x$ layer in the gate stack has reduced the D_{it} and N_{bt} below the target level of $5x10^{11}$ /eV.cm² and $3x10^{10}$ /cm² (target at 10 years operation: $\Delta V_{fb}<30mV$ at 125°C) respectively. From physical analysis, it was found that $LaSiO_x$ can stabilize the interaction of the IL layer with the InGaAs substrate. An implant free $In_{0.53}Ga_{0.47}As$ n-MOSFET was fabricated with this gate stack and for the first time, a III-V gate stack meets the reliability target for advanced technology nodes with a max operating V_{ov} of 0.6 V. In addition, an excellent electron mobility ($\mu_{eff,peak}=3531$ cm²/V-s), low $SS_{lin}=71$ mV/dec and an EOT of 1.15 nm were obtained. We also report the scaling potential of this stack to 1 nm EOT without any loss in performance, reliability and further reduction of the sub-threshold swing ($SS_{lin}=68$ mV/dec).

1. Introduction

III-V compound semiconductors have already been widely investigated as an alternative material to Silicon for future Metal - Oxide-Semiconductor Field Effect Transistors (MOS-FETs). MOS devices based on III-V compound semiconductor materials like InGaAs, with excellent performance, have been demonstrated on both 2-inch InP substrates and on a 300mm Si platform [1-5]. But, in order to become a commercially viable alternative to existing Si technologies, two major issues are yet to be solved: 1) the defective interface between III-V materials and the high-k oxide layers, and 2) the high density of defects in the most commonly investigated high-k oxides, Al_2O_3 and HfO_2.

Both of these issues have been shown to affect almost every aspect of device performance, such as the capacitance [6-11], reliability of III-V based devices [11-14], on-state electrical parameters like mobility [15], and trans-conductance [16, 17]. As a result, it becomes extremely important to reduce their impact in order to make commercial III-V MOSFET technology a reality. In this regard, the use of a new IL layer in combination with an Al_2O_3/HfO_2 or just HfO_2 stack was reported, showing improved performance and a significant reduction of the oxide traps. However, with the requirement of a thin Al_2O_3 layer for gate stack thermal stability and due to the wide defect distribution in Al_2O_3, the total density of charging oxide traps was still higher than the target (10 years reliability target: flat-band voltage shift, $\Delta V_{fb} < 30$ mV at 125°C. Assuming a typical time power law exponent of 0.13, the failure criterion projects effective density of charged defects, ΔN_{eff} to $< 3x10^{10}/cm^2$ at operating field, for an Effective Oxide Thickness (EOT) of 1 nm [18, 19]).

*Abhitosh Vais, Email: Abhitosh.Vais@imec.be

This paper is an extension of the work presented in *IEEE Symposium on VLSI Technology 2017*. In this paper, we demonstrate and comprehensively discuss on the process of replacing Al_2O_3 in the $IL/Al_2O_3/HfO_2$ gate stack with $LaSiO_x$. Furthermore, we elaborate on the improvements, thereof, in the interface properties and reliability for operation at $V_{DD}=0.8$ V, which exceeds the operating target for future technology nodes without any loss in the electrical performance of the MOSFET device reported earlier.

2. Experimental details

Capacitors were made starting from an MBE grown 300 nm n-type $In_{0.53}Ga_{0.47}As$ on 2-inch n-type InP substrates. The targeted Si doping in the $In_{0.53}Ga_{0.47}As$ grown layer was ~$1x10^{17}$ at/cm^3. Prior to the gate stack deposition, the substrates were cleaned in a 2M HCl solution for 5 min at room temperature and subsequently rinsed in de-ionized water. H_2S pretreatment and deposition of the IL (Inter Layer), $LaSiO_x$, Al_2O_3 and HfO_2 were done in an ASM Pulsar® 3000 ALD reactor. The H_2S pretreatment is done in-situ prior to the high-κ deposition. The IL used here is an ALD based layer with a κ-value ~6. The H_2S treatment and IL film deposition are both done at 250°C. Al_2O_3 and HfO_2 were deposited by ALD TMA/H_2O and $HfCl_4/H_2O$ at 300°C. $LaSiO_x$ is deposited by an ALD process at 250°C using the following precursors: LoLaPrime® (AIR Liquide), $SiCl_4$, H_2O. A 42% La concentration in the $LaSiO_x$ is obtained by using a La:Si pulse ratio of 1:5. A full ALD cycle consists of : [[SiCl$_4$/purge/H2O/purge] x 5 + [LolaPrime/purge/H2O/purge] x 1] x n. Under these conditions, a GPC of 0.1754 nm is achieved.

Dot capacitors were fabricated by sputtering 80 nm TiN using physical vapor deposition (PVD). A dry-etch process was used to etch the metal gate. An anneal in forming gas (10% H_2/N_2) was done prior to metal deposition (PDA) or post metal deposition (PMA) at 400°C for 5 min.

Implant-free quantum-well devices were also fabricated to study the impact of these passivation schemes on the device performance and reliability. A schematic diagram of the quantum-well MOSFET is shown in Fig.1. The stack for MOSFETs consists of a 15 nm un-intentionally doped $In_{0.53}Ga_{0.47}As$ channel layer, a 3 nm InP etch stop layer and a 50 nm n+ $In_{0.53}Ga_{0.47}As$ (Si-doped, $1x10^{19}$ cm^{-3}) layer, all grown on a 2-inch semi-insulating InP substrate. Fig. 2 describes the process flow used in this work. The detailed device fabrication can be found in [20]. Prior to gate stack deposition, a digital etch (cycles of HCl and H_2O_2) was used to clean and smoothen the InGaAs channel surface. The different gate stacks that were studied are depicted in Table I.

I-V and C-V measurements were used to study the impact of the gate stack on the performance of the devices. I-V measurements were carried out using HP4156 parameter analyzer and C-V measurements were performed using a HP4284A-LCR meter. The interface defect density, D_{it}, is extracted using the conductance method from the parallel equivalent conductance G_p [21] at room temperature. The amount of oxide traps (ΔN_{eff}) are measured by C-V hysteresis measurements (forward and reverse sweep of the gate voltage) at different maximum stress voltages and the corresponding shift in flat band voltage, ΔV_{fb}, measured between the forward and backward trace is converted into a sheet charge (ΔN_{eff} (E_{ox})). The flat-band voltage is the gate voltage

corresponding to an inflection point in the double derivative of the normalized capacitance-voltage data at a frequency of 100 kHz [21,22]. The ΔN_{eff} at target field as well as the voltage acceleration
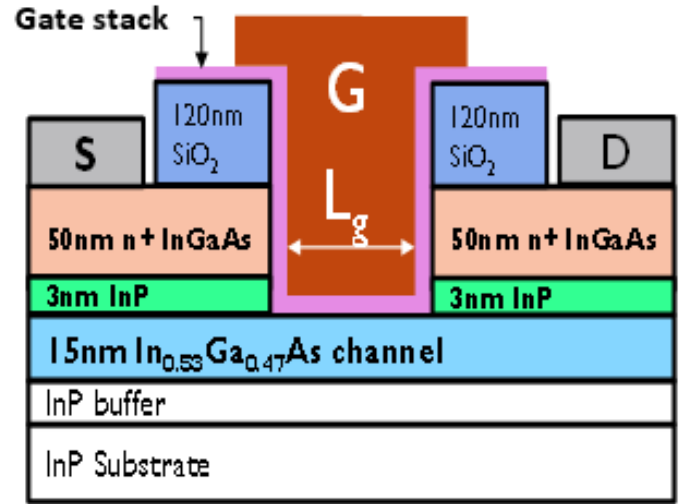
Figure. 1. Schematic representation of the MOSFET device (Dimensions are not to scale).

Mesa isolation by wet etching
SiO_2 hard mask deposition
Gate opening in SiO_2
Gate recess by wet etching
Surface preparation using optimized digital etch process
Gate stack deposition
PVD TiN deposition and gate patterning by RIE
PVD Mo/Al S/D contact by lift-off
Post Metal anneal (PMA)
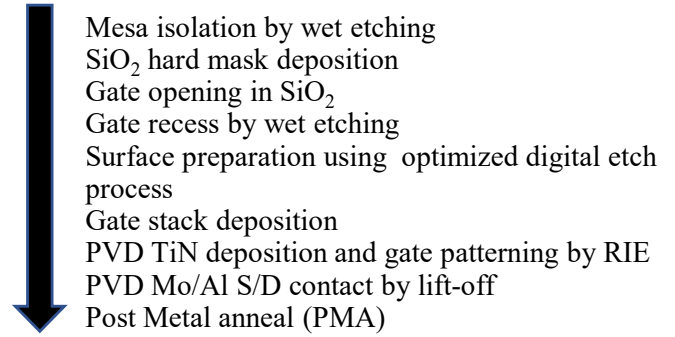
Figure 2. Process flow used in this work for MOSFET fabrication.

Table 1: Split table of III-V gate stacks

Sample ID	Dielectric
Ref	3 nm HfO_2
A	1 nm IL/3 nm HfO_2
A'	0.5 nm IL/3 nm HfO_2
B	1 nm IL/1 nm Al_2O_3/3 nm HfO_2
B'	1 nm IL/0.2 nm Al_2O_3/3 nm HfO_2
C	1 nm IL/1 nm $LaSiO_x$/3 nm HfO_2
C'	1 nm $LaSiO_x$/3 nm HfO_2

factor (γ, i.e. the slope of the log-log plot of ΔN_{eff} vs E_{ox}) is reported as a figure of merit to study the oxide trap behavior [23]. Although this method does not take into account the influence of stress temperature and stress time into account, it helps in identifying the trends between different gate stack/pretreatments. Bias temperature instability (BTI) characterization using extended Measure-Stress-Measure (eMSM) technique was used to evaluate the reliability of the MOSFET devices [23]. Finally, Time-Of-Flight Secondary Ion Mass Spectrometry (TOFSIMS) was used to study the chemistry of the stacks.
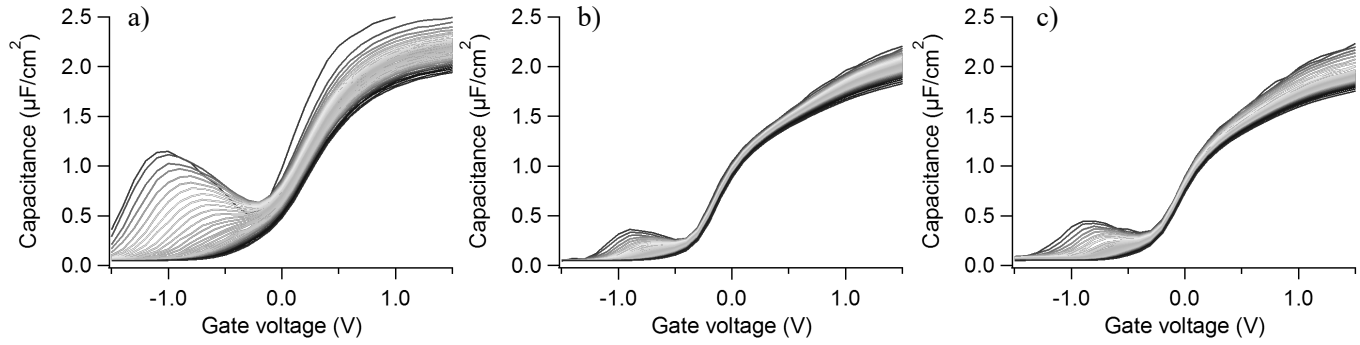
Figure 3. Measured RT CV's of gate stacks a) 'ref', b) A and c) B' described in Table I at 31 frequencies between 1kHz and 1MHz. The MOSCAP devices received a PDA at 400°C, 5' in Forming gas.

3. Results and discussion

3.1. The IL layer development and thermal stability

Fig. 3 shows the multi-frequency (1 kHz - 1 MHz) C-V for various samples in the split table of Table I, measured at room temperature. The reduction in the D_{it} bump and in frequency dispersion when inserting the IL (3-A,B') between the InGaAs channel and the HfO_2 layer clearly shows the improvement. Fig. 4 shows the result of the extraction of various electrical parameters from the measured C-V data. The extracted mid-gap D_{it} drops by more than a factor of 2 when inserting the IL accompanied with a negative flat band voltage, V_{fb}, shift (Fig. 4a,b). An H_2S treatment prior to the HfO_2 deposition has a similar effect: it reduces the D_{it}

and shifts the V_{fb} by -0.15 V. Compared to an H_2S treatment, the insertion of 1 nm of the IL layer shifts the V_{fb} to even more negative values and a larger reduction in D_{it} is observed. In addition, the IL layer reduces the ΔN_{eff} by 1 order of magnitude and increases the γ-value to 2.5. Although 0.5 nm of the IL layer shows similar trends, an optimum is reached when using 1nm of the IL layer. This result suggests that 1 nm of IL is needed to form a closed layer passivating the interface. When adding a thin (0.2 nm) Al_2O_3 cap on top of the IL layer, the D_{it} bump and frequency dispersion increase slightly (Fig. 3-B'). One can also observe that the V_{fb} shifts to more positive values and this goes hand in hand with a slight increase in ΔN_{eff} and reduction of the γ value. From these observations, a hypothesis can be put forward. Fig. 5 depicts
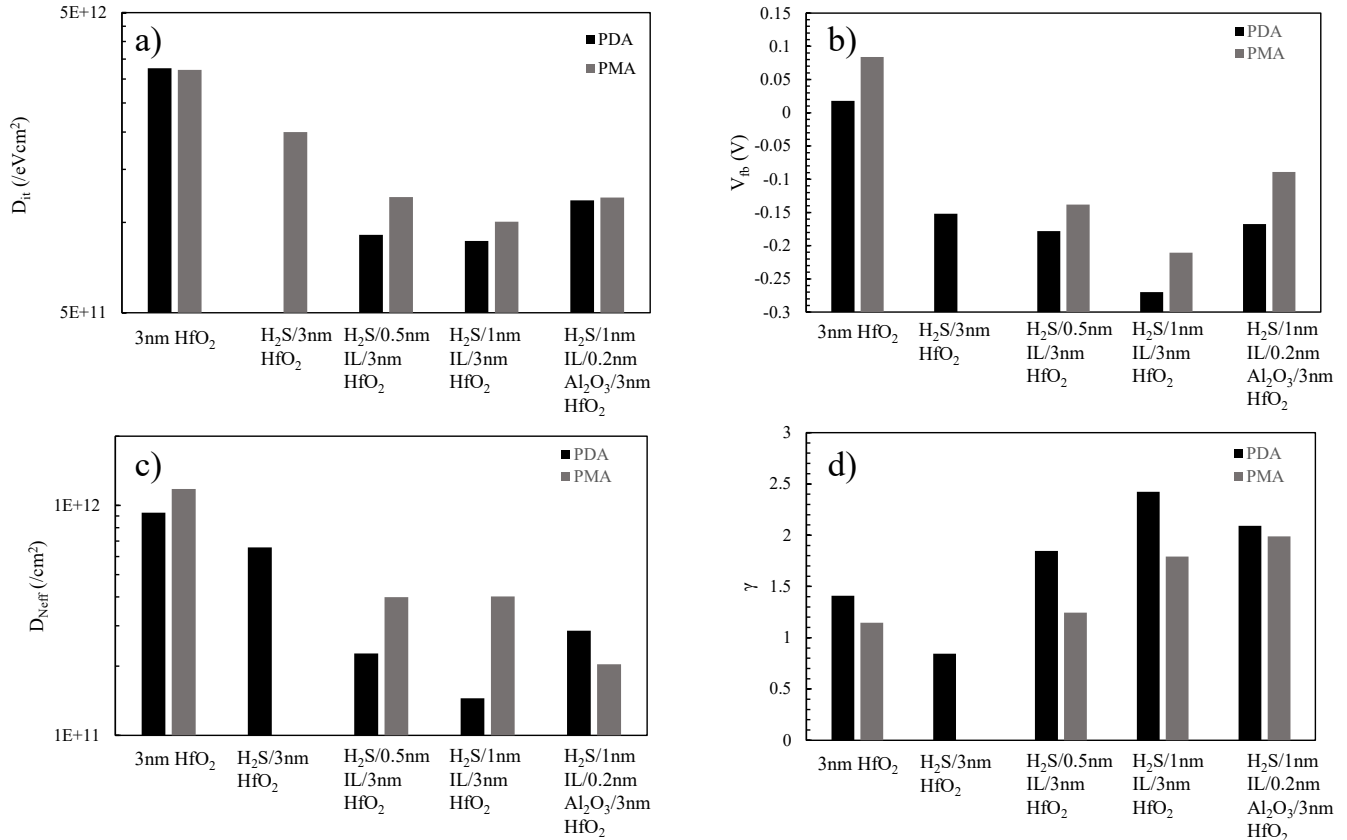
Figure 4. Electrical parameters extracted for gate stacks 'ref', A, A', B' described in Table I. a) Midgap D_{it} extracted from RT CV's. b) V_{fb} c) ΔN_{eff} extracted from RT CV hysteresis by converting the measured ΔV_{fb} into an effective charge sheet at the interface by: $\Delta N_{eff} = \Delta V_{fb} * C_{ox}/q$, and reported at charging equivalent field of 3.5 MV/cm [13]. V_{start} condition was taken at $V_g = -1.5V$. d) Field acceleration exponent, γ, i.e. the slope of the log-log plot ΔN_{eff} vs oxide field which is a measure of the misalignment between the channel carrier energy-i.e. InGaAs E_f- and defect levels in the high-*k*. Filled bars represent the data after PDA, shaded bars represent the data after PMA.
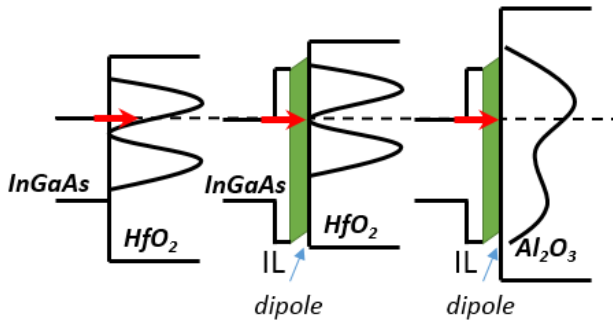
Figure 5. Defect band model to explain the observed changes in the γ value and total amount of charged oxide traps at a given field. The IL passivation layer is creating a dipole shifting the misalignment between InGaAs E_f and defect levels in the high-κ. In Al_2O_3 a partial overlap between deep and shallow traps is existing [9] and an almost uniform defect band is present. Therefore, the dipole created is not as effective as compared to the situation where deep and shallow traps are well separated as is proposed for the stack with HfO_2.

a cartoon of a possible explanation. Defect band modeling [24,25] has shown that although peak defect densities in Al_2O_3 are lower, a partial overlap of the defect bands induces an almost uniform distribution of defect levels around the InGaAs conduction band edge, E_c. On the contrary, HfO_2 on InGaAs shows a minimum defect density slightly below the channel E_c (~0.2 eV below E_c), which can be exploited for improving gate stack reliability by introducing an interface dipole. The assumption is that the IL layer gives rise to such a dipole shifting the defect levels up with respect to the InGaAs E_c and shifting the minimum level into the device operating range. However, dipole engineering is not effective when a broad defect band without a clear minimum is present as is the case with Al_2O_3.

From the results shown till now, one can conclude that the 1 nm IL/3 nm HfO_2 stack shows an optimum for all electrical parameters although the D_{it} and N_{bt} levels are still above the targeted values (5 x 10^{11} /eV.cm^2 and 3 x 10^{10} /cm^2 respectively). However, when applying a PMA, we observe that the 1 nm IL/3 nm HfO_2 stacks degrades slightly i.e. a D_{it}, ΔN_{eff} increase and a reduction in the gamma value is observed in combination with a positive V_{fb} shift (Fig. 4-shaded bars). On the contrary, the stack with the Al_2O_3 cap does not degrade when a PMA is added. This result implies that although the initial gate stack properties are degraded when adding an Al_2O_3 cap, this cap can help in stabilizing the gate stack upon a PMA.

In conclusion, the IL approach provides an improved interface quality and increases the misalignment between the InGaAs channel E_f and the defect bands in the high-κ. The improved interface quality was proven to increase the performance of

InGaAs MOS devices on both planar as well as nanowire architectures. However, the introduction of a cap layer is needed to provide a good thermal stability. Although Al_2O_3 can provide a higher stability for thermal processing, the unfavorable defect band distribution in the Al_2O_3 does not allow for dipole engineering. In the next part, we show that a $LaSiO_x$ cap can provide both an improved thermal stability and a more favorable defect band distribution.

3.2. Replacement of Al_2O_3 stabilization cap by $LaSiO_x$

Fig. 6 shows the multi-frequency C-V for various gate stacks in Table I and demonstrates the impact of the insertion of $LaSiO_x$ between the IL passivation and HfO_2. All the dot capacitors were measured after PMA at 400°C in forming gas anneal for 5 minutes. Although the CV's from the 1 nm IL/3 nm HfO_2 stack (Fig. 6A) and the 1 nm IL/1 nm $LaSiO_x$/3 nm HfO_2 stack (Fig. 6C) look similar, one can notice that the frequency dispersion is further reduced when $LaSiO_x$ is inserted between the IL and HfO_2. As was discussed in the previous section, an Al_2O_3 cap degrades the gate stack properties slightly. Although the difference between the stacks with the IL are subtle, a large difference can be observed when the IL is removed (Fig. 6C'). These results imply that both the IL and the $LaSiO_x$ cap are needed to optimize the gate stack properties.

The D_{it}, extracted at mid gap, and capacitance equivalent thickness (CET) are shown in fig. 7. The insertion of the IL (A) reduces the mid gap D_{it} as was already reported in the previous section. The replacement of Al_2O_3 (B) by $LaSiO_x$ (C) further reduces the D_{it} and a reduction of the CET is also observed (CET=1.55 nm, EOT=1.15 nm). In addition, the total density of charging oxide traps drops to a value close to 1 x 10^{10} /cm^2 at E_{ox}
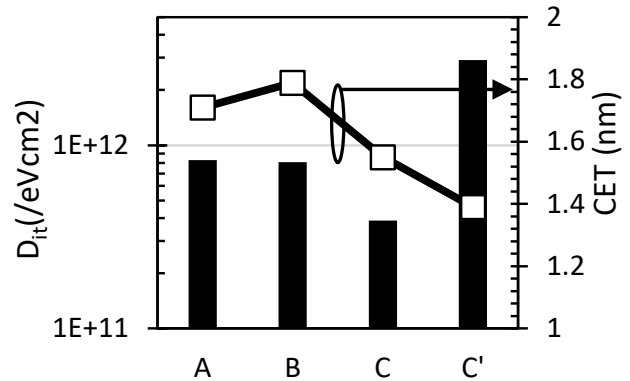
Figure 7: Mid gap D_{it} (bars) and CET (dots) extracted from the C-V data presented in fig. 6. Replacement of Al_2O_3 (B) by $LaSiO_x$ (C) reduces the D_{it} to less than 4 x 10^{11}/cm^{2}eV. The CET drops to 1.55 nm.
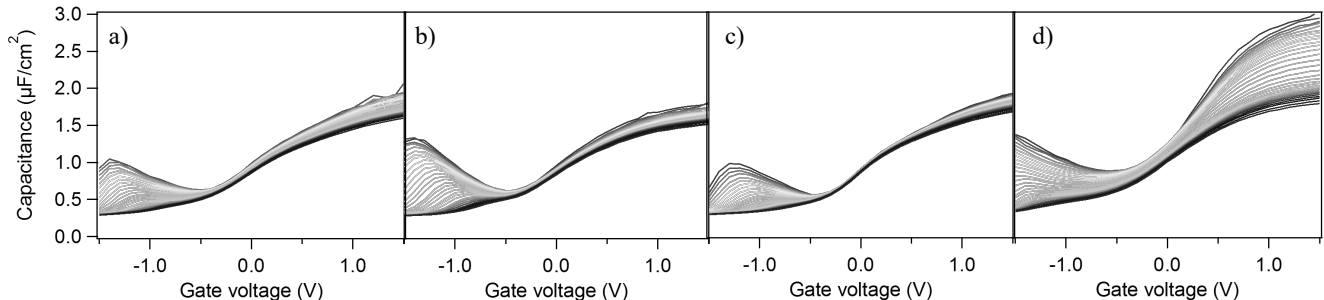
Figure 6. RT MF-CV's measured from dot capacitors of A) 1 nm IL/3 nm HfO_2, B) 1 nm IL/1 nm Al_2O_3/3 nm HfO_2, C) 1 nm IL/1 nm $LaSiO_x$/3 nm HfO_2, C') 1 nm $LaSiO_x$/3 nm HfO_2. Frequencies between 1 kHz-1 MHz (red-blue) were measured with logarithmic spacing. All capacitors received a PMA in forming gas for 5 minutes.
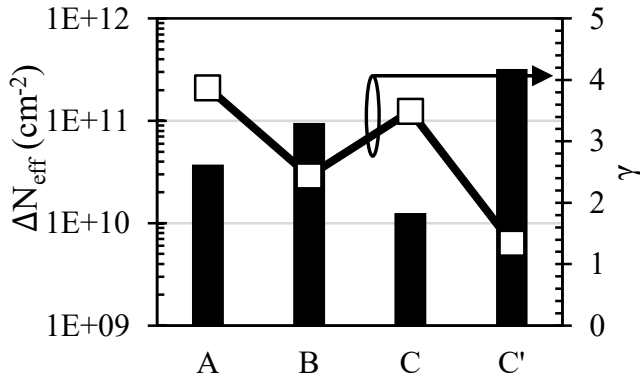
Figure 8: ΔN_{eff} (bars, left axis) extracted from C-V hysteresis ($V_{start}=V_{fb}$ - 0.5 V) by converting the measured ΔV_{fb} into an effective charge sheet at the interface by: $\Delta N_{eff}=\Delta V_{fb}*C_{ox}/q$, and reported at charging equivalent field of 3.5 MV/cm [6]. Field acceleration exponent, γ, (dots, right axis) i.e. the slope of the log-log plot ΔN_{eff} vs oxide field which is a measure of the misalignment between channel carrier energy-i.e. InGaAs E_f and defect levels in the high-*k* [18].

= 3.5 MV/cm (Fig. 8) *meeting the target for 10 years operation* (10 years reliability criterion: ΔV_{fb} < 30 mV at 125°C. Assuming a typical time power law exponent of 0.13, the failure criterion projects ΔN_{eff} to less than 3 x 10^{10} /cm² at operating field, for an EOT of 1 nm). The reduction of oxide traps below the target level is the result of the combination of the IL layer and LaSiO$_x$ cap: while the IL layer increases the field dependence of the ΔN_{eff} ($\gamma \sim$ 3.5), which is beneficial only for low voltage operation, the LaSiO$_x$ cap consistently reduces the density of charging defects irrespective of the operating voltage.

Fig. 9 shows the forward and backward sweep of gate voltage for the C-V hysteresis measured at different V_{max} for 1 nm IL/1 nm LaSiO$_x$/3 nm HfO$_2$ stack. At low overdrive voltages, the hysteresis is negligible. Only at higher overdrive voltages, a clear hysteresis can be measured. Notice also that the starting voltage for these measurements was carefully chosen at V_{fb} - 0.5 V. As was reported in [25], the C-V hysteresis measurements can largely depend on the V_{start} condition, as is also the case for gate stack under evaluation (Fig. 10). The more negative V_{start} condition reflects the discharging of deep traps that can take place during OFF-state, when the device is biased deep into the depletion and is an equally important criterion to take into account. Measuring at $V_{start}=V_{fb}$ - 0.5 V shows that oxide trap densities are within target. Starting from a more negative V_g, the oxide trap density increases and γ decreases. This observation indicates that this gate stack has a low

Figure 10: ΔN_{eff} as a function of applied E_{ox} for the stack with 1 nm IL/1 nm LaSiO$_x$/3 nm HfO$_2$ for 2 different V_{start} conditions. V_{start}=-1.5V and $V_{start}=V_{fb}$-0.5V. The large dependence on the V_{start} conditions points to the presence of deep traps i.e. located at energies below InGaAs E_f at flatband.

amount of electron traps, but a more negative V_{start} charges hole traps, thereby increasing the total amount of traps during the measurement and decreasing the voltage dependence.

Therefore, BTI measurements were performed at both positive and negative stress conditions to capture the max V_{ov} for both stress conditions (Fig. 11). From these CV-BTI measurements, one can see that the 1 nm IL/1 nm LaSiO$_x$/3 nm HfO$_2$ stack meets the

Figure 11: Max V_{ov} measured from positive and negative CV-BTI on MOS capacitors. For a target V_{DD}=0.8V, the max pos V_{ov} target (full line) is 0.53V and the max neg V_{ov} target (dashed line) is -0.26V (i.e. 1/3*V_{DD}).

Figure 9. 100 kHz CV up and down traces of the 1 nm IL/1 nm LaSiO$_x$/3 nm HfO$_2$ gate stack measured at 3 different V_{max}. V_{start} was V_{fb}-0.5V.
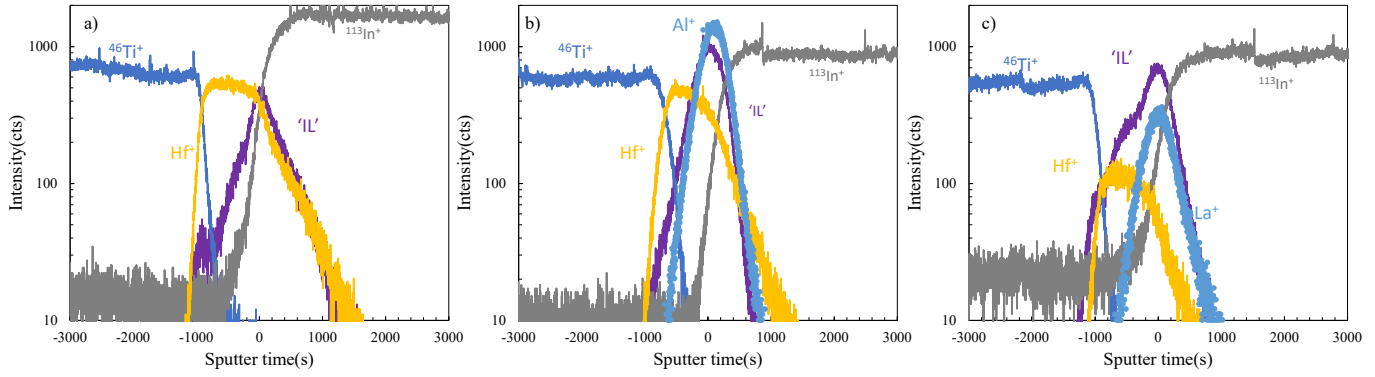
Figure 12. TOFSIMS positive ion profiles from MOSCAPs of stacks a) A, b) B and c) C from Table I. All layers are aligned at the peak of the IL atom profile.

max V_{ov} target (for a $V_{DD}=0.8$ V) for both stress conditions. When removing the IL layer, none of the targets are met. As III-V materials will be introduced in technologies operating at $V_{DD}=0.5$V and resulting target $V_{ov}=0.33$ V (=2/3*V_{DD}), this implies that these MOSCAP $In_{0.53}Ga_{0.47}As$ devices with the 1 nm IL/1 nm $LaSiO_x$/3 nm HfO_2 gate stack surpass the IIIV reliability requirements.

Stabilization of the IL by the Al_2O_3 or $LaSiO_x$ cap is confirmed from the TOFSIMS profiles measured on dot capacitors that received a PMA (Fig. 12). These profiles show the mixing of the Al atoms (from the Al_2O_3) or La atoms (from the $LaSiO_x$) with the IL. Diffusion of the IL atoms down into the InGaAs substrate is clearly present if no cap layer is deposited between IL and the HfO_2 (Fig. 12a). When an Al_2O_3 or $LaSiO_x$ cap is added on top of the IL (Figs. 12b and 12c) , a steep slope of the IL elements down into the substrate is observed. The hypothesis is that the mixing of the Al or La atoms with the IL atoms is increasing the stability of the IL. Although the chemistry from both Al_2O_3 and $LaSiO_x$ cap layers can provide stabilization of the interaction of the IL with the substrate, the electrical characteristics of the $LaSiO_x$ cap are more favorable.

Modeling of the shallow and deep defect bands using the approach as described in [26], was done and explains the effect of the $LaSiO_x$ cap (Fig. 13). The introduction of the IL layer narrows the distribution of the shallow defect band as well as shifts the mean energy of that band to shallower energies. Adding $LaSiO_x$ in between the IL and HfO_2 further reduces the charging defect density. This reduction can be attributed to the stabilization that is provided by the $LaSiO_x$ cap.

3.3. MOSFET device characteristics

Figs. 14 and 15 show the I_{ds}-V_{gs} and g_m-V_{gs} curves of the devices fabricated with stacks B (Al_2O_3 cap) and C ($LaSiO_x$ cap). Excellent device characteristics in the form of high I_{on}/I_{off} ratio (> $2x10^6$, > $8x10^5$ at $V_{ds}=0.05$ V and 0.5 V respectively) and low SS_{lin} (< 75 mV/dec) were extracted for both the gate stacks. The device characteristics for stacks B, C, C', and C'' are given in Fig. 16 and 17. The on-state current increases (Fig. 16a and b) and the SS_{lin} decreases (Fig. 17) when Al_2O_3 (B) is replaced by $LaSiO_x$ (C) between the IL layer and HfO_2. When scaling the HfO_2 to 2 nm (C''), a CET of 1.46 nm (EOT of 1.06 nm) is achieved with a
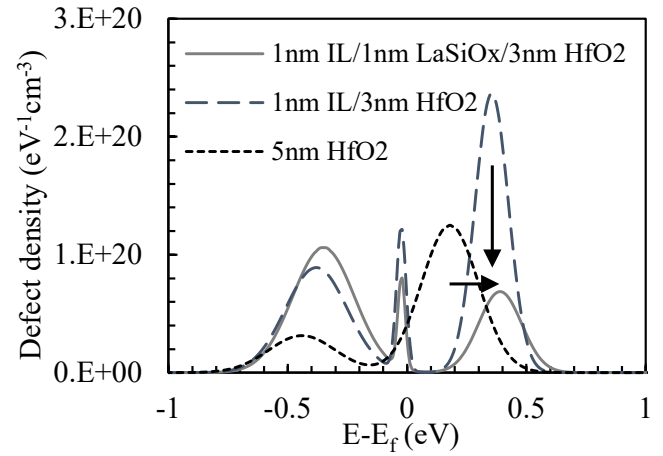
Figure 13: Defect density profile obtained by modeling the ΔV_{th} vs V_{gate} as described in [25]. Insertion of the IL layer causes narrowing of the shallow defect band, thus enhancing γ and shifting the mean defect level of this band to shallower energy [19] but increases the density of defects at deeper energies. Insertion of $LaSiO_x$ consistently reduces the total defect density (both shallow and deep energies).

further improvement in the on-characteristics, and with a reduction in the SS_{lin} to 68 mV/dec. Note that the stack without the IL layer (1 nm $LaSiO_x$/3 nm HfO_2 stack only, C'), exhibits high on-current compared to the stack with the 1 nm IL/1 nm Al_2O_3/3 nm HfO_2 but
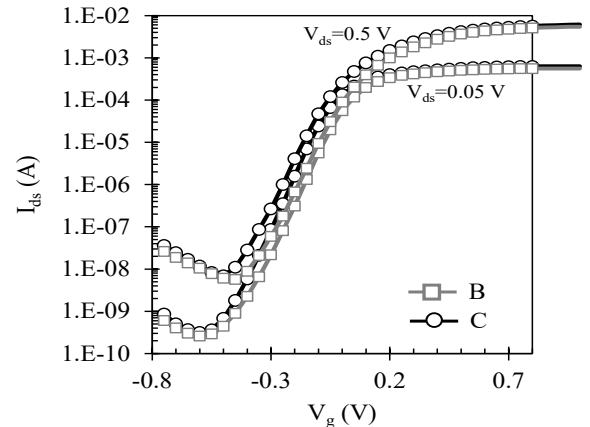
Figure 14: I_{ds}-V_{gs} for L_g=10 μm, W=100 μm device.

Figure 15: G_m-V_{gs} for L_g=10μm, W=100μm device.

Figure 16: I_{on} extracted at V_t+0.5V and for V_{ds}=0.05V (a) and V_{ds}=0.5V (b), W of the devices = 100μm, Lg=10μm. Replacement of Al_2O_3 (B) by $LaSiO_x$ (C) in the gate stack improves I_{on}. Further improvement can be seen when the gate stack is scaled down (C'').
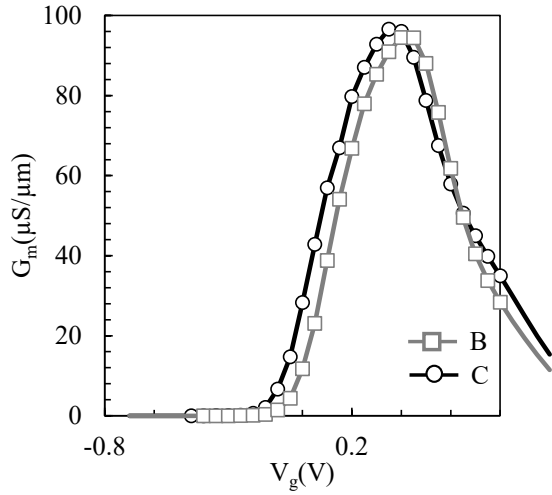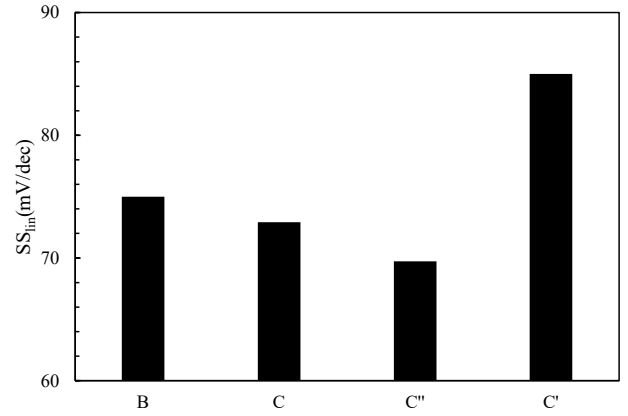
Fig. 17 SS_{lin} for various gate stacks. Lg=50 μm. SS_{lin} reduces by replacing Al_2O_3 (B) with $LaSiO_x$ (C) in the gate stack. Scaling the HfO_2 (C'') thickness leads to further reduction.

Figure 18 $V_{t, lin}$ for various gate stacks. L_g=50 μm. Slight Vt_{lin} reduction by replacing Al_2O_3 (B) with $LaSiO_x$ (C) in the gate stack is observed.

impact of insertion of the IL layer is on the peak mobility (μ_{eff} at peak = 3531 cm^2/V.s), the replacement of the Al_2O_3 by $LaSiO_x$ increases the electron mobility at high charge carrier density.
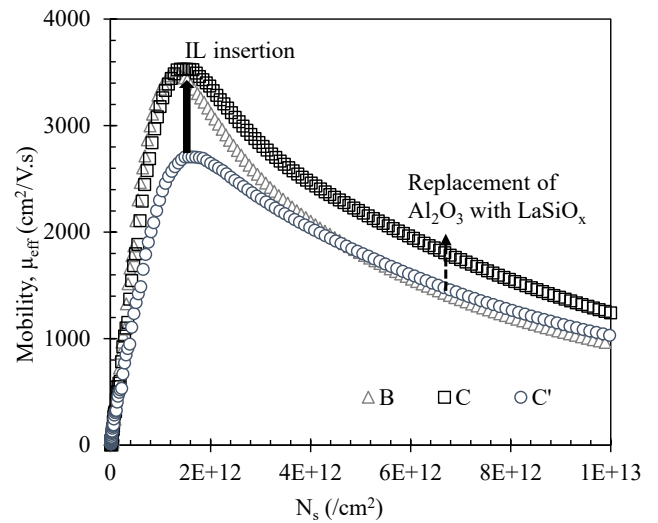
Figure 19 Improvement in the mobility at low N_s (due to the introduction of IL) and high N_s (as a result of replacement of Al_2O_3 with $LaSiO_x$). Mobility was extracted from L_g=50 μm devices.

also a high SS_{lin} reflecting a higher mid gap D_{it} extracted from the MOS capacitor C-V's (Fig. 7 and 8). Only slight changes in V_{th_lin} (~50 mV) are observed when changing the stabilization cap (Fig. 18).

The mobility data (Fig. 19) reveal that while the dominant

This observation suggests a reduction in the roughness using LaSiO$_x$ in the gate stack instead of Al$_2$O$_3$. In addition, BTI measurements on MOSFETs confirm the good reliability observed on simple MOS capacitor structures. ΔN_{eff} and γ values were extracted under both positive and negative stress conditions (Fig. 20). Under both stress conditions, we can observe a reduction of ΔN_{eff} when adding the IL layer and further reduction is achieved by insertion of the LaSiO$_x$ cap. In parallel, the value of γ increases when adding the IL layer but levels off when inserting the LaSiO$_x$. This confirms both the
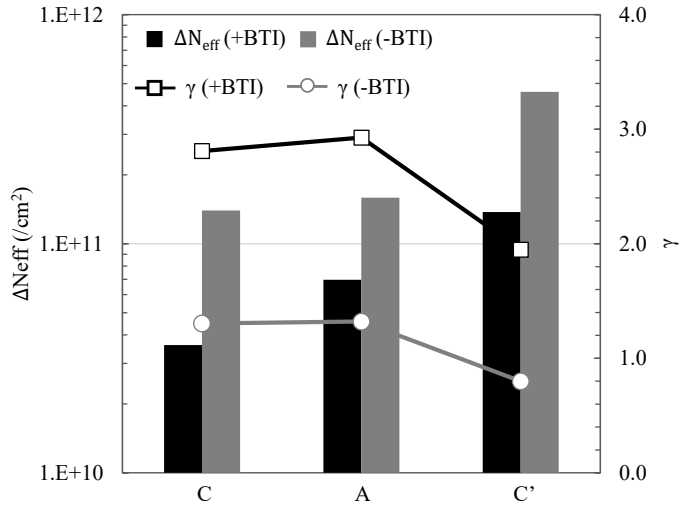
Figure 20 ΔN_{eff} and γ extracted from BTI measurements on MOSFETs applying both positive and negative stress conditions. The data from MOSFETs confirm the trends observed on MOSCAPs.

observation in measurement data and modeling results from the MOSCAP data i.e. the IL layer provides the misalignment between the defect levels in the high-κ and the InGaAs E$_f$ while the LaSiO$_x$ reduced the total charging oxide defects. At negative stress conditions, the differences are less pronounced.

The extrapolated max V$_{ov}$ and V$_{un}$ at 10 years are shown in Fig. 21. Benchmarking the max V$_{ov}$ of the newly proposed gate stack against Si gate stacks (Fig. 22) shows that by using the optimized gate stack III-V devices can meet even the more stringent reliability targets used for a scaled Si technology.
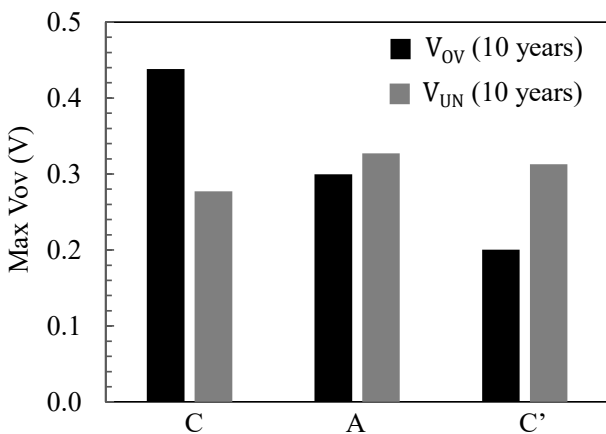
Figure 21. Max V$_{ov}$ and V$_{un}$ from BTI measurements on MOSFETs with different gate stacks.
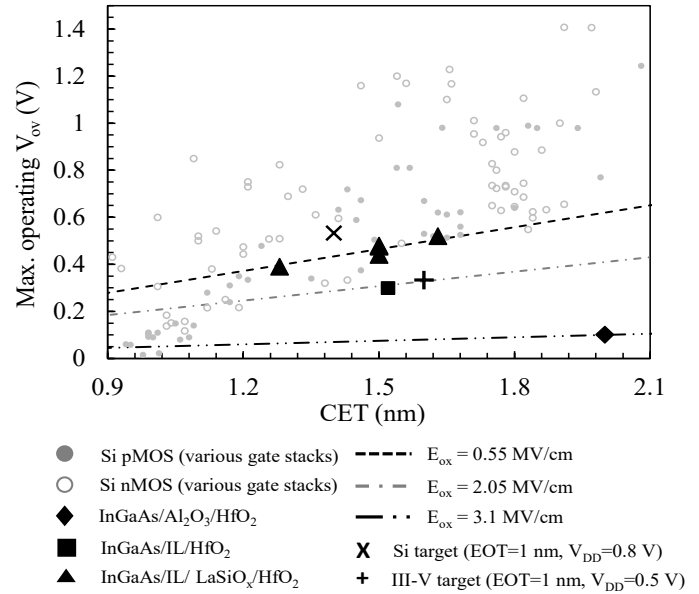
Figure 22. Max. operating overdrive voltage as a function of CET for various gate stacks on Si (p- and n-MOS) and InGaAs. The devices with the IL/LaSiO$_x$/HfO$_2$ gate stack surpass the reliability target for III-V semiconductor based technology.

4. Conclusions

We have shown in this work the improved interface and reliability properties of the IL layer. A D$_{it}$ and ΔN_{eff} reduction are accompanied with a negative V$_{fb}$ shift and an increase in the value of γ. This suggests the formation of an interface dipole. When a favorable defect band distribution is present as is the case for HfO$_2$, dipole engineering can be applied to misalign the defect band with the InGaAs channel E$_f$. However, the thermal stability of this stack is limited and a severe intermixing of the IL layer with the InGaAs substrate at typical thermal budget of a IIIV gate stack, degrades the initial improvement seen from the IL layer.

A cap is needed to stabilize this interaction and both Al$_2$O$_3$ and LaSiO$_x$ can act as a stabilization layer by La or Al mixing with the IL-layer atoms. While the physical properties are very similar, the electrical properties in terms of oxide traps are very different. The different behavior can be explained by the presence of a broad defect band in Al$_2$O$_3$ making this cap layer not suitable for dipole engineering. On the other hand, LaSiO$_x$ shows a more favorable defect band distribution and a further reduction of the oxide defects is observed. In addition, the interface properties are improved and low CET ($\sim$1.5 nm) was maintained.

Excellent device characteristics were demonstrated combining the IL passivation layer with the bi-layer stack consisting of 1 nm LaSiO$_x$/3 nm HfO$_2$. III-V devices with record mobility and record reliability performance at scaled EOT were demonstrated. In addition, we show that this stack can be further scaled down to 1nm EOT without loss of performance or reliability.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] S. Sioncke, J. Franco, A. Vais, V. Putcha, L. Nyns, A. Sibaja-Hernandez, R. Rooyackers, S. Calderon Ardila, V. Spampinato, A. Franquet, J.W. Maes,

Q.Xie, M.Givens, F. Tang, X. Jiang, M. Heyns, D. Linten, J. Mitard, A. Thean, D. Mocuta and N. Collaert, "First demonstration of ~3500 cm²/V-s electron mobility and sufficient BTI reliability (max V_{ov} up to 0.6 V) In$_{0.53}$Ga$_{0.47}$As nFET using an IL/LaSiO$_X$/HfO$_2$ gate stack", in IEEE Symposium on VLSI Technology 2017, pp. T38-T39, 2017. https://doi.org/10.23919/VLSIT.2017.7998192

[2] A. Vais, A. Alian, L. Nyns, J. Franco, S. Sioncke, V. Putcha H. Yu, Y. Mols, R. Rooyackers, D. Lin, J.W. Maes, Q. Xie, M. Givens, F. Tang, X. Jiang, A. Mocuta, N. Collaert, K. De Meyer, A. Thean, "Record mobility (μ$_{eff}$~ 3100 cm²/Vs) and reliability performance (V$_{ov}$~ 0.5 V for 10yr operation) of In$_{0.53}$Ga$_{0.47}$As MOS devices using improved surface preparation and a novel interfacial layer" in IEEE Symposium on VLSI Technology, pp. 140-141. 2016. https://doi.org/10.1109/VLSIT.2016.7573410

[3] C. B. Zota, F. Lindelöw, L.-E. Wernersson, and E. Lind. "InGaAs nanowire MOSFETs with I$_{ON}$= 555 μA/μm at I$_{OFF}$= 100 nA/μm and V$_{DD}$= 0.5 V" in IEEE Symposium on VLSI Technology, pp. 1-2. IEEE, 2016. https://doi.org/10.1109/VLSIT.2016.7573418

[4] N. Waldron, S. Sioncke, J. Franco, L. Nyns, A. Vais, X. Zhou, H. C. Lin, G. Boccardi, J. W. Maes, Q. Xie and M. Givens, F. Tang , X. Jiang , E. Chiu, A. Opdebeeck, C. Merckling, F. Sebaai, D. van Dorp, L. Teugels, A. Sibaja Hernandez, K. De Meyer, K. Barla, N. Collaert, A. Thean, "Gate-all-around InGaAs nanowire FETS with peak transconductance of 2200 μS/μm at 50 nm L$_g$ using a replacement Fin RMG flow" in IEEE International Electron Devices Meeting (IEDM), pp. 31-1. IEEE, 2015. https://doi.org/10.1109/IEDM.2015.7409805

[5] R. Suzuki, N. Taoka, M. Yokoyama, S. Lee, S. H. Kim, T. Hoshii, T. Yasuda, W. Jevasuwan, T. Maeda, O. Ichikawa, N. Fukuhara, M. Hata, M. Takenaka, and S. Takagi, " 1-nm-capacitance-equivalent-thickness HfO$_2$/Al$_2$O$_3$/InGaAs metal-oxide-semiconductor structure with low interface trap density and low gate leakage current density," Appl. Phys. Lett. **100**(13), 132906–1–132906-3 (2012). https://doi.org/10.1063/1.3698095.

[6] F Palumbo, M Eizenberg, "Degradation characteristics of metal/Al$_2$O$_3$/n-InGaAs capacitors", Journal of Applied Physics, 115 (1), 014106 (2014). DOI: https://doi.org/10.1063/1.4861033

[7] G. Jiao, C. Yao, Y. Xuan, D. Huang, P.D. Ye, and M.-F. Li, "Experimental Investigation of Border Trap Generation in InGaAs nMOSFETs With Al2O3 Gate Dielectric Under PBTI Stress", IEEE Trans. Electron Dev., 59, pp. 1661-67 (2012). https://doi.org/10.1109/TED.2012.2190417

[8] I Krylov, B Pokroy, M Eizenberg, and D Ritter, "A comparison between HfO$_2$/Al$_2$O$_3$ nano-laminates and ternary Hf$_x$Al$_y$O compound as the dielectric material in InGaAs based metal-oxide-semiconductor (MOS) capacitors", Journal of Applied Physics, 120 (12), 124505 (2016). https://doi.org/10.1063/1.4962855

[9] K. Tang, R. Droopad, and P. C. McIntyre. "Border Trap Density in Al$_2$O$_3$/InGaAs MOS: Dependence on Hydrogen Passivation and Bias Temperature Stress", ECS Trans., 69 (5), pp. 53-60 (2015). https://doi.org/ 10.1149/06905.0053ecst

[10] C. L. Hinkle, M. Milojevic, B. Brennan, A. M. Sonnet, F. S. Aguirre-Tostado, G. J. Hughes, E. M. Vogel, R. M. Wallace, "Detection of Ga suboxides and their impact on III-V passivation and Fermi-level pinning", Appl. Phys. Lett. 94, 162101 (2009). https://doi.org/10.1063/1.3120546

[11] D. Lin, A. Alian, S. Gupta, B. Yang, E. Bury, S. Sioncke, R. Degraeve, M. L. Toledano, R. Krom, P. Favia, H. Bender, M. Caymax, K. C. Saraswat, N. Collaert, A. Thean, "Beyond interface: The impact of oxide border traps on InGaAs and Ge n-MOSFETs", in Proceedings of IEEE International Electron Device Meeting Technical Digest, 28.3.1 (2012). https://doi.org/10.1109/IEDM.2012.6479121

[12] S. Deora, G. Bersuker, W.-Y. Loh, D. Veksler, K. Matthews, T. W. Kim, R. T. P. Lee, R. J. W. Hill, D.-H. Kim, W.-E. Wang, C. Hobbs, P. D. Kirsch, "Positive Bias Instability and Recovery in InGaAs Channel nMOSFETs", IEEE Transactions on Device and Materials Reliability, 13(4), 507 (2013). https://doi.org/10.1109/TDMR.2013.2284376

[13] J. Franco A. Alian, B. Kaczer, D. Lin, T. Ivanov, A. Pourghaderi, K. Martens, Y. Mols, D. Zhou, N. Waldron, S. Sioncke, T. Kauerauf, N. Collaert, A. Thean, M. Heyns, and G. Groeseneken, "Suitability of high-k gate oxides for III-V devices: a PBTI study in In$_{0.53}$Ga$_{0.47}$As devices with Al$_2$O$_3$", in Proceedings of IEEE International Reliability Physics Symposium Technical Digest, 6A.2.1 (2014). https://doi.org/ 10.1109/IRPS.2014.6861098

[14] N. Conrad, M. Si, S. H. Shin, J. J. Gu, J. Zhang, M. A. Alam, P. D. Ye, "Low-frequency noise and RTN on near-ballistic III–V GAA nanowire MOSFETs", in Proceedings of IEEE International Electron Devices Meeting, 20.1.1 (2014). https://doi.org/10.1109/IEDM.2014.7047086

[15] N. Taoka, M. Yokoyama, S. H. Kim, R. Suzuki, S. Lee, R. Iida, T. Hoshii, W. Jevasuwan, T. Maeda, T. Yasuda, O. Ichikawa, N. Fukuhara, M. Hata, M.Takenaka and S. Takagi, "Impact of Fermi Level Pinning Due to Interface Traps Inside the Conduction Band on the Inversion-Layer Mobility in InGaAs Metal–Oxide–Semiconductor Field Effect Transistors". IEEE Trans. Dev. Mat. Rel. 13, 456 (2013). https://doi.org/10.1109/TDMR.2013.2289330.

[16] S. Stemmer, V. Chobpattana, and S. Rajan, "Frequency dispersion in III-V metal-oxide-semiconductor capacitors", Appl. Phys. Lett. 100, 233510 (2012). https://doi.org/10.1063/1.4724330.

[17] S. . Johansson, M. Berg, K. M. Persson and E. Lind, "Transconductance method for characterization of border traps", *IEEE Trans. Electron Devices*, vol. 60, no. 2, February 2013. https://doi.org/ 10.1109/TED.2012.2231867

[18] J. Franco, B. Kaczer, Ph. J. Roussel, J. Mitard, S. Sioncke, L. Witters, H. Mertens, T. Grasser, and G. Groeseneken, "Understanding the suppressed charge trapping in relaxed-and strained-Ge/SiO$_2$/HfO$_2$ p-MOSFETs and implications for the screening of alternative high-mobility substrate/dielectric CMOS gate stacks" in Proceedings of IEEE International Electron Devices Meeting, 15.2.1 (2013). https://doi.org/ 10.1109/IEDM.2013.6724634

[19] J. Franco, A. Vais, S. Sioncke, V. Putcha, B. Kaczer, B-S. Shie, X. Shi, R. Mahlouji, L. Nyns, D. Zhou, and N. Waldron, "Demonstration of an InGaAs gate stack with sufficient PBTI reliability by thermal budget optimization, nitridation, high-k material choice, and interface dipole" in 2016 IEEE Symposium on VLSI Technology, , pp. 1-2., 2016. https://doi.org/ 10.1109/VLSIT.2016.7573371

[20] A. Alian, M. A. Pourghaderi, Y. Mols, M. Cantoro, T. Ivanov, N. Collaert, and A. Thean, "Impact of the channel thickness on the performance of ultrathin InGaAs channel MOSFET devices", in Proceedings of IEEE International Electron Device Meeting, pp.16.6.1-16.6.4, 2013. https://doi.org/ 10.1109/IEDM.2013.6724644

[21] E. H. Nicolian and J. R. Brews, MOS Phys. and Tech., p. 215-216, Wiley Interscience, New York (2003).

[22] R. Winter, J. Ahn, P. C. McIntyre, and M. Eizenberg, "New method for determining flat-band voltage in high mobility semiconductors", Journal of Vacuum Science & Technology B **31**, 030604 (2013). https://doi.org/ 10.1116/1.4802478

[23] D. K. Schroder, Semiconductor material and device characterization. John Wiley and Sons, 2006, ISBN: 978-0-471-73906-7.

[24] B. Kaczer, T. Grasser, P.J. Roussel, J. Martin-Martinez, R. O'Connor, B. O'Sullivan, and G. Groeseneken, "Ubiquitous relaxation in BTI stressing— New evaluation and insights" in proceedings of IEEE International Reliability Physics Symposium, 2008, pp. 20-27, 2008. https://doi.org/ 10.1109/RELPHY.2008.4558858

[25] A. Vais, J. Franco, H.-C. Lin, N. Collaert, A. Mocuta, K. De Meyer, and A. Thean, "Impact of starting measurement voltage relative to flat-band voltage position on the capacitance-voltage hysteresis and on the defect characterization of InGaAs/high-k metal-oxide-semiconductor stacks", Appl. Phys. Lett. 107, 223504 (2015). https://doi.org/ 10.1063/1.4936991

[26] V. Putcha, J. Franco, A. Vais, B. Kaczer, S. Sioncke, D. Linten and G. Groeseneken, "Impact of slow and fast oxide traps on In$_{0.53}$Ga$_{0.47}$As device operation studied using CET maps", in proceedings of IEEE International Reliability Physics Symposium 2017, XT-8.1 (2017). https://doi.org/ 10.1109/IRPS.2018.8353603

Parametric Co-variance Assignment for a Class of Multi-variable Stochastic Uncertain Systems: Output Feedback Stabilization Approach

Qichun Zhang[*]

School of Engineering and Sustainable Development, De Montfort University, Leicester, LE1 9BH, UK

A R T I C L E I N F O

A B S T R A C T

This paper presents a novel parametric co-variance assignment strategy for multi-variable stochastic uncertain systems. Based upon the explicit parametric design and reduced-order closed-form co-variance model, the variances and co-variances of the system outputs can be assigned artificially using output feedback while the effect of the system uncertainties can be minimized by optimizing the free parameters. In addition, the stability of the closed-loop system has been analyzed and an illustrative numerical example is given to demonstrate the effectiveness of the presented strategy. As a summary, the contributions of this paper include the reduced-order co-variance model, the co-variance error based performance criterion and the parametric control design with stability analysis.

1 Introduction

Co-variance analysis permeates almost all of system theory [1]. Based on the co-variance analysis, the probabilistic properties among random signals can be described. Naturally, the associate co-variance control problem became one of the most significant research topics for multi-variable stochastic systems. Since the system identification, Kalman filtering, stochastic distribution control and fault diagnosis are widely used in practice, the co-variance estimation and control are significant to all of these research areas and applications [2, 3, 4, 5]. In addition, the co-variance is also an ideal tool to analyze the performance of the stochastic systems for the probabilistic decoupling analysis [6, 7] and neural interaction analysis [8, 9].

During the past two decades, the main result of co-variance control is based on the Lyapunov equation while several conditions and controllers have proposed the control design for the co-variance assignment using the determined control signals [10, 11]. However, this controller is designed by Lyapunov equation without closed-form formulation. Since the closed-form model of state co-variance [12] presented in 2007, reduced-order co-variance model [13] has been presented based on eigen-decomposition to solve output co-variance assignment (OCA) problem while the co-variance model

was presented for the stochastic system with parametric uncertainty. However, all the existing results did not consider the control design with parametric uncertainties of the stochastic systems. Considering the uncertainties of the parameters, the robust controller has been designed in [14, 15, 16]. All the mentioned controllers can achieve good performance, however these methods have not been used to deal with the co-variance analysis. To the best of our best knowledge, there is no existing solution to the parametric output-feedback co-variance assignment for stochastic uncertain systems with the stability analysis. Therefore, it is significant to develop a simply output co-variance assignment (OCA) control law for the implementation of the complex dynamic multi-variable stochastic system with uncertainties.

In this paper, the stochastic uncertain multi-variable systems have been investigated while the co-variance assignment is very difficult since the uncertainties would affect the stability of the closed-loop systems. Based upon the investigated uncertain multi-variable model, the transformed co-variance model can be obtained firstly, and then the controller and linear observer can be designed and analyzed. In particular, the sufficient conditions are given for the convergence of the observer, the stabilization of the controller and

---
[*]Q. Zhang. Email:qichun.zhang@dmu.ac.uk

stabilization of the closed-loop system, respectively. In the end, the control input can be obtained by reversing the transformation. Using this control strategy, the design procedure is also given. Furthermore, the parameters of the controller and observer can be optimized using the parametric state feedback [17, 18] and entropy-based performance criterion. Using the presented control strategy, the optimal output feedback control law is obtained for OCA and the performance has been verified by the numerical simulation.

The rest of the paper is organized as follows. In Section 2, the formulation is given including the model formulation, transformed co-variance model and control objective. The parametric control strategy is developed while the convergence of the linear observer, the stabilization of the parametric state feedback controller and the stabilization of closed-loop uncertain system are analyzed in Section 3. Moreover, the parameter optimization and design procedure are also given in this section. Section 4 and Section 5 present the results of numerical simulation and the conclusions, respectively.

2 Formulation

Suppose that the complex industrial dynamic process can be modeled by the following stochastic uncertain multi-variable system.

$$
\begin{aligned}
dx(t) &= (A + \Delta A(t))x(t)dt + (B + \Delta B(t))u(t)dt + Dd\beta_t \\
y(t) &= Cx(t)
\end{aligned}
\tag{1}
$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^s$ and $y \in \mathbb{R}^m$ are the system state vector, input vector and output vector, respectively. β_t is the p-dimensional Wiener process. m, n and s are positive integers while system matrices A, B, C D and parameter uncertainties $\Delta A(t)$, $\Delta B(t)$ are of appropriate dimensions.

Notice that the model (1) can be rewritten as follows:

$$
\begin{aligned}
dx(t) &= ((A + \Delta A(t))x(t) + (B + \Delta B(t))u(t) + Dw(t))dt \\
y(t) &= Cx(t)
\end{aligned}
\tag{2}
$$

where w is a standard Gaussian white noise. Without loss of generality, assume that the investigated system model (1) satisfies the following assumptions.

Assumption A1: Assume that the Gaussian noise vector $w(t)$ satisfies

$$
E\{w(t)\} = 0
$$

$$
E\{w(t)w^T(t)\} = 0
$$

$$
E\{w(t)w^T(\tau)\} = Q\delta(t - \tau)
\tag{3}
$$

where $\delta(\cdot)$ is the Dirac delta function.

Assumption A2: Similar to the assumption of the noise, assume that the control signal is restricted by

$$
E\{u(t)\} = 0
$$

$$
E\{u(t)u^T(t)\} = 0
$$

$$
E\{u(t)u^T(\tau)\} = U(t)\delta(t - \tau)
\tag{4}
$$

where $U(t) \overset{\Delta}{=} E\{u(t)u^T(t)\}$.

Based on the definition of the co-variance matrix, the state and output co-variance matrices of the given system are given by

$$
\begin{aligned}
P_x(t) &\overset{\Delta}{=} E\{x(t)x^T(t)\} \\
P_y(t) &\overset{\Delta}{=} E\{y(t)y^T(t)\}
\end{aligned}
\tag{5}
$$

In addition, the co-variance matrices can be rewritten as follows:

$$
\begin{aligned}
P_x(t) &= V_x \Lambda_x(t) V_x^T \\
P_y(t) &= V_y \Lambda_y(t) V_y^T \\
U(t) &= V_u \Lambda_u(t) V_u^T \\
Q &= V_q \Lambda_q V_q^T
\end{aligned}
\tag{6}
$$

where Λ_x, Λ_y, Λ_u and Λ_q are real diagonal matrices. V_x, V_y, V_u and V_q are associated orthogonal matrices. All of the matrices are with the same dimensions as the associated vectors.

Therefore, the reduced-order closed-form co-variance model can be obtained following the vectorization operation.

$$
\begin{aligned}
\dot{\lambda}_x(t) &= (A_{\text{cov}} + \Delta A_{\text{cov}})\lambda_x(t) + (B_{\text{cov}} + \Delta B_{\text{cov}})\lambda_u(t) + D_{\text{cov}}\lambda_q \\
\lambda_y(t) &= C_{\text{cov}}\lambda_x(t)
\end{aligned}
\tag{7}
$$

where λ_x, λ_y, λ_u and λ_q are the diagonal elements of matrices Λ_x, Λ_y, Λ_u and Λ_q, respectively. The coefficient matrices can be calculated using Hadamard product as follows:

$$
\begin{aligned}
A_{\text{cov}} &= A_\Lambda \circ A_\Lambda \\
B_{\text{cov}} &= B_\Lambda \circ B_\Lambda \\
D_{\text{cov}} &= D_\Lambda \circ D_\Lambda \\
C_{\text{cov}} &= C_\Lambda \circ C_\Lambda \\
\Delta A_{\text{cov}} &= \Delta A_\Lambda \circ \Delta A_\Lambda + 2(A_\Lambda \circ \Delta A_\Lambda) \\
\Delta B_{\text{cov}} &= \Delta B_\Lambda \circ \Delta B_\Lambda + 2(B_\Lambda \circ \Delta A_\Lambda)
\end{aligned}
\tag{8}
$$

while $\Delta A_\Lambda = V_x^T \Delta A V_x$, $\Delta B_\Lambda = V_x^T \Delta B V_u$, $\Delta C_\Lambda = V_y^T \Delta C V_x$, $A_\Lambda = V_x^T A V_q$, $B_\Lambda = V_x^T B V_q$, $C_\Lambda = V_x^T C V_q$ and $D_\Lambda = V_x^T D V_q$. Moreover, reserving the diagonal elements of A_Λ which forms a diagonal matrix A_{diag}, then $A_{\text{cov}} = 2A_{diag}$. Similarly, $\Delta A_{\text{cov}} = 2\Delta A_{diag}$.

Thus, the control objective is to develop a new control strategy to assign the state of the transformed model (7) so that the covariances and variances of the system outputs can be assigned simultaneously.

To achieve the mentioned control objective, the following assumption should be taken into account:

Assumption A3: The pair $(A_{\text{cov}}, B_{\text{cov}})$ is controllable.

Assumption A4: The admissible parameter uncertainties are of the norm-bounded form

$$
\begin{bmatrix} \Delta A_{\text{cov}}(t) & \Delta B_{\text{cov}}(t) \end{bmatrix} = M \begin{bmatrix} \Xi_1(t)N_1 & \Xi_2(t)N_2 \end{bmatrix}
\tag{9}
$$

In Eq. (9), M, N_1 and N_2 denote the structure of the uncertainties which are known real constant matrices

with proper dimensions. $\Xi_1(t)$ and $\Xi_2(t)$ are unknown time-varying matrices which respectively meet the following conditions.

$$\Xi_1^T(t)\Xi_1(t) \le I, \Xi_2^T(t)\Xi_2(t) \le I \quad (10)$$

In addition, the following lemma [19] has been recalled here which can be used to analyze the convergence and stabilization of the presented control strategy.

Lemma 1. *Given any real constant matrices X and Y with proper dimensions. Then there exists a constant $\xi > 0$, such that the following inequality holds.*

$$X^T Y + Y^T X \le \xi X^T X + \xi^{-1} Y^T Y \quad (11)$$

3   Control Strategy

The reference co-variance matrix can be rewritten as follows:

$$R = V_r \Lambda_r V_r^T \quad (12)$$

Since the diagonal matrix r can be arranged as vector λ_r, the co-variance assignment problem transfers to state tracking problem using the presented reduced-order co-variance model if we set $V_y = V_r$.

To track the desired state co-variance vector, the integrator should be considered in the control scheme. The error vector $e_y(t) = \lambda_r \lambda_y$ is treated as the extended state and substitutes the error into the closed-loop system.

Then, the closed-loop system in the state-space form can be obtained as follows:

$$\begin{bmatrix} \dot{\lambda}_x(t) \\ \dot{\bar{e}}_y(t) \end{bmatrix} = (\bar{A} + \Delta\bar{A}) \begin{bmatrix} \lambda_x(t) \\ \bar{e}_y(t) \end{bmatrix} + (\bar{B} + \Delta\bar{B})\lambda_u(t)$$
$$+ \bar{D}\lambda_q + \begin{bmatrix} 0 \\ \lambda_r \end{bmatrix}$$
$$\lambda_y(t) = \bar{C} \begin{bmatrix} \lambda_x(t) \\ \bar{e}_y(t) \end{bmatrix}$$

$$(13)$$

where

$$\bar{A} = \begin{bmatrix} A_{\text{cov}} & 0 \\ -C_{\text{cov}} & 0 \end{bmatrix}, \bar{B} = \begin{bmatrix} B_{\text{cov}} \\ 0 \end{bmatrix}, \bar{C} = \begin{bmatrix} C_{\text{cov}}^T \\ 0 \end{bmatrix}^T$$
$$\bar{D} = \begin{bmatrix} D_{\text{cov}} \\ 0 \end{bmatrix} \Delta\bar{A} = \begin{bmatrix} \Delta A_{\text{cov}} & 0 \\ 0 & 0 \end{bmatrix}, \Delta\bar{B} = \begin{bmatrix} \Delta B_{\text{cov}} \\ 0 \end{bmatrix}$$

$$(14)$$

and

$$\dot{\bar{e}}_y(t) = e_y(t) \quad (15)$$

We can further present the closed-loop system in compact format as follows:

$$\dot{\bar{x}}(t) = (\bar{A} + \Delta\bar{A})\bar{x}(t) + (\bar{B} + \Delta\bar{B})\lambda_u(t) + \bar{D}\lambda_q + \bar{\lambda}_r$$
$$\lambda_y(t) = \bar{C}\bar{x}(t) \quad (16)$$

where

$$\bar{x}(t) = \begin{bmatrix} \lambda_x(t) \\ \bar{e}_y(t) \end{bmatrix}, \bar{\lambda}_r = \begin{bmatrix} 0 \\ \lambda_r \end{bmatrix} \quad (17)$$

Notice that $\bar{D}\lambda_q$ and $\bar{\lambda}_r$ are real constants which means that if the following system is stabilized then the variance and co-variance assignment can be completed.

$$\dot{\bar{x}}(t) = (\bar{A} + \Delta\bar{A})\bar{x}(t) + (\bar{B} + \Delta\bar{B})\lambda_u(t)$$
$$\lambda_y(t) = \bar{C}\bar{x}(t) \quad (18)$$

Based upon the transformed system model 18, the control strategy can be divided into two parts: observer-based output feedback design and the parametric optimization for uncertainty compensation.

3.1   State feedback design

The linear state feedback controller can be determined by the nominal linear model and the control law is described by

$$\lambda_u(t) = K\bar{x}(t) \quad (19)$$

where the gain matrix K can be obtained by parametric design [17, 13]. In particular, we have

$$K = [W_1 f_1, \ldots, W_m f_m]$$
$$\times \left[(\lambda_1^* I - A_1)^{-1} B_1 f_1, \ldots, (\lambda_m^* I - A_m)^{-1} B_m f_m \right]^{-1} \quad (20)$$

where modified parameter vectors and closed-loop eigenvalues are denoted by $f_1, \ldots, f_m$ and $\lambda_1^*, \ldots, \lambda_m^*$ which can be considered as free parameters. In the case of a common open-loop and closed-loop eigenvalue, the gain matrix K can be determined by the following equations.

$$A_j = \bar{A} + v_j^0 w_j^{0T}$$
$$W_j = I - \frac{e_i w_j^{0T} \bar{B}}{w_j^{0T} b_i}$$
$$B_j = \bar{B} W_j + v_j^0 e_i^T, j = 1, \ldots, m \quad (21)$$

where v_j^0 and w_j^0 denote the open-loop eigenvectors and eigenrows of A_j. b_i is the i-th column of $\bar{B}$. e_i is a unit vector while the i-th element is 1. In the other case, no common eigenvalue results in $w_j^{0T} b_i = 0$ which leads to

$$A_j = \bar{A}$$
$$W_j = I$$
$$B_j = \bar{B}, j = 1, \ldots, m \quad (22)$$

Substituting control law (19) into the transformed system model (18) yields the closed-loop system:

$$\dot{x}(t) = (A_c + \Delta A_c(t)) x(t) \quad (23)$$

where $A_c = \bar{A} + \bar{B}K$, $\Delta A_c(t) = \Delta\bar{A}(t) + \Delta\bar{B}(t)K$. Thus the following lemma can be proposed.

Lemma 2. *For the uncertain multi-variable system given by (18), with the assumptions and with the control law given by (19), then there exist two positive constants ε_1 and ε_2, so that the equilibrium $\bar{x}(t) = 0$ is stabilized if the following matrix inequality has a positive-definite solution $P = P^T > 0$.*

$$A_c^T P + P A_c + \varepsilon_1 N_1^T N_1 + \left(\varepsilon_1^{-1} + \varepsilon_2^{-1}\right) P M M^T P$$
$$+ \varepsilon_2 K^T N_2^T N_2 K + \varepsilon_3 P^T P < 0 \quad (24)$$

Proof. Consider the Lyapunov function candidate as

$$V_c(\bar{x}) = \bar{x}^T(t) P \bar{x}(t), P = P^T > 0 \quad (25)$$

The time derivative of $V_c(\bar{x})$ along the trajectories of (23) is given as follows.

$$\dot{V}_c(\bar{x}) = \bar{x}^T(t) A_c^T P \bar{x}(t) + \bar{x}^T(t) P A_c \bar{x}(t)$$
$$+ \bar{x}^T(t) \Delta \bar{A}^T(t) P \bar{x}(t) + x^T(t) P \Delta \bar{A}(t) \bar{x}(t)$$
$$+ \bar{x}^T(t) K^T \Delta \bar{B}^T(t) P \bar{x}(t) + \bar{x}^T(t) P \Delta \bar{B}(t) K \bar{x}(t) \quad (26)$$

Let ε_1 and ε_2 be positive constants, the following matrix inequalities hold using Lemma 1.

$$\bar{x}^T(t) \Delta \bar{A}^T(t) P \bar{x}(t) + x^T(t) P \Delta \bar{A}(t) \bar{x}(t)$$
$$= \bar{x}^T(t) (M \Xi_1 N_1)^T P \bar{x}(t) + \bar{x}^T(t) p M \Xi_1 N_1 \bar{x}(t)$$
$$\leq \bar{x}^T(t) (\varepsilon_1 N_1^T N_1 + \varepsilon_1^{-1} P M M^T P) \bar{x}(t) \quad (27)$$

$$\bar{x}^T(t) K^T \Delta \bar{B}^T(t) P \bar{x}(t) + \bar{x}^T(t) P \Delta \bar{B}(t) K \bar{x}(t)$$
$$= \bar{x}^T(t) K^T (M \Xi_2 N_2)^T P \bar{x}(t) + \bar{x}^T(t) P M \Xi_2 N_2 K \bar{x}(t)$$
$$\leq \bar{x}^T(t) (\varepsilon_2 K^T N_2^T N_2 K + \varepsilon_2^{-1} P M M^T P) \bar{x}(t) \quad (28)$$

Substituting these inequalities into the derivative of $V_c(\bar{x})$ with Assumption 4, we have

$$\dot{V}_c(\bar{x}) \leq \bar{x}^T(t) \left(A_c^T P + P A_c\right) \bar{x}(t)$$
$$+ \bar{x}^T(t) \left(\varepsilon_1 N_1^T N_1 + \varepsilon_1^{-1} P M M^T P\right) \bar{x}(t)$$
$$+ \bar{x}^T(t) \left(\varepsilon_2 K^T N_2^T N_2 K + \varepsilon_2^{-1} P M M^T P\right) \bar{x}(t) \quad (29)$$

Since $\dot{V}_c(\bar{x}) < 0$, the proof of lemma 2 is completed. □

3.2 Observer design

Using the linear observer to estimate the states of the model (18), the linear observer can be designed based on the nominal linear model.

$$\dot{\hat{x}}(t) = \left(\bar{A} - L\bar{C}\right) \hat{x}(t) + L \lambda_y(t) + \bar{B} \lambda_u(t) \quad (30)$$

where the estimated vector can be denoted by $\hat{x}$ and L is pre-specified gain matrix of this observer.

Introducing the error of the estimation by

$$e(t) = \bar{x}(t) - \hat{x}(t) \quad (31)$$

and substituting the Eq. (30)-(31) to system model (18). The closed-loop model can be described by

$$\dot{e}(t) = A_o e(t) + \Delta A_c(t) \bar{x}(t) \quad (32)$$

where $A_o = \bar{A} - L\bar{C}$. Similar to Lemma 2, Lemma 3 is given as follows.

Lemma 3. *For the uncertain multi-variable system given by (18), with the assumptions and with the linear observer given by (30), then there exists two positive constants ε_1 and ε_2, so that the estimation error $e(t)$ converges to zero if the following matrix inequalities have a positive-definite solution $P = P^T > 0$.*

$$A_o P + P A_o < 0 \quad (33)$$

$$\varepsilon_1 N_1^T N_1 + \left(\varepsilon_1^{-1} + \varepsilon_2^{-1}\right) P M M^T P + \varepsilon_2 K^T N_2^T N_2 K < 0 \quad (34)$$

Proof. Consider the Lyapunov function candidate as

$$V_o(e) = e^T(t) P e(t), P = P^T > 0 \quad (35)$$

The time derivative of $V_o(e)$ along the trajectories of (31) is given by the following equation.

$$\dot{V}_o(e) = e^T(t) (A_o P + P A_o) e(t)$$
$$+ \bar{x}^T(t) \Delta \bar{A}^T(t) P \bar{x}(t) + \bar{x}^T(t) P \Delta \bar{A}(t) \bar{x}(t)$$
$$+ \bar{x}^T(t) K^T \Delta \bar{B}^T(t) P \bar{x}(t) + \bar{x}^T(t) P \Delta \bar{B}(t) K \bar{x}(t) \quad (36)$$

Similar to the proof of Lemma 2, we have

$$\dot{V}_o(e) \leq e^T(t) (A_o P + P A_o) e(t)$$
$$+ \bar{x}^T(t) \left(\varepsilon_1 N_1^T N_1 + \varepsilon_1^{-1} P M M^T P\right) \bar{x}(t)$$
$$+ \bar{x}^T(t) \left(\varepsilon_2 K^T N_2^T N_2 K + \varepsilon_2^{-1} P M M^T P\right) \bar{x}(t) \quad (37)$$

which ends the proof □

3.3 Output feedback design

Combining the parametric state feedback controller and the designed observer, the output feedback controller can be obtained for the system (18).

$$\lambda_u(t) = K \hat{x}(t) \quad (38)$$

which leads to the closed-loop dynamics as follows.

$$\dot{\bar{x}}(t) = A_c \bar{x}(t) + \Delta A_c(t) \bar{x}(t) - \left(\bar{B} + \Delta \bar{B}(t)\right) K e(t) \quad (39)$$

Furthermore, the stability of the closed-loop control design can be guaranteed by the following theorem.

Theorem 4. *For the uncertain multi-variable system given by (18), with the assumptions and with the control law given by (38) using the observer (30), then there exists two sets of positive constants $\varepsilon_i, i = 1, \ldots, 2$ and $\varepsilon_j, j = 4, \ldots, 7$, so that the equilibrium $\bar{x}(t) = 0$ is stabilized if the following matrix inequalities have positive-definite solution $P_1 = P_1^T > 0, P_2 = P_2^T > 0$.*

$$\varepsilon_4 K^T B^T B K + \varepsilon_5 K^T N_2^T N_2 K + A_o P_2 + P_2 A_o + \varepsilon_8 P_2^T P_2 < 0 \quad (40)$$

$$A_c^T P_1 + P_1 A_c + (\varepsilon_1^{-1} + \varepsilon_2^{-1} + \varepsilon_4^{-1} + \varepsilon_5^{-1}) P_1 M M^T P$$
$$+ \left(\varepsilon_6^{-1} + \varepsilon_7^{-1}\right) P_2 M M^T P_2 + (\varepsilon_2 + \varepsilon_7) K^T N_2^T N_2 K$$
$$+ \varepsilon_3 P_1^T P_1 + (\varepsilon_3^{-1} + \varepsilon_8^{-1}) \gamma^2 I_1 + (\varepsilon_1 + \varepsilon_6) N_1^T N_1 < 0 \quad (41)$$

Proof. Consider the Lyapunov function candidate as

$$V(\bar{x}(t), e(t)) = \bar{x}^T(t) P_1 \bar{x}(t) + e^T(t) P_2 e(t) \qquad (42)$$

The time derivative of $V(\bar{x}(t), e(t))$ along the trajectories of (39) is shown as follows.

$$
\begin{aligned}
\dot{V}(\bar{x}(t), e(t)) &= \bar{x}^T A_c^T P \bar{x} + \bar{x}^T P A_c \bar{x} \\
&+ \bar{x}^T \Delta \bar{A}^T P \bar{x} + \bar{x}^T P \Delta \bar{A} \bar{x} \\
&+ \bar{x}^T K^T \Delta \bar{B}^T P \bar{x} + \bar{x}^T P \Delta \bar{B} K \bar{x} - e^T K^T \bar{B}^T P \bar{x} \\
&- \bar{x}^T P \bar{B} K e - e^T K^T \Delta \bar{B}^T P \bar{x} - \bar{x}^T P \Delta \bar{B} K e \\
&+ e^T(t)(A_o P + P A_o) e(t) \\
&+ \bar{x}^T(t) \Delta \bar{A}^T(t) P \bar{x}(t) \\
&+ \bar{x}^T(t) P \Delta \bar{A}(t) \bar{x}(t) + \bar{x}^T(t) K^T \Delta \bar{B}^T(t) P \bar{x}(t) \\
&+ \bar{x}^T(t) P \Delta \bar{B}(t) K \bar{x}(t) \qquad (43)
\end{aligned}
$$

Let ε_4 and ε_5 be positive constants, the following matrix inequalities hold using Lemma 1.

$$
\begin{aligned}
&- e^T(t) K^T \bar{B}^T P \bar{x}(t) - \bar{x}^T(t) P \bar{B} K e(t) \\
&\leq \varepsilon_4 e^T(t) K^T \bar{B}^T \bar{B} K e(t) + \varepsilon_4^{-1} \bar{x}^T(t) P M M^T P \bar{x}(t) \quad (44)
\end{aligned}
$$

$$
\begin{aligned}
&- e^T(t) K^T \Delta \bar{B}^T(t) P \bar{x}(t) - \bar{x}^T(t) P \Delta \bar{B}(t) K e(t) \\
&\leq \varepsilon_5 e^T(t) K^T N_2^T N_2 K e(t) + \varepsilon_5^{-1} \bar{x}^T(t) P M M^T P \bar{x}(t) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (45)
\end{aligned}
$$

Substituting these inequalities into the derivative of $V(\bar{x}(t), e(t))$ and using Lemma 3, we have

$$
\begin{aligned}
\dot{V} \leq \bar{x}^T(t) &[A_c^T P_1 + P_1 A_c + (\varepsilon_1 + \varepsilon_6) N_1^T N_1 \\
&+ (\varepsilon_2 + \varepsilon_7) K^T N_2^T N_2 K \\
&+ (\varepsilon_1^{-1} + \varepsilon_2^{-1} + \varepsilon_4^{-1} + \varepsilon_5^{-1}) P_1 M M^T P_1 \\
&+ (\varepsilon_6^{-1} + \varepsilon_7^{-1}) P_2 M M^T P_2] \bar{x}(t) \\
&+ e^T(t)(\varepsilon_4 K^T \bar{B}^T \bar{B} K + \varepsilon_5 K^T N_2^T N_2 K \\
&+ A_o P_2 + P_2 A_o) e(t) \qquad (46)
\end{aligned}
$$

which leads to the conditions and the proof has been completed. $\qquad\qquad\qquad\qquad\qquad \square$

3.4 Parametric optimization

Notice that all the free parameters can be adjusted without changing the stability of the system, which means that the optimization of the parameters can be done for various performance criteria. In this paper, we design the controller based on the nominal model then the optimization can be considered as the uncertainty compensation using the following performance criterion.

$$J = \int_0^t R_1 E\{e_y(t)\} + R_2 H\{e_y(t)\} dt \qquad (47)$$

where real positive R_1 and R_2 stand for the weights. $H\{\cdot\}$ denotes the entropy. To simplify the performance criterion, the entropy can further be replaced equivalently by information potential [20, 21].

Then the optimal free parameter f_i can be obtained by gradient descent once the eigenvalues λ_i^* are pre-specified.

$$f_{i,j+1} = f_{i,j} + \mu \left. \frac{dJ}{df_i} \right|_{f_i = f_{i,j}} \quad i = 1, \dots, m \qquad (48)$$

where j denotes the optimization searching iteration index. μ stands for the pre-specified step.

Once the control law λ_u is obtained, the control input $u(t)$ for system (1) can be calculated by reversing the transformation. In particular, Λ_U can be obtained by λ_u and we have

$$u(t) = \left[V_u \hat{\Lambda}_u(t) V_u^T \right]^{\frac{1}{2}} \xi(t) \qquad (49)$$

where $\xi(t)$ denotes the standard Gaussian white noise.

Remark 1. *The actual control law is non-linear though the transformed system model is linear.*

Remark 2. *Based on the dual principle, the observer gain matrix can be also obtained using proposed optimization approach. Meanwhile, the optimization operation can also be replaced by multi-objective optimization algorithms then the weights can be neglected.*

Remark 3. *Only a few elements of the parameter vectors f_i affect the control performance directly. Therefore, in order to determine the free parameters quickly, trial and error method can be used and the performance criterion can verify the manually selected parameters simply.*

3.5 Design procedure

The procedure of the proposed control strategy is summarized as follows:

Step1 Transfer the system to co-variance assignment model.

Step2 Setup the initial free parameters of the controller.

Step3 Use the numerical approach to optimize the performance criterion (47), by computing the mean, the entropy and gradient descent, then the optimal parameters are obtained.

Step4 Update the feedback gain matrix of the control law (19, 49), and verify it by the conditions of Lemma 2 to guarantee stability of the system, if the conditions hold, then go to next step, otherwise, return to Step 1.

Step5 Obtain the feedback gain matrix of the observer by dual principle and verify it by Lemma 3.

Step6 Verify the optimal parameters by Theorem 4, and if the conditions can be satisfied, then complete the procedure, otherwise, return to Step 2.

Step7 reverse the transformation and obtain the control input by (49).

4 A Numerical Simulation

To verify this new model and the control algorithms proposed in this paper, one numerical example is presented in this section.

The original model can be shown as below:

$$A = \begin{bmatrix} -0.5 & -0.3 \\ 0.1 & -0.2 \end{bmatrix}, B = \begin{bmatrix} 2 & 0.3 \\ 0.1 & 4 \end{bmatrix}$$
$$C = \begin{bmatrix} 0.7 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}, D = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}$$

The co-variance matrix of disturbance white noises is

$$\begin{bmatrix} 0.2 & 0.1 \\ 0.1 & 0.2 \end{bmatrix}$$

To assign the co-variance matrix, we choose the reference co-variance matrix as

$$\begin{bmatrix} 0.2 & 0.1 \\ 0.1 & 0.3 \end{bmatrix}$$

The parametric uncertainties are given as follows:

$$\Xi_1(t) = \Xi_2(t) = I \sin t$$

$$N_1 = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}, N_2 = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.1 \end{bmatrix}$$

Simply, we have

$$\lambda_r = \begin{bmatrix} 0.1382 \\ 0.3618 \end{bmatrix}$$

We can further obtain the eigen-vector of the variables as follows:

$$V_r = V_y = V_x = V_u = V_q = \begin{bmatrix} -0.8507 & 0.5257 \\ 0.5257 & 0.8507 \end{bmatrix}$$
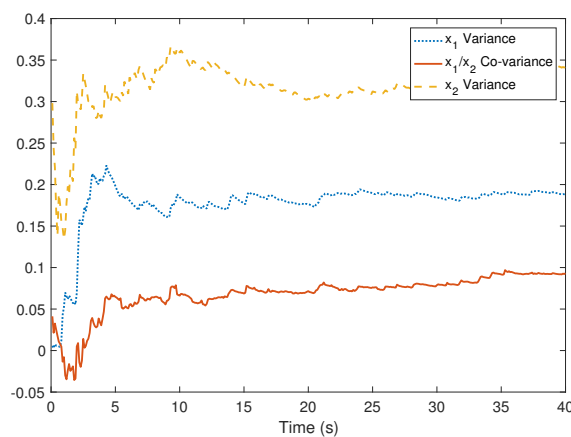
Figure 1. The measured co-variance and variances of the system outputs.

Following the presented control strategy, the simulation results have been shown by Fig. 1-4. Particularly, Fig. 1 indicates the co-variance and variances of the system outputs. Comparing to the reference co-variance matrix, the practical system outputs achieve the assignment with uncertainties. Meanwhile, Fig.

2 shows that the system outputs are stable if the transformed co-variance system design is stabilized while the states and control inputs of the transformed co-variance model have been given by Fig. 3 and Fig. 4. It has been shown that the eigenvalues of the co-variance matrix are tracking the reference eigenvalues of the reference co-variance matrix and the practical control input can be obtained by reversing the transformation using the values of the designed control input of the transformed co-variance model.
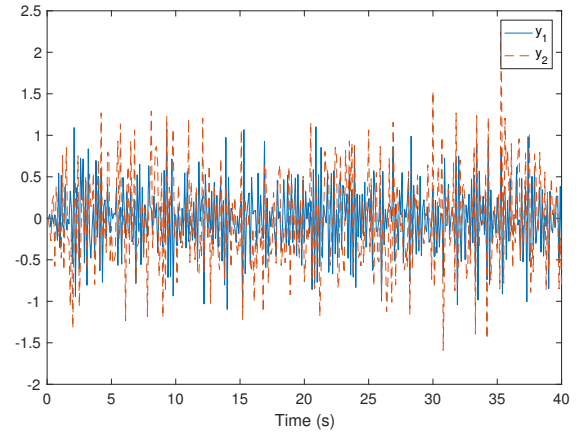
Figure 2. The outputs of the system.
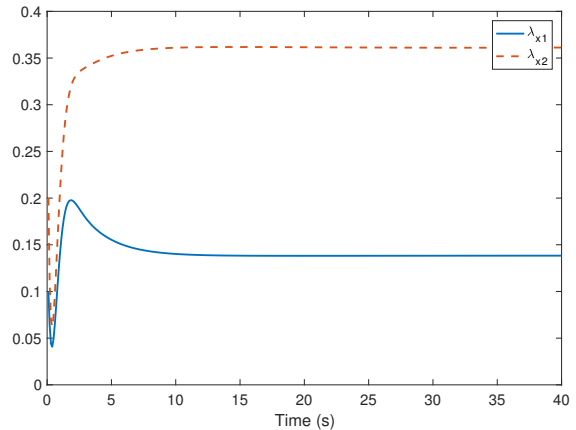
Figure 3. The state assignment of the transformed co-variance model.
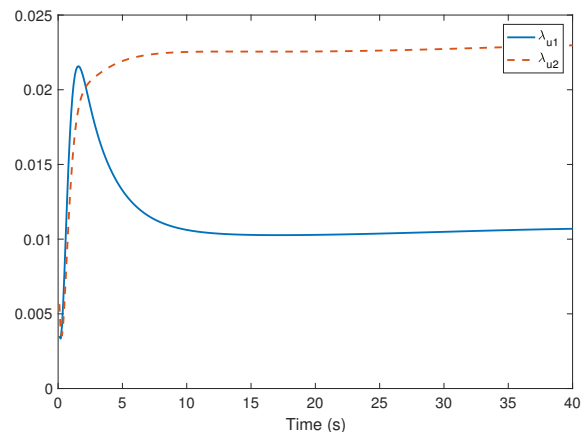
Figure 4. The designed control input of the transformed co-variance model.

5   Conclusion

This paper investigates the co-variance assignment strategy for a class of multi-variable stochastic uncertain systems. Combining the reduced-order co-variance model and parametric feedback, the control strategy is obtained by output feedback stabilization. In particular, the transformed model is given firstly with the extended co-variance assignment error. Then the linear observer is designed to estimated the extended state of the transformed model. After that, the output feedback is obtained via parametric optimization. Meanwhile the theoretical analysis is given to guarantee the robustness, stabilization and convergence of the closed-loop systems. Based on the results of the numerical simulation, the effectiveness of the presented control strategy has been verified while the control objectives have been achieved. Since the co-variance assignment is widely used in practical systems, such as paper-making process, the industrial applications using the presented control strategy will be the potential extension as a future work.

Conflict of Interest   The authors declare no conflict of interest.

References

[1] A. Hotz and R. E. Skelton, "Covariance control theory," *International Journal of Control*, vol. 46, no. 1, pp. 13–32, 1987.

[2] K. Liu, K. Li, and C. Zhang, "Constrained generalized predictive control of battery charging process based on a coupled thermoelectric model," *Journal of Power Sources*, vol. 347, pp. 145–158, 2017.

[3] M. Ren, J. Zhang, M. Jiang, M. Yu, and J. Xu, "Minimum ($hphi$)–entropy control for non-gaussian stochastic networked control systems and its application to a networked dc motor control system," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 1, pp. 406–411, 2015.

[4] L. Guo, L. Yin, H. Wang, and T. Chai, "Entropy optimization filtering for fault isolation of nonlinear non-gaussian stochastic systems," *IEEE Transactions on Automatic Control*, vol. 54, no. 4, pp. 804–810, 2009.

[5] L. Yao, J. Qin, H. Wang, and B. Jiang, "Design of new fault diagnosis and fault tolerant control scheme for non-gaussian singular stochastic distribution systems," *Automatica*, vol. 48, no. 9, pp. 2305–2313, 2012.

[6] Q. Zhang, J. Zhou, H. Wang, and T. Chai, "Minimized coupling in probability sense for a class of multivariate dynamic stochastic control systems," in *Decision and Control (CDC), 2015 IEEE 54th Annual Conference on*.   IEEE, 2015, pp. 1846–1851.

[7] ——, "Output feedback stabilization for a class of multivariable bilinear stochastic systems with stochastic coupling attenuation," *IEEE Transactions on Automatic Control*, vol. 62, no. 6, pp. 2936–2942, 2017.

[8] Q. Zhang and F. Sepulveda, "A statistical description of pairwise interaction between nerve fibres?" in *Neural Engineering (NER), 2017 8th International IEEE/EMBS Conference on*.   IEEE, 2017, pp. 194–198.

[9] ——, "A model study of the neural interaction via mutual coupling factor identification," in *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE*.   IEEE, 2017, pp. 3329–3332.

[10] K. Yasuda and R. E. Skelton, "Covariance controllers: A new parameterization of the class of all stabilizing controllers," in *American Control Conference, 1990*.   IEEE, 1990, pp. 824–829.

[11] K. M. Grigoriadis and R. E. Skelton, "Minimum-energy covariance controllers," *Automatica*, vol. 33, no. 4, pp. 569–578, 1997.

[12] H. Khaloozadeh and S. Baromand, "State covariance assignment problem," *IET control theory & applications*, vol. 4, no. 3, pp. 391–402, 2010.

[13] Q. Zhang, Z. Wang, and H. Wang, "Parametric covariance assignment using a reduced-order closed-form covariance model," *Systems Science & Control Engineering*, vol. 4, no. 1, pp. 78–86, 2016.

[14] Z. Wang, B. Huang, and H. Unbehauen, "Robust reliable control for a class of uncertain nonlinear state-delayed systems," *Automatica*, vol. 35, no. 5, pp. 955–963, 1999.

[15] T. Shen and K. Tamura, "Robust h/sub/spl infin//control of uncertain nonlinear system via state feedback," *IEEE Transactions on Automatic Control*, vol. 40, no. 4, pp. 766–768, 1995.

[16] Q. Zhang and X. Yin, "Observer-based parametric decoupling controller design for a class of multi-variable non-linear uncertain systems," *Systems Science & Control Engineering*, vol. 6, no. 1, pp. 258–267, 2018.

[17] G. Roppenecker, "On parametric state feedback design," *International Journal of Control*, vol. 43, no. 3, pp. 793–804, 1986.

[18] G. Roppenecker and J. O'reilly, "Parametric output feedback controller design," *Automatica*, vol. 25, no. 2, pp. 259–265, 1989.

[19] I. R. Petersen, "A stabilization algorithm for a class of uncertain linear systems," *Systems & Control Letters*, vol. 8, no. 4, pp. 351–357, 1987.

[20] Q. Zhang and A. Wang, "Decoupling control in statistical sense: minimised mutual information algorithm," *International Journal of Advanced Mechatronic Systems*, vol. 7, no. 2, pp. 61–70, 2016.

[21] Y. Zhou, Q. Zhang, H. Wang, P. Zhou, and T. Chai, "Ekf-based enhanced performance controller design for nonlinear stochastic systems," *IEEE Transactions on Automatic Control*, vol. 63, no. 4, pp. 1155–1162, 2018.

Using of the Flipped Classroom Learning with a Workshop Activity in Object-oriented Analysis and Design Course

Pakawan Pugsee[*]

Innovative Network and Software Engineering Technology Laboratory, Department of Mathematics and Computer Science, Faculty of Science, Chulalongkorn University, 10330, Thailand

A R T I C L E I N F O	A B S T R A C T

A lot of research reported that the active learning, especially the flipped classroom learning has higher performance than the lecture-based learning called the traditional learning. Therefore, this research is experimented on the object-oriented analysis and design course. The objectives are to try the flipped classroom learning with the workshop activity in order to analyze learners' behavior and to evaluate the effectiveness of the learning system. The aim of the research is to determine the suitability of the flipped classroom learning on the object-oriented analysis and design course, and to find the advantages and limitations of such learning method. The scope of this research is to test on the object-oriented analysis and design course, Faculty of Science, Chulalongkorn University. CourseVille developed by Chulalongkorn University is used as the learning management system. Moreover, the learning materials are Thai lecture videos, and English documents, such as English PowerPoint slides. Furthermore, the examples of activities in the classroom are Thai discussion and group activities, including the learner presentation. The research mythology is to collect data from learners composed of the behavior of the learners, learners' test scores, the result of the workshop activity and the output of term project. In addition, the collected data from the questionnaires about learner opinions with the classroom observation result are used for analyzing the learner attitude of the flipped classroom learning, and the learners' scores in the flipped classroom learning are compared to those in the traditional learning. In conclusion, the research results show that the 90 percentage of learners agree that the flipped classroom learning is suitable for the object-oriented analysis and design course, and it has higher performance that the traditional lecture-based learning system.

1. Introduction

At present, the learner-centered teaching method combined with the active learning is accepted widely that mainly focuses on learners, and can increase the effectiveness of learning and teaching system. One famous active learning method is the flipped classroom learning focused on the learners as the center and this system changes the traditional learning system like the lecture-based teaching from listening lecture contents in the classroom and doing homework or exercises after in-class time to the learners can study by themselves from the knowledge resource like out-of-school learning, such as teaching videos, online learning materials. On the other hand, the learners will practice in-class time to study,

and the instructors will provide assignments, problem questions or case studies for learners to have a brainstorm and exchange ideas for finding those answers under the guidance of the instructors in the classroom. In addition, the learners and the instructors can join learning activities in the classroom to improve their knowledge, abilities, and skills.

In the previous academic year, a traditional object-oriented analysis and design (OOAD) course was the lecture-based learning which instructors will lecture the lessons in the classroom. After that, the class spends the rest of time for the questions of the case study to practice object-oriented analysis and design skill. Therefore, there is too less time to discuss and share opinions among learners and instructors in the classroom for

[*]Pakawan Pugsee, Bangkok, Thailand, +66 2218 5170 & pakawan.p@chula.ac.th

implementation. Moreover, learners cannot make the decision or conclude the different answers for problems or questions about study contents, when learners do group work outside the class time. A few reasons are that some learners cannot apply the course content to real world problems, or they are not able to compromise their different ideas generated from the understanding content. Moreover, they cannot achieve the best solutions for the view of object-oriented analysis and design.

For these limitations of traditional OOAD course, the instructors should try the flipped classroom learning in this course to analyze results of using this new learning system for OOAD course. In addition, the workshop activity outside the classroom were included to practice the object-oriented analysis and design skill. The learners' behavior of preparing for class and evaluation scores was analyzed to find the results of using the flipped classroom learning with a workshop activity in OOAD course. Learners' opinion was also investigated for concluding that the flipped classroom learning system can encourage learners to learn and practice analysis and design skill more or not.

This research paper is an extension of work originally presented in conference [1]. Proposed research details will be described in the following sections. The flipped classroom learning system and research about it are explained in Section 2. The research methodology is demonstrated in Section 3. Next, the research results with some discussions and conclusions are described in Section 4 and Section 5, respectively. Then, the final section is the suggestions.

2. The Flipped Classroom Learning System

The flipped classroom [2-3] is a learning combination between online studying and teaching in the classroom. This learning system is different from the traditional lecture-based learning that the learners do some practices (e.g. class discussions, presentations or assignments) in class, after they listened to lectures before in-class time outside the classroom. So, the learners can study the courses' subjects by learning the online content lecture videos or other online media resources outside the classroom, and then they will do in-class activities (e.g. having discussions, finishing exercises) in the classroom to more understand the subject. It is also helpful for learners to develop their knowledge and skills to compromise on the interests the individual learners.

Moreover, online social community has currently be contained a lot of knowledge and useful information for learning. This leads to instructors and learners to have valuable resources to teach and study. Instructors have more powers to generate learning experiences for learners. Learners will also understand the subject concept deeply from these experiences, thus the knowledge is not restricted to that has been transferred from the instructors [4]. Furthermore, there is some research about the development of Hybrid learning systems like integrated learning and teaching in the classroom [5]. Online tools can assist learners to do assignments, for examples, finding the answers or the guidance for problem solving in class. In addition, discussions on topics related class lessons or case studies can be run as online groups and learners can search information from other resources for generating knowledge besides course materials.

Therefore, the flipped classroom [6-9] is new learning and teaching system that can improve the academic exam scores of learners and encourages learners to exchange ideas and discuss and about different course contents. In addition, the flipped classroom learning is flexible enough to integrate with other methods for achieving more effectiveness of learning and adapt to the class topics as much as possible. Nevertheless, there is the limitation of individual learners' time that they must spend to study the preparation before in-class time attending. For example, the research [10] applied the flipped classroom learning in computer programming course that learners had to study the definition and syntax of programming languages before class and get their hands-on practices for deeper learning taking place in the classrooms. It was found that the flipped classroom allows for more interaction during the face-to-face meetings, and for both instructors and learns to gain experience from the deeper learning in the classroom. However, the learners need to take time for preparation before class and practicing in class regularly and learn on their own.

The effectiveness of the flipped classroom results from the article [11] was discovered that this learning system can include more content than the traditional lecture because of design and problem-solving scores for the engineering course. Learners' scores in the flipped classroom are higher than learners' scores in the lecture-based class only, when the types of questions are separated into problem solving or no), although learners in both groups were similar in overall scores with insignificant differences. Moreover, this paper also analyzed the student perception which expressed that most of learners like to study in the flipped classroom learning system, including using teaching online videos, while over half of the learners in the lecture class want to practice more. In the same way of the research [10], the learners must take more time to learn by themselves, though doing case studies and projects for in-class time flipped classroom learning can help the learners to understand the content deeply.

Other research [12-14] found that the flipped classroom learning has the positive impacts on the educational system and has higher performance than the traditional learning in different issues. The paper [12] adopt the flipped classroom learning to the first-year vocational college students in Taiwan. Although there is no significant effect of the flipped classroom on the certification examination of learners, the face survey interviews of learners can be concluded that using a flipped classroom to enhance student engagement can promote active learning activities both inside and outside of the classroom. In addition, the finding guided that group activities are the main goal of the in-class portion of the flipped classroom, but activities of individual learner can be helpful for persons who need more individual reflective time to learn. The study [13] compared the effectiveness of flipped classroom learning with that of traditional lecture-based learning for selected topics in the college level technology course. The results of a learner survey revealed that the learners awesomely favored use of the flipped classroom teaching style because this system supports collaboration and hands-on activities learning for solving problem during the class time. Additionally, the flipped classroom approach made a statistically significant difference to the self-efficacy in a technology integration classroom and made learners' confidence to apply knowledge more.

Another research [14] used surveying and interviewing to identify the performance of the flipped classroom learning. The results discovered that the flipped classroom can engage the learners to study more in class, and can lead to more perseverance for learning. Learners of the flipped classroom also were successful for better in productivity and the more responsibility of the learner because of various learning activities, e.g. watching videos, doing exercises and homework. Consequently, there are interesting remark points that the flipped classroom learning, focusing on learner-centered can make the learners to be satisfied with learning and teaching style. Unfortunately, the learners, who cannot be adapted to the flipped classroom learning system, will be faced with trouble and their grades may be decreased as well.

The concept and practical cases of related research [2-14] has given all benefits of the flipped classroom learning. Meanwhile, a major limitation is that the learners must need time to study on their own outside the classroom and preparing before class. Additionally, collecting suggestions of the learners in OOAD course in the previous academic year explored the learners thought that they should achieve the learning objectives of the course and can be applied to real world cases, when they have time to train the critical thinking skill for analyzing and design more besides learning object-oriented design principles, including trying analysis and design thinking on practical issue problems. These are the reasons why the objectives of this proposed research extending the research [1] is to analyze the results of using the flipped classroom learning system in OOAD course. The research [1] studied on the learners' satisfaction of this learning system and discovered the advantages and limitations of the flipped classroom learning classroom. This proposed research analyzed and reported the learners' behavior with learner opinions in detail to confirm the improvement of the learning intention and critical thinking skills of learners. Moreover, the effectiveness of the flipped classroom learning was evaluated by a comparison of learners' total scores and term project scores in two different academic years, which OOAD course in the current academic year has applied the flipped classroom learning, while in the previous academic year used the traditional lecture-based learning. Furthermore, learners' opinions about a workshop activity and the suitability of the flipped classroom learning for OOAD courses were surveyed and made a conclusion.

3. Research Methodology

3.1. Research questions

The researcher tried out the flipped classroom learning system to analyze the learners' behavior and opinions to find out whether this system encourage learning intention and critical thinking skill or not, including suitability for OOAD course. Moreover, learners' scores of the flipped classroom learning were compared to those of the traditional lecture-based learning to evaluate the effectiveness of this learning system. Furthermore, problems or obstacles in both instructor and learner views were reported for the challenges of using the flipped classroom in learning and teaching management system.

3.2. Participants and Research Instruments

There was collecting data from 20 students enrolled in object-oriented analysis and design course which were three-year and four-year undergraduate students in Chulalongkorn University. The accumulated grade point average (GPAX) of learners is displayed in Table 1. The learners' behavior, both in-class time and out-of-class time were collected for observation and questionnaire survey. Consequently, the class attendance scores, in-class assignment scores, class preparing scores, individual academic exam scores and the term project scores of the student groups were focused to analyze the effectiveness of the flipped classroom learning system. Additionally, the answers of a questionnaire about the flipped classroom learning were summarized to gain information on learners' attitudes and to conclude the results of the flipped classroom learning system.

Table 1. The Accumulated grade point average of learners

Accumulated grade point average (GPAX)	Learner (persons)	Ratio of all learners
< 2.00	0	0%
2.00 - 2.50	3	15%
2.50 - 3.00	10	50%
3.00 - 3.50	4	20%
> 3.50	3	15%

3.3. Course Activities

The course activities of the flipped classroom learning system in this proposed research are explained in Figure 1. There were data collecting of the learners' behavior, both in-class time in the classroom by observations and out-of-class time outside the classroom by a survey. The examples of course activities were watching lecture videos, doing exercises, and answering the questions about the subjects. In the classroom, there were group activities about case studies or interesting topics related to subject contents, such as discussions and presentations. Moreover, the progress of term project continued to be processed and reviewed by another group learners and the instructors in the classroom (in-class time). Furthermore, the learners gave some presentations both pair and group work for subject contents or term project works in a few weeks.

In addition, a workshop activity about analysis and design in object-oriented view was added into this course in the current academic year. The workshop issue is a real-world problem about environmental problem-solving arranged outside the classroom. All learner groups had to gather requirements, analyze and design the system to solve environmental issues.

3.4. Course Materials

The teaching materials are Thai lecture videos and English documents, e.g. PowerPoint slides created by the instructors. In addition, some online multimedia, contents, or documents which related to the subject lessons can included in the course materials, such as YouTube videos, online documents on websites, pdf file, and other online knowledge resources. Table 2 shows the number of lecture videos and their length in each week.

3.5. Course Environments

CourseVille [12], which developed by Faculty of Engineering, Chulalongkorn University, is the Learning Management System (LMS) for this OOAD course.

Instructors prepare before class study time
- search multimedia resources and online documents related to class lessons
- create learning materials, e.g. lecture videos, PowerPoint slides
- assign out-of-class tasks for learners before attending classes and design in-class learning activities for learners
- publish class materials and tasks to learners

For out-of-class time, learners prepare themselves before class study time
- view lecture videos and study PowerPoint slides
- finish exercises or answer the questions after viewing lecture videos and studying PowerPoint slides
- give some opinions or topic issues about class lessons
- prepare presentation for the assigned topics about class lessons

For in-class time, instructors review the lesson in lecture videos and lecture the additional content lessons

For in-class time, instructors and learners do learning activities together
- summarize main points of lessons
- answer exercises and questions with discuss on the learner answers
- do more exercises using critical thinking and exchange the reasons of answers
- discuss topics related to class lessons, and case studies
- do working group activities about class lessons or the term project
- present information of tasks for the assigned topics or the term project

Figure 1. Course activities with the flipped classroom learning system.

Table 2 The features of lecture videos.

Week	Number of episodes	Length per episode (minutes)	Total length (minutes)
1	-	-	-
2	1	5.49	5.49
3	2	23.46 26.09	49.55
4	1	22.15	22.15
5	2	7.18 9.14	16.32
6	1	17.29	17.29
7	Midterm Exam		
8	-	-	-
9	2	14.16 13.09	27.25
10	1	10.15	10.15
11	1	11.46	11.46
12	-	-	-
13	1	12.05	12.05
14	-	-	-
15-16	Final Exam		

All course data on CourseVille are be stored in the cloud. Additionally, some activities can connect to Facebook to engage learners for acknowledging learning activities. For examples, CourseVille are logged by Facebook account, and links courses to the Facebook Group. General LMS services are also provided by CourseVille as follows:

- Multimedia course materials (e.g. lecture videos, video files, PowerPoint slides, pdf file documents, YouTube link) can be uploaded by instructors and can be downloaded or accessed by learners.

- News and announcements about the course can be announced to learners by instructors and can be gotten by learners.

- All teaching materials and tasks published by the instructors, can be acknowledged to the learners via the system which link to Facebook.

- Instructors and learners can raise questions and post topics to exchange their own ideas or opinions though the system.

- Instructors can assign tasks or exercises, while learners can do or answer online by the system. In addition, all learner submitted tasks can be scored online with the system by instructors

- Instructors and learners can check and record the class attendance of learners.

- Instructors can use some automatic classroom tools on the system, e.g. student randomizer, countdown timer, and Instant Q (lets instructors to create simple questions/polls to which students can respond suddenly).

4. Research Results

The proposed research experimented on the instructor and learners of OOAD course, and collected data from 20 students enrolled in this course. The results of using the flipped classroom learning in OOAD course are separated into five parts.

- The results of analysis and study on learners' behavior when using the flipped classroom learning were the analysis of learners' preparation before class, learners' class attendance, including the assignment and task submission.

- The results of effectiveness evaluation of using the flipped classroom learning were focused on academic exam scores by knowledge assessment, the term project scores which were compared to the behavior of learners, i.e. class attending, lecture video viewing and PowerPoint slide learning before class, and finishing all tasks through the LMS.

- The results of analysis and study on the learners' opinions about the flipped classroom learning and the workshop activity were summarized from the questionnaire about the flipped classroom learning with a workshop activity.

- The report of problems or obstacles in the flipped classroom learning and teaching management

- The summary of the suitability of the flipped classroom learning in OOAD course.

4.1. The Results of Analysis and Study on Learners' Behavior

In this research, interesting learners' behavior is separated into 3 main activities: preparing before class (watching lecture video and studying class documents), attending class, and submitting the assignments in class and after class). These data were collected from observations and behavior survey questions.

The learners' behavior of watching lecture videos and studying class documents before class are shown in Table 3. According to

learners' behavior survey results, while most of the learners (75% of all learners) watched almost of lecture videos (80%-100% of all videos) before class, nobody studied class documents before class more than 80% of all documents. Additionally, 45% and 35% of all learners studied the documents about 20%– 40% and 40%–60%, respectively. When the number of learners who watched lecture videos were compared to those who studied class documents, Figure 2 expressed that most of the learners from all 20 students watched the video 80-100% of all videos, but study class documents about 20-60% of all documents. The x-axis of bar graph represents the percentage of the number of watched lecture videos to the number of total lecture videos compared to the percentage of the number of class documents to the number of total class documents, and the y-axis denotes the number of students. Consequently, there were only 20% of all learners who usually study documents during video viewing. Therefore, most of the learners give precedence to watching the video more than study text documents before classes. In addition, the results concluded that most learners did not realize the importance of studying the document during the video viewing.

Table 3. Behavior of watching lecture videos and studying documents before class.

Behavior of watching lecture videos and studying documents before class		Learner (persons)	Ratio of all learners
% of the number of watched lecture videos to the number of total lecture videos	0% - 20%	-	0%
	20% - 40%	-	0%
	40% - 60%	2	10%
	60% - 80%	3	15%
	80% - 100%	15	75%
% of the number of class documents to the number of total class documents	0% - 20%	1	5%
	20% - 40%	9	45%
	40% - 60%	7	35%
	60% - 80%	3	15%
	80% - 100%	-	0%
% of the number of watched lecture videos with studying the document to the number of total lecture videos	0% - 20%	3	15%
	20% - 40%	2	10%
	40% - 60%	6	30%
	60% - 80%	5	25%
	80% - 100%	4	20%
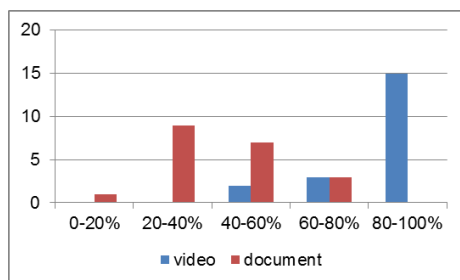
Figure 2. The number of learners watching lecture videos compare to those who studied class documents.

Moreover, the behavior analysis of learners' video watching as Table 4 found that approximately 50% of learners focused to view and listen intently to the whole part of the video. Additionally, 30% of learners view and listen through the end without serious study. Referring to the lecture videos during each week on Table 2, Table 5 also demonstrates the number of learners who did not watch lecture videos, and watched lecture videos through the end,

including viewing only some parts of the lecture videos. Unfortunately, there were few learners not informing viewing types during some weeks, so this information is "No record".

Table 4. The characteristics of watching lecture videos before class.

The characteristics of watching lecture videos	Learner (persons)	Ratio of all learners
only listen through the end	1	5%
only listen intently some part of the video	-	0%
view and listen through the end	6	30%
view and listen intently some part of the video	2	10%
view and listen intently the whole part of the video	11	55%

Table 5. The video viewing types of learners for each week.

Week	Not view	View the end	View not end	No record	Total (persons)
1	-	-	-	-	-
2	3	6	7	4	20
3	4	5	8	3	20
4	-	12	6	2	20
5	1	18	-	1	20
6	1	15	4	-	20
7			Midterm Exam		
8	-	-	-	-	-
9	3	8	6	3	20
10	-	17	1	2	20
11	2	14	1	3	20
12	-	-	-	-	-
13	-	18	-	2	20
14	-	-	-	-	-
15-16			Final Exam		

According to the features of the lecture videos on Table 1 and learner's viewing behavior on Table 5, during the week that had only a single video episode, learners would watch through the end more than the week that had more video episodes. In addition, learners would obviously watch less through the end during the week with the length of all videos is about or more than 20 minutes. Then, if there is more than one video episode during the week, more learners would watch through the end of the shorter length of the episode. In the same way, if the length of a video episode is about or less than 10 minutes, learners would watch through the end of every episode more.

Furthermore, the results from the questionnaire surveyed learners' time duration preparing for class found that the preparation time by learners 55% of all learners around about 12-24 hours before class as shown in Table 6, and nobody also do prepare more than 1 day. Another result discovered that the popular time to prepare for the class was during the late evening to midnight as displayed in Table 7. It was not surprising that most learners preparing themselves in the night before the morning classes.

Table 6. The preparation time before class.

Time duration	Learner (persons)	Ratio of all learners
before class less than 3 hours	4	20%
before class 3 - 12 hours	4	20%
before class 12 - 24 hours	11	55%
before class 1 day	1	5%
before class more than 1-2 days	-	0%

Table 7. The time of preparing for class.

Time	Learner (persons)	Ratio of all learners
0.00 - 4.00 a.m.	2	10%
4.00 - 8.00 a.m.	2	10%
8.00 - 12.00 a.m.	-	0%
0.00 - 4.00 p.m.	-	0%
4.00 - 8.00 p.m.	5	25%
8.00 - 12.00 p.m.	11	55%

Table 8. The class attendance of individual learners.

Learners no.	Not attend	In time	Late	No record	Total (weeks)
1	-	12	-	-	12
2	2	9	1	-	12
3	-	12	-	-	12
4	4	7	1	-	12
5	-	11	-	1	12
6	1	5	6	-	12
7	-	5	5	2	12
8	1	-	11	-	12
9	1	11	-	-	12
10	1	6	5	-	12
11	-	9	3	-	12
12	-	12	-	-	12
13	-	12	-	-	12
14	4	6	1	1	12
15	2	8	2	-	12
16	-	12	-	-	12
17	-	11	1	-	12
18	7	2	2	1	12
19	-	6	4	2	12
20	4	3	5	-	12
Total	11.25%	65%	19.58%	2.92%	100%

Table 9. The time of coming in class.

Time	Learner (persons)	Ratio of all learners
8.00 - 8.30 a.m.	4	20%
8.30 - 9.00 a.m.	8	40%
9.00 - 9.30 a.m.	6	30%
9.30 - 10.00 a.m.	1	5%
10.00 - 10.30 a.m.	-	0%
10.30 - 11.00 a.m.	1	5%

To observe learners' behavior in class, attendances in the classroom of individual learners were checked for 12 weeks displayed in Table 8. The OOAD course class time was 3 hours per week and started at 8.00 a.m. on Monday. The learners who came in the classroom before 9.00 a.m. were on time, otherwise they were late. The results found that 65% of all attendance checking are in time because most learners attended classes before 9.00 a.m. whether they missed the class. However, there was one learner always late for class, and another learner usually missed the class.

In addition, the learners' common time coming in class is illustrated in Table 9. Based on the time attendance, there are 60% of all learners attending classes before 9 a.m. and participating the class activities in time. Although there were a group of late participants (30% of all learners), they did not late more than a half an hour. Therefore, most learners realize that it is important to attend class and practice themselves in the flipped classroom learning.

By the way, all assignment submissions of learners recorded on CourseVille were composed of exercises or topic discussions in class and individual assignments after class. Table 10 presented the number of learners who submitted the assignments, and the rest of them who did not submit the assignments. The results show that almost all the learners submitted assigned works every piece, except few learners.

Table 10. Individual class attendance.

Assignment no.	Submission	Not submission	Total (persons)
1	17	3	20
2	13	7	20
3	13	7	20
4	19	1	20
5	20	-	20
6	20	-	20
7	17	3	20
8	16	4	20
9	20	-	20
10	16	4	20
Total	85.5%	14.5%	100%

In conclusion, all learners' behavior expressed that most learners gave corporation for managing the flipped classroom learning. They took their time to watch lecture videos for preparing before class and attended class regularly. Although few learners missed the class or did not study before class, they did their best when joined the group activities in class. In addition, learners' preparation for class by themselves, attending class to do in-class time activities, and doing all individual assignments, were critical thinking skill training. Therefore, the flipped classroom learning can stimulate learners' intention and encourage critical thinking skill of learners.

4.2. The Results of Effectiveness Evaluation of Using the Flipped Classroom Learning

The results of effectiveness evaluation of using the flipped classroom learning are separated into 2 main parts. The first part is the comparison among evaluation scores of course exams and activities that are the midterm exam, the final exam, the term project, the assignments, the class attendance, and the video viewing before class. The second part is the comparison between evaluation scores of learners in the current academic year with the flipped classroom learning, and in the previous academic year with traditional lecture-based learning.

There were comparisons between all scores of individual learners which consisted of the assessment scores of academic knowledge of each learner (the total scores of the midterm and the final academic exams), and the scores of each course activity in the flipped classroom learning system that were watching lecture video in preparation before in-class time, the class attendance for joining group activities, and individual assignment or task submission for in-class time). All scores of individual learners were inconsistent and difficult to find their relations shown in Figure 3. The x-axis denotes the individual student number (1 - 20) and the y-axis represents the scores of learners composed of the midterm and final exam scores (out of 45 points), the class attendance scores (out of 12 points), the individual assignment scores (out of 10 points), and the lecture video viewing (out of 9 points).
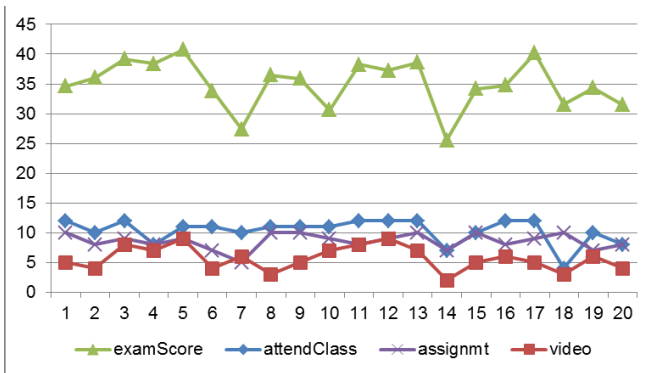
Figure 3. The exam scores of individual learners compare to each course activity.

Meanwhile, it was found that almost learners' exam scores were consistent with the summation scores of all course activities in the flipped classroom learning, when all course activity scores, which were watching lecture videos, attending class, and submitting assignments, were combined into the total scores of course activities.

The comparison result has been revealed in Figure 4, which the x-axis and y-axis mean the same parameters of Figure 3. The two-line graphs in Figure 4 demonstrated that learners, who did not prepare before class by watching lecture videos, did not join in-class learning activities by attending the class, and did not submit their works, would have low academic exam scores. And there were only a few learners who did all course activities, but still got low academic exam scores.
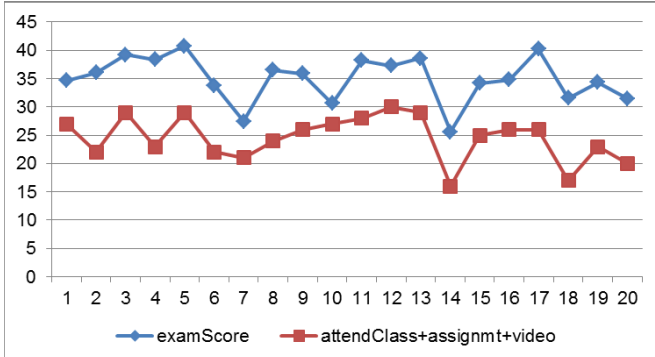
Figure 4. The academic exam scores of individual learners compared to the total scores of three course activites.

Moreover, the correlation analysis of academic exam scores with total scores of three course activities was performed to confirm the relationship between learners' academic exam scores and the total scores of learners' course activities. The analysis result displays that the academic exam scores are significantly related to the summation of all three course activities as in Figure 5. Therefore, the result has been concluded that the learner academic knowledge is not based on only one ability or one skill trained by only one activity but generated from the integrity of all learning and training activities. Knowledge improvement of individual learners requires studying and practicing all activities that are preparation before studying in-class time, paying attention to the discussion or group activities of in-class time, and training on course exercises or assignments.

Correlations			
		examScore	attendClass +assignmt+ video
examScore	Pearson Correlation	1	.731**
	Sig. (2-tailed)		.000
	N	20	20
attendClass+assignmt +video	Pearson Correlation	.731**	1
	Sig. (2-tailed)	.000	
	N	20	20
**. Correlation is significant at the 0.01 level (2-tailed).			

Figure 5. The correlation between the exam scores of individual learners and the total scores of three course activities.

Furthermore, the flipped classroom learning increases time in class for discussing related topics, practicing on analysis and design skills, and doing the term project in the classroom under the guidance and management of the instructors. Therefore, the comparison results as in Figure 6 expresses that the term project scores of individual learners (out of 30 points) were related to their class attendance scores, except a few learners having the good term project score without attending class regularly. One cause of these scores is the term project was a group work, so the other members of the group could operate and review the work of absent members. The x-axis and y-axis of Figure 6 were the same meaning as Figure 3 and Figure 4. The correlation analysis of the term project and the class attendance was also displayed in Figure 7.
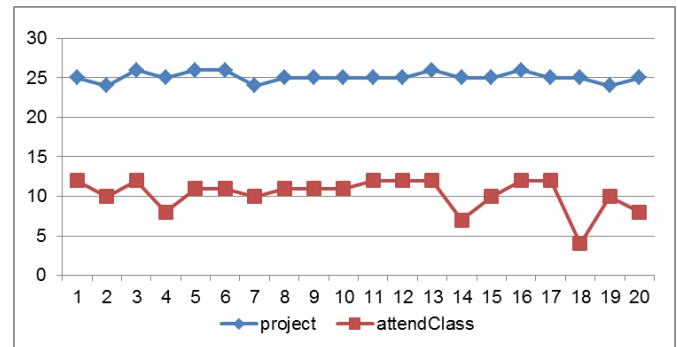
Figure 6. The term project score of individual learners compare to the class attendance

Correlations				
			project	attendClass
Spearman's rho	project	Correlation Coefficient	1.000	.469*
		Sig. (2-tailed)		.037
		N	20	20
	attendClass	Correlation Coefficient	.469*	1.000
		Sig. (2-tailed)	.037	
		N	20	20
*. Correlation is significant at the 0.05 level (2-tailed).				

Figure 7. The correlation between the term project scores of individual learners and the class attendance

To evaluate the performance of the flipped classroom learning, the total scores of grading and the term project scores of learners in the current academic year were compared to those of learners in the previous academic year by the independent t-test. These scores

are also used to measure critical thinking skills of learners in this research. Figure 8 shows the term project scores and the total scores of the current academic year with the flipped classroom learning (the table at the top of the figure) and the previous academic year with the traditional learning (the table at the bottom of the figure).

Descriptive Statistics					
Current	N	Minimum	Maximum	Mean	Std. Deviation
project	20	24.00	26.00	25.1000	.64072
TotalScores	20	68.83	95.10	85.2350	7.29223

Descriptive Statistics					
Previous	N	Minimum	Maximum	Mean	Std. Deviation
project	10	22.00	26.00	24.2000	1.39841
TotalScores	10	73.07	88.07	80.1610	5.80154

Figure 8. the term project scores and the total scores of both academic years.

The average total scores of both academic years and the independent t-test results are demonstrated in Figure 9. The result found that the average total scores of learners in the current academic year were higher than those of learners in the previous academic year.

Group Statistics					
year		N	Mean	Std. Deviation	Std. Error Mean
Total Scores	current	20	85.2350	7.29223	1.63059
	previous	10	80.1610	5.80154	1.83461

Independent Samples Test								
		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Total Scores	Equal variances assumed	.264	.612	1.913	28	.066	5.07400	2.65244

Figure 9. The t-test results of the total scores of students in current academic year and the previous academic year

According to Figure 9, the p-value of Levene's test was greater than 0.05 (Sig. $0.612 > \alpha$), that meant the variances of two learner groups (students in the current and the previous academic years) are equal. Additionally, the p-value of t-test was less than 0.05 (Sig. $0.066/2 < \alpha$), then there was a mean difference between total scores of learners in the current academic year and the previous academic year. Therefore, the average total scores of learners in the flipped classroom learning were significantly higher than those in the traditional learning.

Another result was that the average term project scores of learners in the current academic year were higher than those of learners in the previous academic year. The average term project scores of both academic years and the independent t-test results are explained in Figure 10. Like the total scores, the variances of term project scores were equal across the two learner groups because the p-value of Levene's test was greater than 0.05 (Sig. $0.001 > \alpha$). Then, the t-test result displayed that the p-value of t-test is less than 0.05 (Sig. $0.021/2 < \alpha$), thus there was a mean difference between term project scores of students in the current academic year and

the previous academic year. That meant the average term project scores of learners in the flipped classroom learning were a little bit significantly higher than those in the traditional lecture-based learning.

Group Statistics					
year		N	Mean	Std. Deviation	Std. Error Mean
project	current	20	25.1000	.64072	.14327
	previous	10	24.2000	1.39841	.44222

Independent Samples Test								
		for Equality of		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
project	Equal variances assumed	15.108	.001	2.440	28	.021	.90000	.36888

Figure 10. The t-test results of the term project scores of students in current academic year and the previous academic year

Although there was not much difference in the average term project scores, the effectiveness of the flipped classroom learning has been higher than the traditional learning. The reason is that the number of students in the current academic year were more than those in the previous academic years and all learner groups did the term project with the good quality. Therefore, the result can be concluded that the flipped classroom learning helped to encourage the critical thinking skill for the learners.

4.3. The Results of Analysis and Study on the Learners' Opinions

The learners' opinions were collected by the questionnaire about the features of the lecture video and all course activities in the flipped classroom learning, including a workshop activity. All interesting course activities are watching lecture videos and studying lesson documents before class (out-of-class time), working in groups and discussing together on case studies and related topics about lessons, answering instructors' questions

following to discuss on learners' answers (in-class time), and doing term project with the discussion (in-class time). In addition to the flipped classroom activities, the learner attitudes about this learning system suitability for the OOAD course were summarized from the survey answers.

The opinions about the features of the teaching videos were that 60% of learners (12 students) commented the lecture video is able to have more than one episode per week, while the rest of them (40 % of learners or 8 students) supposed that it should have one episode per week only. When analyzing the length of the video, 85% of learners (17 students) thought that it should be no longer than 20 minutes or less than 10 minutes as detailed in Table 11. In addition, a comparison between the lecture video features and the number of learners who watch through the end is demonstrated in Table 12.

According to Table 12, the lecture video features were analyzed by using the number of learners who watching the video through the end. The opinion of the learners discovered that most of the learners would watch through intentionally, if the length of

one lecture video is about 10 minutes. Meanwhile, learners would noticeably watch less through the end, if the length of lecture videos is more than 20 minutes. The analysis results of the number of episodes and the length of the video per one episode are also consistent with the number of learners who watching the video until the end of each week.

Table 11. The learner opinion about the appropriate length for one episode of lecture video.

The length for one episode of video	Learner (persons)	Ratio of all learners
less than 10 min.	8	40%
10 - 20 min.	9	45%
20 - 30 min.	2	10%
more than 30 min.	1	5%

Table 12. Comparison between the lecture video features and the number of learners who watch through the end.

Week	Number of episodes	Total length of lecture videos (minutes)	Number of learners viewing through the end (persons)
1	-	-	-
2	1	5.49	6
3	2	49.55	5
4	1	22.15	12
5	2	16.32	18
6	1	17.29	15
7	Midterm Exam		
8	-	-	-
9	2	27.25	8
10	1	10.15	17
11	1	11.46	14
12	-	-	-
13	1	12.05	18
14	-	-	-
15-16	Final Exam		

To ensure the performance of the flipped classroom learning, there were seven survey questions about the impacts of the learning course activities in the flipped classroom learning for assessing learners' opinions.

The effectiveness evaluation results are the average attitude scores from all learners' opinions that is classified in five levels:

1 = A disagree   2 = B quite disagree   3 = C partly agree
4 = D quite agree   5 = E agree

Figure 11 shows the number of learners who commented on each survey question and the conclusions of the learners' opinions that agree or disagree on each topic are displayed as follows:

- (1) Watching lecture videos before in-class time helps learners to understand the subjects more easily. Mean score = 4.25

- (2) Studying lesson documents before in-class time helps learners to understand the subjects more easily. Mean score = 3.90

- (3) Case study and group discussion in class make learners to understand the subjects deeply. Mean score = 4.00

- (4) In-class working group and discussing together make learners to understand the subjects deeply. Mean score = 3.85
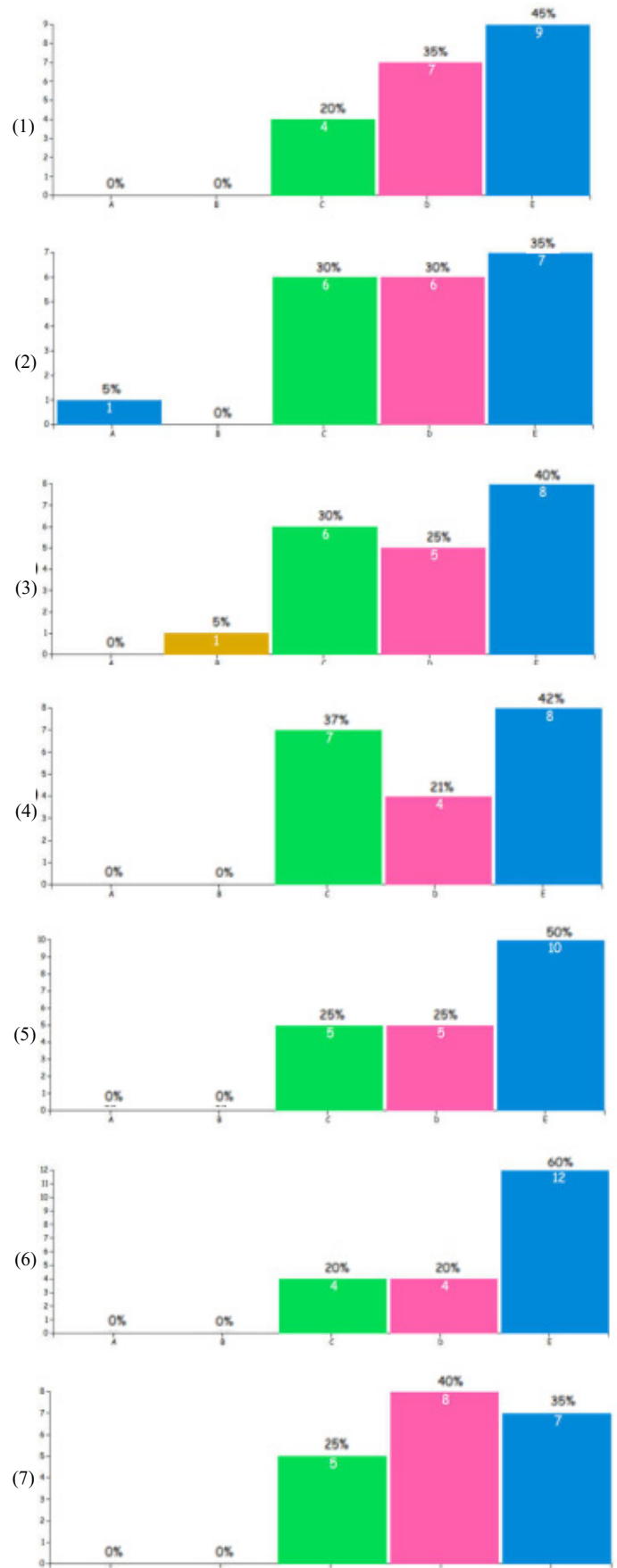
Figure 11. The learner opinions about the flipped classroom learning activities

- (5) Doing term project in class study time with discussion together helps learners to develop more quality of the term project. Mean score = 4.25

- (6) Answering questions with discussion on the learners' answers help the learner to improve their knowledge. Mean score = 4.40

- (7) The flipped classroom learning helps to develop critical thinking, analysis and design skills more. Mean score = 4.10

According to the bar graphs in Figure 11, most of the learners agree that study preparation before in-class time (out-of-class time) and in-class time activities in the flipped classroom learning make learners to understand the lesson contents easier and deeper, develop a better quality of term projects, and improve their own knowledge. In conclusion, most of the learners are agreed and quit agree that the flipped classroom learning helps to develop critical thinking skill, including analysis and designing skills better than the traditional lecture-based learning.

By the way, the opinion assessment of the learners about the object-oriented analysis and design workshop was evaluated by the questionnaire on five levels like effectiveness assessment. The average scores of learners' opinion for each topic are shown as follows:

- (1) You have learned how to analyze with critical thinking the overview of preserving the underwater environment. Mean score = 3.84

- (2) You have learned to identify the problem of coral damage. Mean score = 4.11

- (3) You understand the causes of problem in coral damage. Mean score = 4.26

- (4) You have gotten experiences in interviewing to gather requirements for problem solving about coral damage. Mean score = 3.84

- (5) You have gotten experiences in the analysis and design for the real-world problem. Mean score = 4.16

- (6) You have practiced on real world problem solving. Mean score = 4.05

- (7) You understand the subject contents more by joining this workshop activity. Mean score =3.95

- (8) You have trained on group working to create good quality works. Mean score =4.16

Referring to the average scores of each survey question of the workshop activity, most of the learners agreed that learners have learned how to analyze with critical thinking and identify the problem, including to find out and understand the causes of the problem. Moreover, the learners had gotten experiences in gathering requirements, analysis and design, and solving the real-world case study. Furthermore, most learners can understand the subjects deeply by doing workshop activities and practice themselves in working groups for creating good quality works.

4.4. The Report of Problems or Obstacles

There were two different question types of data collection for problems and obstacles of the flipped classroom learning. The first

question with answer choices were asked for the most important problem or obstacle in the individual learners' attitude. And the second question with the short answers was independent writing for the question "What are your difficulties or obstacles in the flipped classroom learning?".

According to the selected answers by learners, 55% of the learners (11 students) thought that the most important problem or obstacle in the flipped classroom learning system is habit and self-responsibility of individual learners. Meanwhile, some learners (4 students) are not able to understand the content in the lecture video or the PowerPoint slide that are the most important problem for them in this learning system. However, there are two students told that their most important problem is no time to study the preparation before class, and only one student who thought that unfriendly LMS, or no time to attend class, or conflicts on discussions or group works is the most important problem.

For the independent writing answers, there were five students left empty and three students wrote "no problem". The comments of the rest students were grouped as follows:

- Sometimes there is no time for study preparation before in-class time, so it caused the less understanding and not learn much.

- The instructors upload some lecture videos before class study time less than one week, if we are busy, we cannot watch them in advance.

- Cannot understand all the content, and need to study more or ask more in class study time.

- Cannot get the answers to any questions.

- Need time to adjust to this learning system because of the unfamiliar learning system.

- Viewing the lecture video is very troublesome because of the too small screen.

- Get bored with watching the lecture videos through the end because some lecture videos are too long.

- Need the instructors to upload and publish all lecture videos at the beginning course period because we can totally learn when we have a spare time.

- Problems with the internet connection lead to other problems. For example, sometimes cannot display the lecture videos, cannot access to the LMS, and miss the class announcements

Other problems or obstacles of the flipped classroom learning in the instructor view are displayed as follows:

- The learners lack time to watch the video or study the lesson document before in-class time, and if there are no scores for self-study preparing activities, the learners may not concentrate on viewing the lecture video or learning class documents in advance.

- There is insufficient time for the instructors to prepare learning materials, such as lecture videos and PowerPoint slides for the whole subject contents of the course in advance, so sometimes they has been published before classes less than one week.

- There is the variant of the lesson contents so some lessons in some weeks cannot be understood by viewing the lecture videos with a length of shorter than 20 minutes.

4.5. The Summary of Suitability of the Flipped Classroom Learning

The assessment results of the survey answers from the questionnaire were analyzed and the mean score of the appropriateness of the flipped classroom learning system was calculated to summarize the suitability of the flipped classroom learning for this course. The rating given by the learners on five levels as below:

1 = A inappropriate   2 = B quite inappropriate
3 = C moderately appropriate   4 = D quite appropriate
5 = E appropriate

The mean score for the appropriateness of the flipped classroom learning was 4.40. The result means almost of the learners believed that the flipped classroom learning system has appropriated for this course, as detailed in Figure 12.
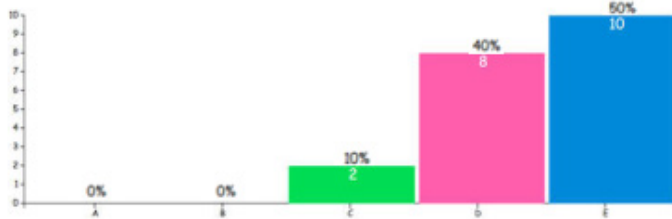
Figure 12. The number of learners who answered the suitability of the flipped classroom learning.

According to Figure 12, there were 50% of learners (10 students) and 40% of learners (8 students) thought that the flipped classroom learning has been appropriate and has been quite appropriate to the OOAD course. Therefore, these results show that the flipped classroom learning is suitable for the OOAD course with the aggregation of 90 percent of learners.

Moreover, the reasons which support their attitudes are explained as listed below:

- There can be the diversity in the thinking process in this course to define the functionality of the system and design the system. Therefore, explaining, dialoguing, questioning and answering among learners, including between learners and instructors can help in learning as well. The discussion between learners and instructors helps them to have coincident ideas.

- The content of this course is quite many subjects. The flipped classroom learning helps learners to understand the lesson before class, have time to exchange ideas and work in groups more in class.

- This course focuses on the practical so learners have to think and practice by themselves before they consult or discuss with others. In addition, learners can learn, think, ask the question and find the answer by themselves to use their own ability for learning.

- This course is not only focused on academic content, but also emphasize on practical application. It should have activities

in the classroom instead of teaching the content which learners can read from documents or find information from outside resources. It also needs to discuss and practice to understand the contents because the subjects cannot be understood by only reading.

- The flipped classroom learning system is very beneficial. The learners are enabled to practice critical thinking and self-study skills. Watching the video before class makes an easier preparing and more intention. Discussion motivates the learners to be active, not sleepy in the classroom.

- If learners cannot attend classes or some contents may be difficult, they are able to study some subjects of the video viewing outside classes which may make more understanding. The flipped classroom learning system helps to save the time of teaching in lecture class duration, so the interesting subjects can be learned more.

Furthermore, the advantages of the flipped classroom learning that the learners gave opinions on their own side are listed below:

- Learners can study the lessons before class that make easier understanding when they come to learn in class. If they don't understand the lessons, or have any questions, they can ask the instructor in the classroom. Beside previous advantages, they pay more attention in class because they make understanding the contents before class.

- Learners have selves' critical thinking before class, so they can remember and understand the contents more, don't have to read much for exams.

- Learners have practiced critical thinking and done practical works under the care and guidance of the instructor.

- Learners can understand the lessons more and work better on the project because the traditional teaching system spent the whole period for lecturing, no chance to ask questions about the project, and no advisor to check whether their works are right or not.

- Learners can have the lecture video to watch after class and prepare for tests that make reviewing the lessons easily. Additionally, in the case of missing the class, learners still study from watching the video and gain some knowledge.

- Learners can discuss and exchange ideas in the class without wasting the teaching time.

- A non-stressful environment for learning in class.

In conclusion, although there are problems or obstacles in the flipped classroom learning because of limitation out-of-class time for learners preparing a lesson before class and for instructors providing teaching materials, including specifying in-class time activities, there are more benefits to use this new learning system in the OOAD course with many supported reasons by learners' opinions. Therefore, the flipped classroom learning has been suitable for the OOAD course in this research results.

5. Conclusions

Using of the flipped classroom learning in the OOAD course was studied to analyze the results of using this learning system. This research focused on the behavior of learners, learners' scores, and learner opinions to find out the appropriateness of the flipped classroom learning in this course. The results of learners' behavior

analysis were found that most learners are interested in the flipped classroom learning activities and enjoy cooperating with the instructors. In addition, the performance evaluation of the flipped classroom learning for the OOAD course can be concluded that this learning system has been better than the traditional learning system because the average total scores of grading and the average term project scores of learners in the current academic year with the flipped classroom learning were significantly higher than those in the previous academic year with the traditional learning. Therefore, using the flipped classroom learning can increase the learning outcome of the learners and can be useful for learners in the learning intention and critical thinking skill development.

The summary of learners' opinions on the flipped classroom learning with a workshop activity was uncovered that the majority of the learners concurred that the flipped classroom learning activities composed by viewing the lecture video and studying the text documents before in-class time, helps them to understand the class contents effortlessly. Moreover, group discussions of case studies, and doing a term project under the guidance of the instructors in class study time can help understand class contents well and create the better term project. Furthermore, the solutions to the assignments or the answers of exercises which are explained supporting reasons in the classroom make the learners develop their own knowledge as well.

In conclusion, most of the learners agreed that the flipped classroom learning system can help improve critical thinking skill and object-oriented analysis and design skill. The research results also pointed out that the flipped classroom learning system helps the learners to pay more attention because of preparation before class. Meanwhile, the critical thinking skills were cultivated by discussions, group activities in-class time, which can ask questions and get advice from the instructor immediately when in trouble, so it encourages learners to practice and apply the knowledge better. In addition, there were 90% of learners thought the flipped classroom learning is suitable for this course. Therefore, it can be concluded that the flipped classroom learning is suitable for the OOAD course.

6. Suggestions

In spite of the fact that the flipped classroom learning is appropriate to the OOAD course, this learning system requires readiness in both of the instructors and the learners. They should prepare for adjusting themselves to the active learning style with the useful learning management system. Instructors should be able to generate suitable learning materials with valuable knowledge, and to transfer that knowledge to learners. Moreover, learners must have abilities in self-study learning for learning and understanding course content. In addition, the learner need to concentrate on study preparation before in-class time and doing learning group in-class activities to get their knowledge and practice their skills. Furthermore, both instructors and learners should be able to create a friendly atmosphere in the classroom.

In fact, according to problems and obstacles of the flipped classroom learning in views of both instructors and learners, this learning system cannot completely be flipped the classroom style. If instructors have more time to create learning materials with the coordination between different courses in the same field, such as software development, this may increase the achievement of the flipped classroom learning because learners can arrange appropriate times for self-study preparation and cooperating group activities, while the instructor can share available learning resources in order to save time for searching and generating a few duplicate subjects in different courses.

Beside available time of learners and instructors, materials in the course which are prepared or provided by instructors cannot guarantee that all learners are able to study and understand by themselves. The reasons are that the content of each lesson has the different difficulty, and there still are differences among self-learning skill, intelligence, ability and perseverance of each learner. Additionally, the expertise of individual instructors to transfer knowledge in each lesson is different. If there are quality and standard materials, may cause the flipped classroom learning system achievement even better.

Acknowledgment

References

[1] P. Pugsee, "Effects of using flipped classroom learning in object-oriented analysis and design course" in 10th International Conference on Ubi-media Computing and Workshops (Ubi-Media), Pattaya, Thailand, 2017. https://doi.org/10.1109/UMEDIA.2017.8074130

[2] C. Brinton, M. Chiang, "Social Learning Networks: A brief survey" in 48th Annual Conference on Information Science and Systems (CISS), Princeton, NJ, USA, 2014. https://doi.org/10.1109/CISS.2014. 6814139

[3] A. Wallace, "Social learning platforms and the flipped classroom" Int. J. Inf. Educ. Technol., 4(4), 293-296, 2014. https://doi.org/10.7763/IJIET.2014.V4.416

[4] T. Bristol, "Flipping the classroom" Teach. Learn. Nurs., 9(1), 43-46, 2014. https://doi.org/10.1016/j.teln.2013.11.002

[5] Q. Dang, D.D. Gajski, "Bringing in-class online - A hybrid solution" in 4th Interdisciplinary Engineering Design Education Conference (IEDEC), Santa Clara, CA, USA, 2014. https://doi.org/10.1109/ IEDEC.2014.6784674

[6] A. Amresh, A.R. Carberry, J. Femiani, "Evaluating the effectiveness of flipped classrooms for teaching CS1" in IEEE Frontiers in Education Conference (FIE), Oklahoma City, OK, USA, 2013. https://doi.org/10.1109/FIE.2013.6684923

[7] J.M.M. Ferreira, "Flipped classrooms: From concept to reality using Google Apps" in 11th International Conference on Remote Engineering and Virtual Instrumentation (REV), Porto, Portugal, 2014. https://doi.org/10.1109/REV.2014.6784256

[8] M. L. Carrió-Pastor, H. Skorczynska, "Collaborative learning and communication technologies in teaching business english" Procedia Soc. Behav. Sci., 178, 32-37, 2015. https://doi.org/10.1016/j.sbspro. 2015.03.142

[9] K. T. Wissman, K. A. Rawson, "Why does collaborative retrieval improve memory? Enhanced relational and item-specific processing" J. Mem. Lang, 84, 75-87, 2015. https://doi.org/10.1016/j.jml. 2015.05.003

[10] C. P. Rosiene, J. A. Rosiene, "Flipping a programming course: The good, the bad, and the ugly" in IEEE Frontiers in Education Conference (FIE), El Paso, TX, USA, 2015. https://doi.org/10.1109/FIE.2015. 7344151

[11] G.S. Mason, T.R. Shuman, K.E. Cook, "Comparing the effectiveness of an inverted classroom to a traditional classroom in an upper-division engineering course" IEEE Trans. Educ. 56(4), 430-435, 2013. https://doi.org/10.1109/TE.2013.2249066

[12] Y. Chen, L. Chen, "Effects of the flipped classroom model on student performance for vocational college students" in International Conference on Educational Innovation through Technology (EITT), Tainan, Taiwan, 2016. https://doi.org/10.1109/EITT.2016.30

[13] M. Khan, M. Ibrahim, "Flipped classroom in technology courses - impact on personal efficacy and perception based on learning style preferences" in IEEE Integrated STEM Education Conference (ISEC), Princeton, NJ, USA, 2017. https://doi.org/10.1109/ISECon.2017. 7910229

[14] A. Amiri, H. Ahrari; Z. A. Saffar; V. Akre, "The effects of classroom flip on the student learning experience: An investigative study in UAE classrooms" in International Conference on Current Trends in Information Technology (CTIT), Dubai, United Arab Emirates, 2013. https://doi.org/10.1109/CTIT.2013.6749480

[15] CourseVille. "The Social Learning Management System of Chulalongkorn University". https://www.mycourseville.com/

ASTES

A Novel MICS Receiver with FSK Dual Band Demodulator

Mouna Bettaieb*, Saif Benali, Ghazi Bouzid, Hatem Trabelsi

METS research Group, National Engineering School of Sfax, University of Sfax, BP 1173, 3038 Sfax, Tunisia.

A B S T R A C T

A low-complexity dual-band chirp FSK, direct conversion receiver is described in this paper. The receiver is dedicated to be used in the transceiver unit of a medical implantable wireless sensor. The system uses the RF band between 402 and 405 MHz. Two sub-bands frequencies employing chirped pulses are assigned for both binary information. The novelty of this work is the use of a Binary FSK LFM modulator, a direct conversion receiver and a simple and low power non-coherent BFSK envelope detection demodulator. Receiver performances are evaluated for all the input power dynamic range. Receiver front-end parameters are optimized using harmonic balance simulation. In order to improve receiver sensitivity, a low pass filter with controllable bandwidth between 40 and 300 KHz is used to avoid in-band interference. The receiver is able to achieve a noise figure of 5.5 dB, a receiver sensitivity of -93 dBm and a maximum data rate of 100Kbps. The simulated IIP3 and P_{-1dB} are 12.6 dBm and 22.1 dBm respectively. A simple non coherent binary dual band FSK demodulator was used which is based on an envelope detector, integrate & dump, a sampling & hold and a liming circuit. The receiver was co-simulated with the dual band non coherent demodulator. The proposed receiver has a sensitivity of -93 dBm and a BER less than 10^{-3}.

1. Introduction

With the proliferation of portable or implantable devices in wireless healthcare sensor network, energy-efficient radios have become an active area of research.

This paper is an extension of work originally presented in International Conference on Engineering & MIS (ICEMIS 2017) [1].

Today, there is a great demand for low power transceivers dedicated to implanted sensors, thanks to the development of new technologies in microelectronics and the progress of sensor implantation techniques in the human body [2]. This system measures certain physiological parameters of people suffering from chronic diseases and communicates useful information to an information collection center in order to warn medical professionals in case of danger [3-5].

The implanted sensor node must provide continuous operation under battery power for at least 5 years.

WMTS (Wireless Medical Telemetry Services) and MICS (Medical Implantable Communications Services) are two standards used for medical electronic devices. Non implantable devices uses WMTS whereas MICS is dedicated to sensor networks implanted in the human body.

According to the MICS regulation from the Federal Communications Committee (FCC) its frequency band is between 402 MHz and 405 MHz which is divided into 10 channels. Each channel occupies a bandwidth of 300 KHz. The wireless communication range is limited to 2 m and the maximum transmitted power is set at -16 dBm [6]. These frequencies were selected for the following reasons: reasonable antenna size, good propagation characteristics for medical implants in the human body and worldwide availability [4].

In addition, MICS signals are safe for the human body thanks to their low power level. Moreover, this introduces low interference with other users. With this technology we can reach reduced complexity, low power and low cost transceiver.

Figure 1 introduces the block diagram of a typical sensor node [7-8]. Physiological signals are very weak and coupled with noise like body pressure or the amount of glucose in the blood of the patient. Thus, amplification and filtering process are required to increase the signal level and reject the noise. A multiplexer is used to switch between different sensors. The analog signal conversion into digital is ensured by an ADC. The microcontroller processes

*Mouna Bettaieb, National Engineering School of Sfax, BP 1173, 3038 Sfax, Tunisia, bettaiebmouna3@gmail.com

the signals from the sensors unit and executes the communication protocol. The MICS transceiver communicates the formatted data across a wireless link with a portable personal healthcare centre for testing and diagnosis by doctors.

A sensor node can perform communication in both directions. In transmission mode the sensor node sends all the useful sensors data and information regarding its operating status. In receive mode, it receives node configuration, wake up or standby information. It should also be noted that power consumption of a sensor node depends on the duty cycle, the time during which the system is active over a given period.

The use of CMOS technology in healthcare implanted devices is justified since the principal challenge is the minimization of the energy consumption in the sensor nodes in order to integrate the digital and the analog parts together on the same chip. That's why the architecture of the RF wireless sensors, modulation scheme and parameters optimization of different blocks of the transceiver was subjects of many studies and researches as part of the low power consumption challenge [9].
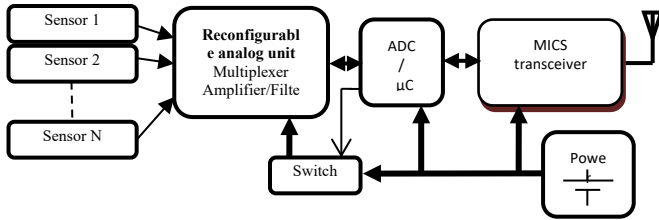
Figure 1. A typical sensor node.

Linear Frequency Modulation (LFM) is widely used in radars, but it has not been yet used for implanted sensor networks in the human body using the MICS band.

The purpose of this paper is to design a simple and low power, 402-405MHz-MICS receiver chain based on the LFM technique with a non-coherent BFSK envelope detection demodulator dedicated for implantable sensor node.

The description of Linear Frequency Modulation technique and transceiver architecture is presented in section 2. In section 3 the receiver design and simulations results will be presented. A co-simulation of the analog receiver with the non coherent dual band FSK demodulator will be shown in section 4. Section 5 presents blocks parameters and performance of the receiver. Section 6 concludes this paper.

2. Chirp FSK MICS modulation

2.1. Mathematical background associated with LFM

Linear Frequency Modulation (LFM) signal or linear chirp modulation signal is considered as a form of spread spectrum technique for wireless communications. Since 1940 chirp modulation has been widely used in radars communication systems because it has the property of pulse compression [9].

LFM has been chosen because it presents performances in terms of distortion, rejection of interference and allows a low

power implementation [10]. Linear Frequency Modulation (LFM) signal or linear chirp signal can be representation by (1):

$$S(t) = A.\cos(2\pi f_0.t + \varphi(t)) \qquad 0 \leq t \leq T \qquad (1)$$

Where f_0 the initial frequency of the chirp,

T is the duration of a single chirp pulse and the instantaneous phase $(2\pi f_0.t + \varphi(t))$ varies linearly with time.

The instantaneous frequency $f(t)$ is given by (2):

$$f(t) = \frac{1}{2\pi}\frac{d}{dt}(2\pi f_0.t + \varphi(t)) = f_0 + \frac{1}{2\pi}\left(\frac{d}{dt}\varphi(t)\right) \qquad (2)$$

If we make $\varphi(t) = 2\pi C_r.t^2 + \varphi_0$ where C_r and φ_0 are constants then $f(t)$ becomes (3):

$$f(t) = f_0 + 2C_r.t = f_0 + R.t \qquad (3)$$

We can notice that the instantaneous frequency is a linear function of time with a slope R given by (4):

$$R = 2C_r = \frac{B}{T} \qquad (4)$$

where B is the bandwidth. R is called chirp rate. Therefore the instantaneous frequency is increasing or decreasing over the chirp pulse duration and the bandwidth B equals the swept frequency.

For increasing instantaneous frequency (up chirp) the LFM signal is given by (5):

$$S(t) = A.cos(2\pi f_0.t + \pi R.t^2 + \varphi_0). \qquad (5)$$

For decreasing instantaneous frequency (down chirp) the LFM signal can be represented by (6):

$$S(t) = A.cos(2\pi f_0.t - \pi R.(t^2 - 2T.t) + \varphi_0) \qquad (6)$$

As represented in Figure 2, the transmission of a bit 1 corresponds to the band B1 (402.9 MHz - 403.2 MHz) and the transmission of a bit 0 corresponds to the band B0 (403.2 MHz - 403.5 MHz).

2.2. Transceiver architecture

The super heterodyne architecture presents problems associated with the use of several frequency transposition stages, such as the image frequency that must be taken into account during the design of the receiver. Moreover, this architecture needs an external IF filter which is a major problem to full integration.

In order to avoid image interference, the choice of the receiver architecture was the direct conversion (or zero IF) receiver. In zero IF architecture we can ensure the hardware minimization because it eliminates the image–reject filter and other IF components. Except Low Noise Amplifier (LNA) all amplification and filtering will take place at very low frequencies. Therefore, it is suited to a monolithic integration.

A Linear Frequency Modulation with Binary FSK scheme is proposed for this sensor node. This modulation has advantages over other techniques such as OOK and PPM. The reasons are: the demodulation operation are simple because it demands a zero threshold detector for the decision circuit and it is less affected by inter-pulse interference (IPI) for the same pulse rate [11].

The only disadvantage is that quadrature local oscillator waveforms are produced at the RF frequency. In consequence, the

power consumed will increase. But the LO frequency is about 403.5 MHz, which is considered relatively low frequency so the power consumption of the quadrature local oscillator is acceptable [12].
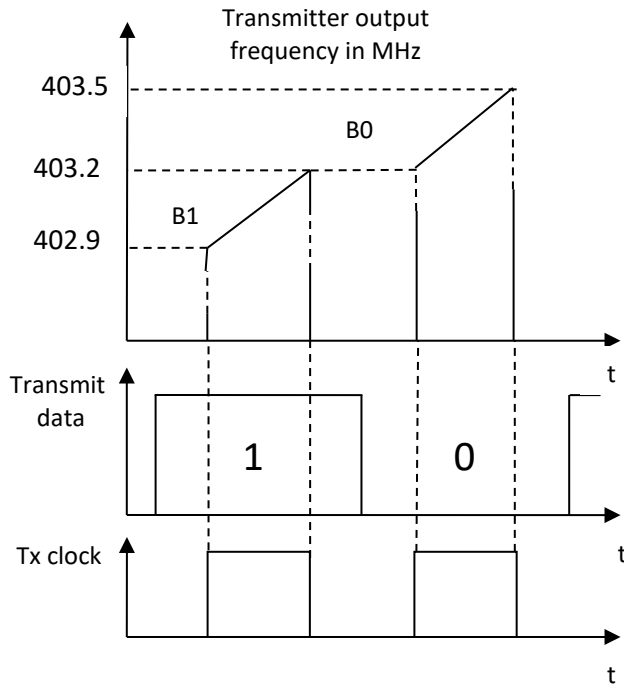
Figure 2. Output frequency of the transmitter in time domain
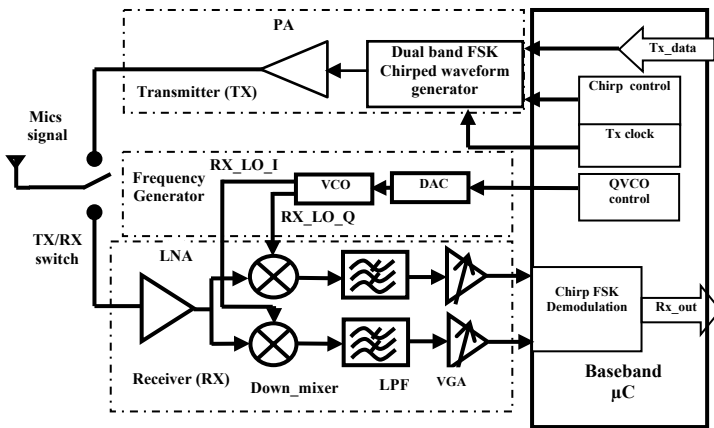
Figure 3. Illustration of zero-IF transceiver architecture.

Figure 3 shows different blocs of the proposed transceiver. The transmitter (TX) is composed of a dual band FSK chirped waveform generator and a power amplifier (PA).

A chirp control and Tx clock signals are used to switch between B0 and B1 signals depending on transmitted binary information and produces the LFM Binary FSK waveform. The power amplifier brings the signal power to the desired level in order to drive the antenna.

The RF band pass filter is not present in this design because it increases the noise figure and the circuit complexity of the receiver. In the other hand band pass filtering supplied by both

antenna and LNA can mitigate the out-of-band signals at the receiver input.

After the antenna a LNA is used to amplify and band pass filter the received signal then the mixer down-converts it using quadrature LO signals. LO tones are produced with a VCO controlled by a Digital to Analog Converter (DAC). The DAC must generate the different voltages so that the VCO will be able to scan the entire RF frequency range from 402 MHz to 405 MHz with steps of 30 KHz. This technique is very useful for improving sensitivity and specially in presence of interferer. However, the system varies the LO frequency with a certain step to sweep the entire RF band and look for a clear channel and then perform the reception and demodulation. Before the VGA a low pass select filter (LPF) passes the desired channel and removes the other channels. Then a dual band chirped binary FSK demodulator based non coherent technique will recover the binary transmitted data [13].

3. Receiver design and simulation

In this work ADS tool is used for simulation and the receiver schematic is presented in figure 4. The aim of this simulation is to determine parameters for LNA, mixer, LPF and VGA that meet MICS and receiver specifications.

For this design the Pulse Rate Frequency (PRF) is set to 100 KHz and the bandwidth B is set to 300 KHz.
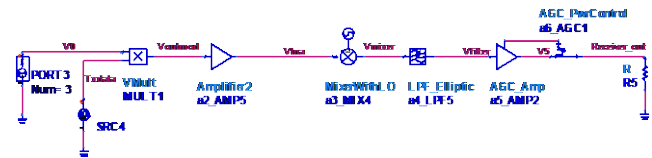
Figure 4. Receiver schematic.
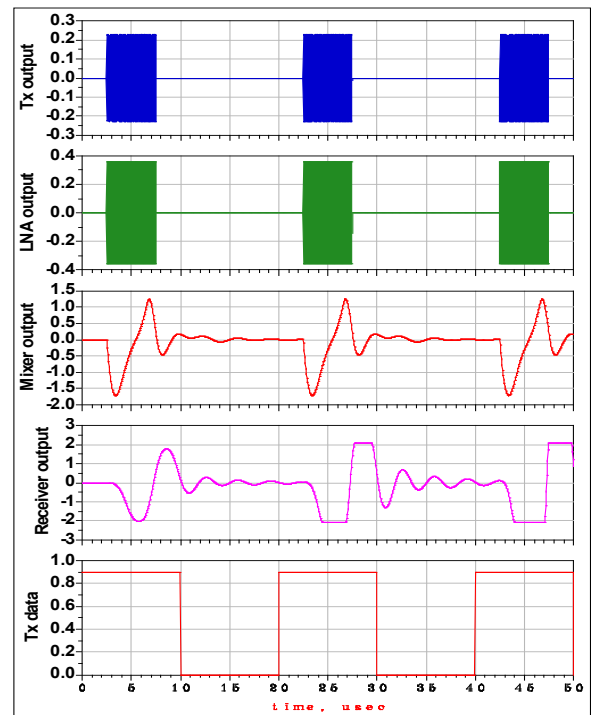
Figure 5. Tx binary data, Tx output, LNA output, mixer output and receiver output in time domain for $t_p=5\mu s$, Pin = -16 dBm and a PRF of 100KHz

3.1. Spectrum at each node of the receiver

A harmonic balance simulation was made for both minimum and maximum receivers input power in order to evaluate receiver performances. Figure 5 presents time domain signals of the transmitted binary data (Tx data), transmitter output (Tx output) as spread spectrum BFSK signal, LNA output, mixer output and receiver output for t_p=5µs, Pin = -16 dBm and a PRF of 100KHz.

Figure 6. Receiver input in frequency domain

Figure 6 presents the spectrum at receiver input for channel number 5. We can see that the bandwidth B=300 KHz and the maximum input power Pin=-16 dBm. Figure 7 shows LNA output spectrum. Mixer output, LPF output and receiver output spectrums are shown in Figure 8 for maximum input power Pin=-16 dBm.
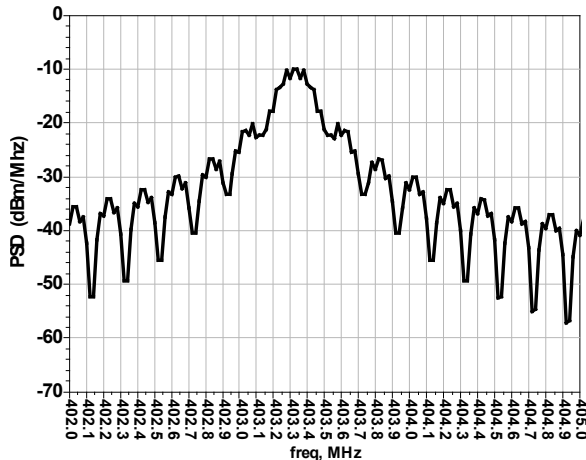
Figure 7. LNA output in frequency domain (Pin=-16 dBm)

It's clear that the result of simulation of the output Power Spectral Density (PSD) aligns FCC mask requirements such as bandwidth 300 KHz. Therefore, the receiver down converts the received signal properly.

Figure 9 presents receiver input in frequency domain for input power Pin=-93 dBm which is the receiver sensitivity. In the same way, Mixer output, LPF output and receiver output spectrums are shown in Figures 10 and 11 for input power Pin=-93 dBm.

From Figures 7 and 11 we can notice that the receiver output dynamic range is between -62 dBm/MHz and 0 dBm/MHz. This power level can drive the chirped Binary FSK demodulator stage.
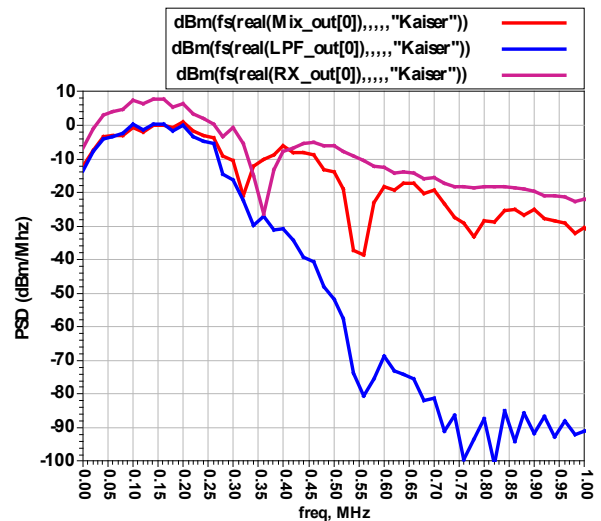
Figure 8. PSD for mixer, LPF and RX outputs, Pin=-16 dBm
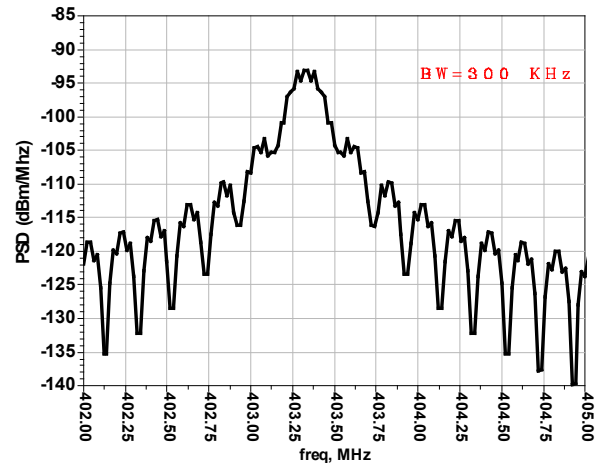
Figure 9. RX input in frequency domain for channel 5, Pin=-93 dBm
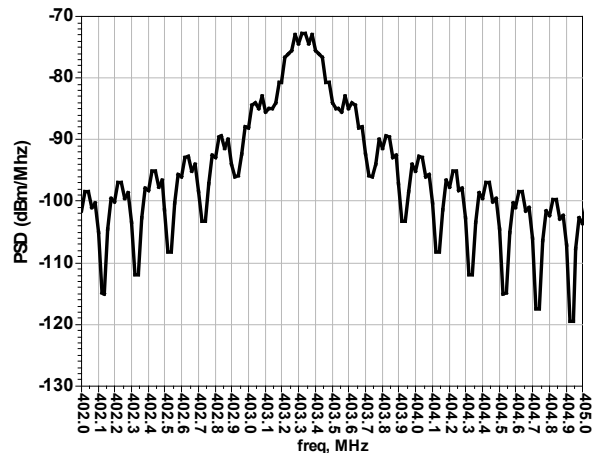
Figure 10. LNA output in frequency domain (Pin=-93 dBm)

3.2. LPF design

A variable bandwidth and sharp roll-off LPF has been used to select the clear channel from the 10 channels used in the MICS standard and reject at maximum interference. In this design we used a 5th order Elliptic filter with roll-off, 2dB rejection loss and

65dB attenuation. A tuning range from 40 KHz to 300 KHz is adopted as shown in figure 12. In addition this variable bandwidth Low Pass Filter can reject any narrowband interference by sweeping its bandwidth. This will improve the receiver performance.
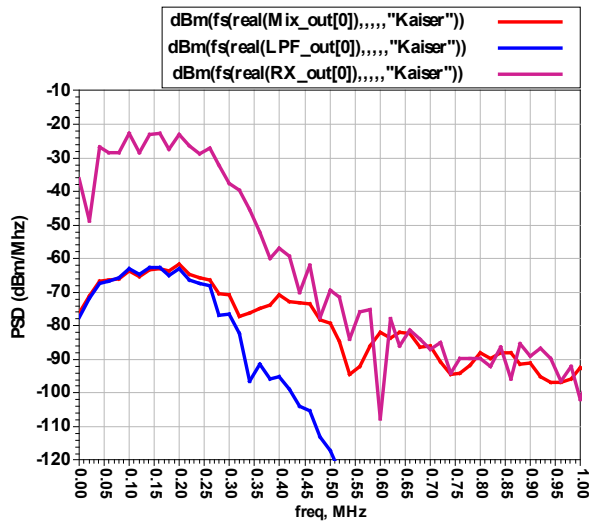
Figure 11. Spectrum for mixer, LPF and RX outputs, at a sensitivity Pin=-93 dBm.

3.3. Budget simulation

A Budget simulation for Pin= -16dBm was conducted which led to the following results: In Figure 13(a) the cascaded gain is 29dB, in Figure 13(b) the cascaded output power is equal to 13dBm and in Figure 13(c) the Noise Figure (NF) is 5.5dB.
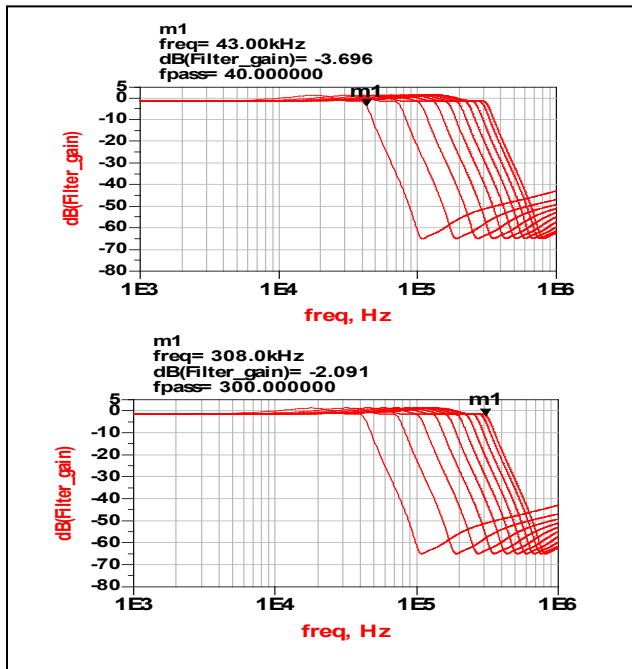
Figure 12. LPF response in frequency domain

For Pin= -93dBm (sensitivity) a Budget simulation was conducted which led to the following results: In Figure 14(a) the cascaded gain is 102dB, in Figure 14(b) the cascaded output power is equal to 10dBm and in Figure 14(c) the Noise Figure (NF) is 5.5dB.

Figure 13(d) demonstrates that the receiver achieves a P_{-1dB} of -22.1dBm and an IIP3 of -12.6dBm. This means that the receiver is linear up to -22.1dBm. The sensitivity of the receiver (-93dBm) is an excellent value compared with Bluetooth or Zigbee communication systems. Therefore this receiver can detect strongly attenuated input signals. A receiver NF of 5.5dB is acceptable to achieve a bit error rate of 10^{-3}.
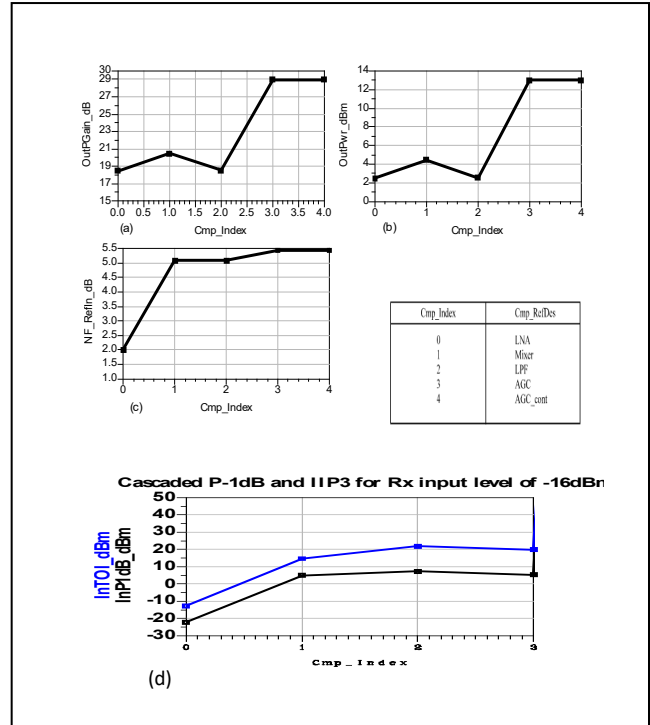
Figure 13. (a) receiver gain, (b) receiver output power, (c) receiver NF, (d) 1dB compression point and IIP3, for Pin=-16dBm.
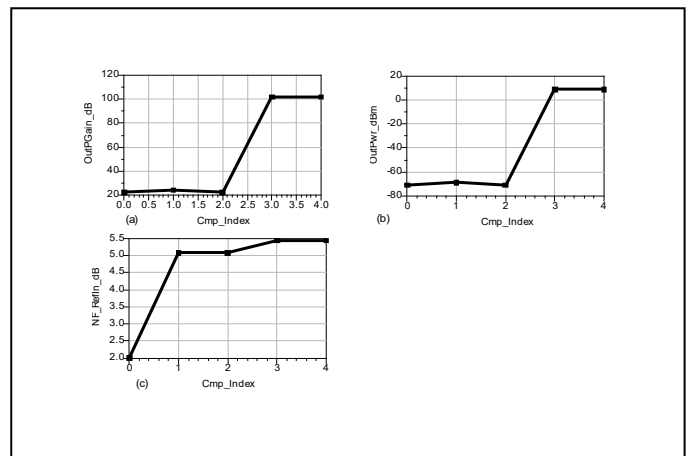
Figure 14. (a) receiver gain, (b) receiver output power, (c) receiver NF, for Pin=-93dBm.

4. FSK dual band Demodulator

The design of demodulator circuit for Binary Frequency Shift Keying (BFSK) modulation request simple circuits that consume low power. Non coherent demodulator based on an envelope detector, integrate and dump, a sampling and hold and a limiting

circuit still provides acceptable performance for low complexity MICS receivers, especially in multipath environments.

(a)

(b)
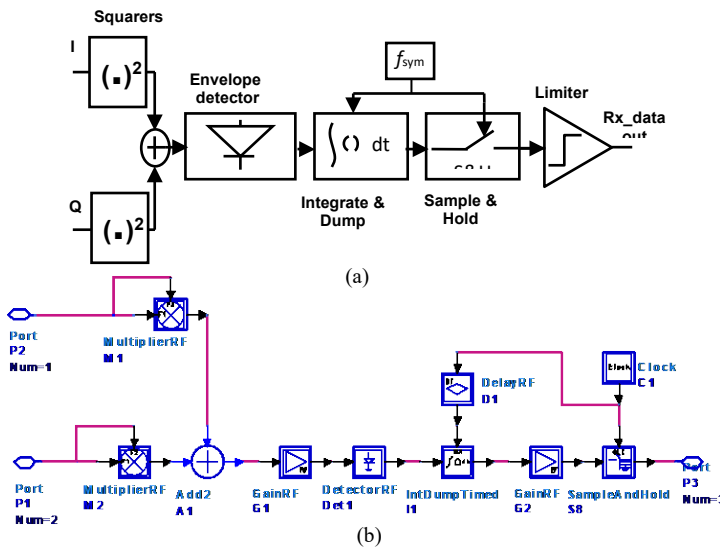
Figure 15. (a) block diagram of the Binary dual band FSK demodulator, (b) Implemented on ADS
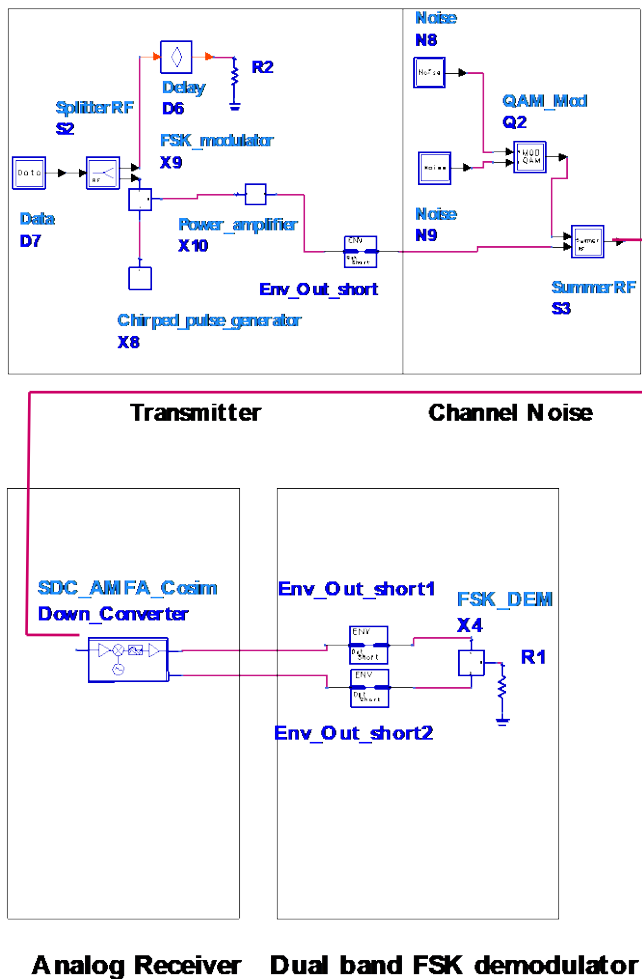
Figure 16. The template used to Co simulate the receiver with the chirped BFSK demodulator

Figure 15(a) presents the proposed FSK dual band demodulator. Advanced Design System tool was used to simulate this demodulator and the corresponding schematic is shown in Figure 15(b). Tow squarer (multiplier RF) are employed to perform energy detection. The I and Q signals are added (Add2). The combined signal is fed into an envelope detector then into integrate and dump block. After sampling-and-hold, a limiting amplifier will retrieve transmitted data. This technique provides acceptable performance for low complexity receivers, especially in multipath environment.

The analog transceiver with the chirped dual band FSK demodulator are co simulated and the corresponding template is shown in Figure 16. AWGN channel noise was used to run this co-simulation.

Figure 17 shows time representation of Tx data and chirped pulses for bit 1 and bit 0.

Figure 18 shows Tx spectrum for bit B1, bit B0 and the dual band Tx output spectrum. We can notice that the Tx output spectrum respects the FCC mask requirements for the MICS band. The emission power is under -16 dBm in order to avoid interferences with other communication services using the same MICS RF band.

The receiver output spectrum after low pass filtering and amplification is represented in Figure 19. The in band power of 0 dBm can drive the dual band FSK demodulator input.
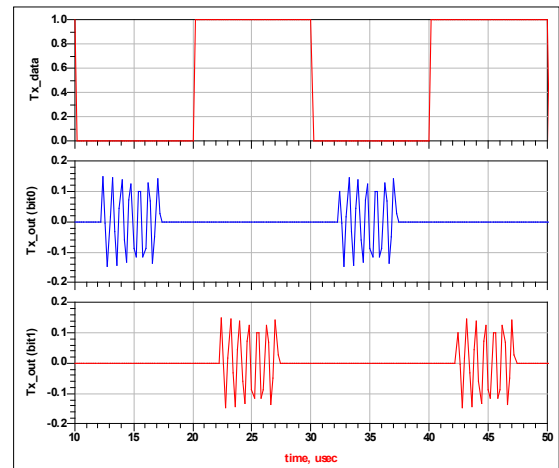
Figure 17. Time representation of Tx data and chirped pulses for bit 1 and bit 0
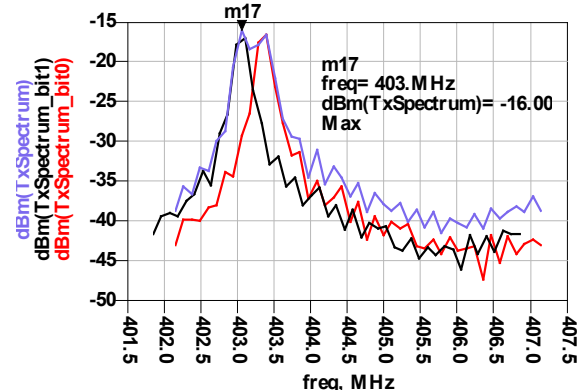
Figure 18. Tx spectrum for t_p=5µs and PRF=100KHz

Figure 20 shows waveforms at the adder output (adder_out), on the output of the envelope detector (env_det_out), after the block of integration (int_out), at the sample and hold output (S&H_out) and at the system output (Rx_dataout).

The co-simulation results show that the binary signal at the output of the demodulator is exactly the same as the delayed binary signal transmitted. Thus proposed demodulator is working properly.
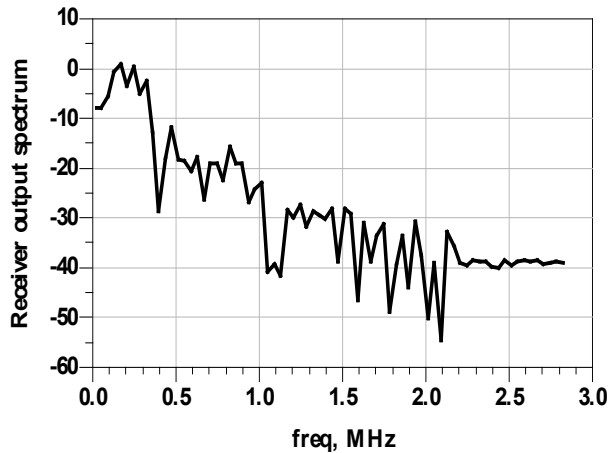
Figure 19. Output spectrum of the direct conversion receiver
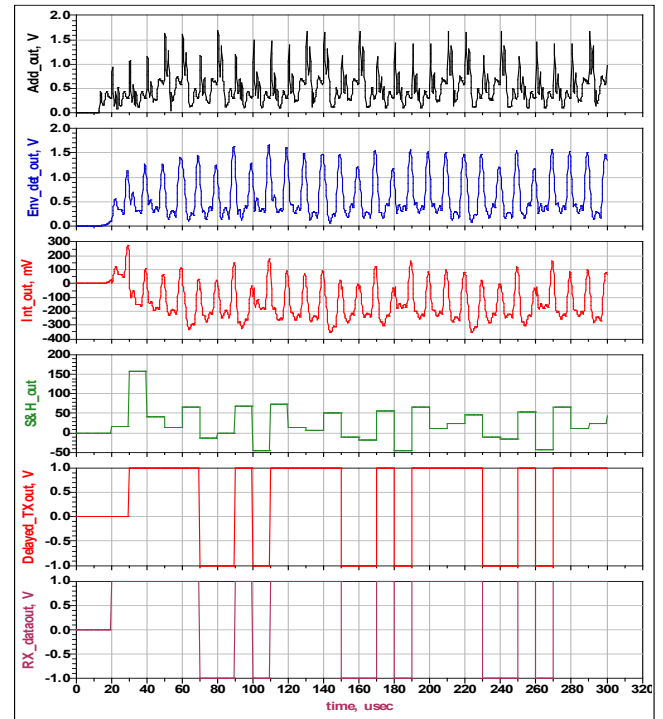
Figure 20. Time domain waveforms at different demodulator nodes

Table 1. Summary of the main technical parameters of the receiver

Receiver performance		Receiver blocks requirements			
Chirp FSK 402-405 MHz	Bandwidth 300 kHz	*LNA*	*Mixer*	*Var-BW-LPF*	*VG-Amplifier*
Noise figure	5.5 dB	S21= 22 dB	Con gain= 7 dB	Fpass= 40-300 kHz	Max_con gain= 80 dB
Gain	102 dB	S11= -13 dB	S11= -10 dB	Order N= 5	S11= -15 dB
IIP3	-12.6 dBm	NF= 2 dB	S22= -10 dB	Max rej= 65 dB	S22= -20 dB
Sensitivity	-93 dBm	P_{-1dB} = -22dBm	S33= -15 dB	ILoss= 2dB	NF= 13 dB
Data rate	100 Kbps	IIP3= -7dBm	NF= 7 dB	-	IIP3= 26 dBm

Table 2. Comparison with previously reported receiver

	This work	[12]		[13]	[14]	[15]	[16]
Frequency	402-405 MHz	402-405 MHz		3-5 GHz	402-405 MHz	433 /868 MHz	402-405 MHz
Modulation	LFM FSK	DBPSK	DQPSK	Dual Band FSK	FSK	FSK /OOK	BFSK
Bandwidth	300 KHz	300 KHz	300 KHz	500 MHz	300 KHz	-	300 KHz
Data Rate	100 Kbps	75.9 /151.8 Kbps	303.6 Kbps	3.2 Mbps	800 Kbps	100 Kbps	100 Kbps
Tx power	<-16 dBm	<-16 dBm	<-16 dBm	<-14 dBm	<-16 dBm	<-16 dBm	<-16 dBm
Max pulse width	5 μs	-	-	75 ns	-	-	-
Max PSD	-16 dBm	-	-	-41.3 dBm	-4.5 dBm	10 dBm	0 dBm
Sensitivity	-93 dBm	-92 dBm	-89 dBm	-	-83 dBm	-111 dBm	-70 dBm
Noise Figure	5.5 dB	6dB	6dB	5.32 dB	-	-	11-40 dB
Gain	102 dB	-	-	99 dB	-	-	35-51 dB
IIP3	-12.6 dBm	-	-	-8.8 dBm	-	-	-

5. Circuit blocks parameters and performance

The harmonic balance simulations results prove that the proposed design respects the constraints imposed by the FCC spectral mask and requirements of MICS standard.

In addition, the performance of the receiver is improved because it is able to reject any narrowband interference by sweeping the bandwidth of the LPF.

This chirped dual band BFSK demodulator can reply for the main challenge that is the low power. So, it is the best candidate for this kind of receiver. With this demodulator we can reach a better sensitivity.

In Table 1 we give a summary of the main technical parameters of the receiver showing its performance. The optimized specifications of the functions constituting the receiver are listed as well.

Table 2 shows a performance summary of the chirped BFSK MICS receiver. A comparison with other works dedicated for healthcare devices are listed as well.

It is proved that the proposed chirped BFSK MICS receiver designed using LFM technique and non-coherent dual band FSK demodulation has a sensitivity and a NF better than those found in the works [12], [14] and [16]. The results found in this design show that the proposed MICS receiver can be realized by simple, inexpensive and low consumption circuits which authorize healthcare implantable devices application.

6. Conclusion

A low-complexity LFM dual-band Binary FSK receiver for medical implantable wireless devices has been presented. A direct conversion architecture without RF band pass filter was used to avoid image interference and reduce circuit complexity. Linear Frequency Modulation technique was used to generate chirped Binary FSK signals. In order to reach low power implementation a simple non coherent demodulator was adopted. It is based on an envelope detector and simple decision circuits. Time domain, frequency domain and budget simulations results has been presented. We demonstrate that the proposed design respects the constraints imposed by the FCC and MICS requirements. A co-simulation of the analog receiver with the proposed non coherent demodulator shows that the receiver can reach a sensitivity of -93 dBm. In addition, the receiver is robust because it is able to reject any narrowband interference by sweeping the bandwidth of the LPF. This design allows the total integration of the sensor node in the same chip using low power CMOS technology. It promotes its use on a large scale in sensor networks implanted in the human body. This will essentially help people with disabilities, people with chronic diseases, people living alone and aged people to overcome their disabilities and live like healthy people.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] M. Bettaieb, H. Trabelsi and M. Masmoudi, "A Novel MICS Chirp FSK Receiver Front-End" in International Conference on Engineering & MIS ICEMIS2017, Monastir, Tunisia, May 2017.

[2] Voskerician G., Shive M.S., Shawgo R.S., Von Recum H., Anderson J.M., Cima M.J., Langer R. "Biocompatibility and biofouling of MEMS drug delivery devices" Biomaterials, 24(11), 2003, pp.1959-1967.

[3] World Population Ageing 2015, Departement of economic and social affairs Population Division.United States.New York, 2015.

[4] H.S. Savci,A. Sula,E. Arvas, "MICS Tranceivers: Regulatory standard and applications", in Proc.of Southest Con,April 2005, pp. 179-182.

[5] J. Ryckaert, M. Badaroglu, V. D. Heyn, G. V. der Plas, P. Nuzzo, A. Baschirotto, S. D'Amico, C. Desset, H. Suys, M. Libois, B. V. Poucke, P. Wambacq, and B. Gyselinckx, "A 16 mA UWB 3-to-5 GHz 20 M pulses/s quadrature analog correlation receiver in 0.18 μm CMOS," in IEEE ISSCC Dig. Tech. Papers, pp. 368–377, 2006.

[6] FCC Rules and Regulations, "MICS Band Plan" , Part 95, jan 2003.

[7] D. C. Daly and A. P. Chandrakasan, "An energy efficient OOK transceiver for wireless sensor networks," IEEE J. Solid State Circuits, vol. 42, no. 5, pp. 1003–1011, 2007.

[8] J. Ayers, N. Panitantum, K.Mayaram, and T. S. Fiez, "A 2.4 GHz wireless transceiver with 0.95 nJ/b link energy for multi-hop battery free wireless sensor networks," in Symp. VLSI Circuits Dig. Tech. Papers, pp. 29–30, 2010.

[9] H. Trabelsi, Gh. Bouzid, F. Derbel and M. Masmoudi, "A 863-870MHz Spread Spectrum FSK Transceiver Design for Wireless Sensor," Microelectronics Journal (MEJ), vol. 41, No 8, pp 465-473, August 2010.

[10] Simanjuntak, Lastri, "A Novel Chirp Slope Keying Modulation Scheme for Underwater Communication", University of New Orleans Theses and Dissertations. Paper 201, December 2004.

[11] R. Roovers, D. M. W. Leenaerts, J. Bergervoet, K. S. Harish, R. C. H. van de Beek, G. van der Weide, H. Waite, Y. Zhang, S. Aggarwal, and C. Razzell, "An interference-robust receiver for ultra-wideband radio in SiGe BiCMOS technology," IEEE J. Solid-State Circuits, vol. 40, pp. 2563–2572, 2005.

[12] BA et al, " a 0.33nJ/bit IEEE802.15.6/Proprietary MICS/ISM Wireless Transceiver With Scalable Data Rate for Medical Implantable Applications", IEEE Journal of Biomedical and Health Informatics, Vol.19 .No.3, May 2015.

[13] H. Trabelsi, I. Barraj and M. Masmoudi "A 3-5 GHz FSK-UWB Transmitter For Wireless Personal Healthcare Applications"; International Journal of Electronics and Communications (IJEC), Vol. 69, pp. 262-273, 2015, DOI:10.1016/j.aeue.2014.09.009.

[14] P. D. Bradley, "An Ultra-Low power, high performance medical implant communication system (MICS) transceiver for implantable devices," in Proc. IEEE Biomed. Circuit Syst. Conf., pp .158-161, December 2006.

[15] V. Peiris, C. Arm, S. Bories, S. Cverveny, F. Giroud, P.Graber, S. Gyger, E. Le Roux, T. Melly, M. Moser, O. Nys, F. Pengg, P.-D. Pfister, N. Raemy, A. Ribordy, P.-F.Ruedi, D. Ruffieux, L. Sumanen, S. Todeschin, and P. Volet, " A 1V 433/868MHz 25kbps-FSK 2kbps-OOK RF transceiver SoC in standard digital 0.18μm CMOS," in Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, pp. 258-259, February 2005.

[16] S. Kim, W. Lepkowski, S. J. Wilk, T. J. Thornton and B. Bakkaloglu, "A Low-power CMOS BFSK Transceiver for Health Monitoring Systems" in *IEEE Biomedical Circuits Syst Conf*. 2011, 157–160. doi:10.1109/BioCAS.2011.6107751.

A Machine Learning Framework Using Distinctive Feature Extraction for Hand Gesture Recognition

Sudipta Saha[*,1], Aninda Saha[2], Zubayr Khalid[1], Pritam Paul[1], Shuvam Biswas[1]

[1]Computer Science and Engineering, Institute of Engineering and Management, Kolkata, 700064, India

[2] Electronics and Communication Engineering, Institute of Engineering and Management, Kolkata, 700064, India

A B S T R A C T

There are more than 7billion people in the world where there are around 500 million people in the world who are denied from normal lifestyle due to physical and mental issue. It is completely fair to say that every person deserves to enjoy a normal lifestyle. While physically and mentally challenged people find suitable way to surpass their limits, thus become able in other ways, researchers always try to find solutions better than the existing one. A complete remedial of such issue is included in advanced medical science, and the amelioration of such issue to a better extent is the challenge for the engineers.

In this work we have focused on hand gestures. Hand gestures are created using the movement of hand and arm, using fingers to create different shapes, using fingers and palm to create different angles. Single or both hands can be used to create different expressions. The main objective of this work is to generate an algorithm that can recognize different patterns of hand gestures with notable accuracy. American Sign Language is one possible reference model that can be used. Images of different hand signs are taken as inputs using a webcam, followed by segmentation of the images using polygon approximation and approximate convex decomposition. Feature extraction is done by recording the unique feature among the various convex segments of the hand. The resultant singularities are then used as extracted feature vectors. This involves training with the obtained features which are approximately unique for different hand gestures. Hence, we will be able to identify sign languages and successively make disabled individuals socially acceptable. This work is an extension of the work entitled "A Machine Learning Framework Using Distinctive Feature Extraction for Hand Gesture Recognition" in 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (IEEE-CCWC).

1.   Introduction

Gesture Technology is an emerging scope in the field of Computer Science, precisely machine learning. The objective of gesture technology is to interpret or recover the human gesture using intelligent machines. Complex mathematical algorithm is used to interpret the gestures. Fundamentally spoken, the mathematical algorithms contain a large set of test and training data, which are eventually used to compare with the current status of the human, thus come out with an appropriate result that can describe the gesture. Gesture is vastly used in human-machine interaction where the machine detects gesture input from the human and responds accordingly. Advanced technology has created a possible way for communication between 2 or more

machines, with no human presence, using gesture input and output. Furthermore, gesture technology can be very useful in human to human interaction, precisely where the persons involved in the communication do not have the same mode of communication. While other body parts are also used in gesture technology, the most used parts are face and hands. In general term, gesture can be divided into 2 types- a) body gesture: where body is used to create expression; and b) hand gesture: where only hand is used to create expression. A sign or symbol is the most commonly used hand gesture [1]. Gesture can be further distinguished on basis of static gesture [2] and dynamic gesture [3]. Face images in gesture technology are mostly used for emotion detection [4] and as a security tool used in gadgets and other devices [5]. Some main application areas for gesture technology include gaming sector, home automation, automotive sector, sign language interpretation.

*Sudipta Saha, Kolkata, India, +918013554858, subho040995@gmail.com

In hand gesture technology, human can interact with machine or with another human through a machine that recognizes the hand gestures created by the human being [6]. Generally still images of hand gestures are sent as inputs to the machines, but there are also machines that can collect gesture data from live video footage. Thanks to advanced technology, touching of the machine is not necessary anymore; strong sensors in the machine can detect hand gestures from distance. Hand gesture technology is verily used in the later two of the previous example. In this domain, we will talk about hand gesture technology and sign language interpretation.

Enjoying the proper lifestyle is the birth right of every human being in this world. Everybody should be able to communicate and share expression with the outer world, but unfortunately not every time people are able to do so. It is always hard for people to communicate with others when mental or physical challenges become a barrier in the way. While it is safe to say that these people are verily able in other ways that help them to communicate, we always try to find a system better than the existing one. The most fundamental and most popular technique used in this field is creating sign language using hand gestures [7]. But the challenges occur when at least one of the people involved in such communications doesn't know the sign language or how to create them. A few numbers of people who don't need sign language for communication actually know this language, resulting in a great difficulty when there is a communication between 2 people where one of them only uses sign language for sharing expression. Based on practical experience, it is almost certain that these types of conversations do not proceed easily unless there is a third person who is used as an interpreter. There exist devices, as for example, data gloves, which are used to solve such problems during conversation. Data glove is an interactive device, basically a glove to wear with sensors on it. These sensors create precise capturing of hand gesture information. Though data glove is generally used in robotics, other application areas also include human interaction. This report refers only to camera based static hand gesture recognition [3]. Static hand gesture recognition technology has a useful application on communication between 2 persons where one is able to use a particular sign language (say ASL or American Sign Language) and other isn't. A hand gesture recognition device is used as the protocol in this communication. As mentioned earlier, the device contains sensors to capture images and to detect gesture data and analyze them. The basic objective of the device is to eliminate the necessity of a third party interpreter, by allowing itself to do the same job. It captures gesture images as input and creates output suitable for non-ASL speaker. Generally the output is in textual format. The device is also used in the reverse situation, that is., taking textual or vocal words as inputs and creating sign language as outputs. Since the algorithms used in a device is based on a particular sign language, it is necessary that mode of communication from the sign language user's side only includes one specific language, such as either Indian Sign Language (ISL)[8] or the American Sign Language (ASL)[9] but not an amalgam of both. The objective of this paper is to illustrate the extraction of unique features from each symbol or gesture in reference to an existing data set (ISL or ASL) and then train the machine with the obtained feature vectors using standard classification models.

Using a huge number of data in test set and training set creates less ambiguity for the machine while taking decision, and as a result it can give a more accurate output. As for example, a specific hand sign can be described by several textual words, and a set of hand signs can be described by 2 or more different sentences. The more there is data in the device's memory based on previous trainings (that is, in training data set), the more accurately the device will be able to decode texts into proper sign languages. There are more necessities to be taken care of in order to get more accurate results. Use of white gloves is employed to eliminate unwanted wrinkles or groves on the palm and fingers. From the threshold image of the figure we obtain its minimum fitting polygon through what we call polygon approximation technique. The next step is Convex Decomposition technique. It segregates each of the convex parts of the hand, which is an essential step to identity the crucial aspects of the gesture. The next and final step is to extract the required unique features. The wrist point in the hand is considered to be the mid-point of the lowermost polygon.

The novelty of the proposed method is listed as follows.

- Extraction of individual features for each and every gesture creates the possibility of notable polarity and a chance of further efficient classification of the obtained data.
- The proposed method is equipped for large scale deployment through recognizing the demand and necessity of real time application.
- The produced results from trained dataset are favorable with adequate acceptance rate.

2. Related Work

Both the Indian Sign Language (ISL) and American Sign Language (ASL) are comprised of a notable list of hand gestures, each with meaningful expression that can be used for any non-verbal communication. As mentioned earlier, the sign languages come real handy when we have the technology to decode a specific sign language to sounds/texts and vice-versa, thus useful in communication where one person is accustomed with sign language communication where as other person use vocal/ textual communication and does not know the sign language. In probability, Markov Model is a concept for modelling randomly changing system. According to this model, the next state of a random system mostly depends on the current state of the system; and not on the previous states. In 1960 William Stokoe developed a concept based on real time Markov-model based systems [10]. In this system, continuous American Sign Language (ASL) generates long combination of words as outputs. A single camera is used in this concept. The concept is still very much effective and is majorly used technique to recognize gestures. However, this implementation requires polygon approximation followed by convex decomposition.

The convex decomposition method is a special case of alternating sum of volumes (ASV) method. ASV can be defined as a series of convex elements stick together in a way such that for elements A, B, C, D, E, ... the series will be: A u B / C u D / E and so on, where u is Union operation and / is Difference Operation [11]. In 1998, Starner proposed the using of convex hulls and set- difference operations [12]. A convergent convex decomposition is proposed that uses a combination of ASV decomposition and remedial partitioning for the non-convergence.

The concept uses ASVP decomposition for image feature extraction, in other words, to identify the unique properties of an image that differentiates the image from others.

Polygon Approximation is very necessary in this context. The simplest idea of the process can be described by 4 steps – Identifying original shape (contour), chopping the shape around the corners, generate approximate convex lines for the chopped segments, merging the segments [13]. The decomposition of polygon with 0 or more holes into convex chops can be done approximately using Approximate Convex Decomposition (ACD) [14]. The algorithm used in this context is simple: iterative resolving of the most significant non-convex feature. Approximate Convex Decomposition method is applicable for both 2D and 3D models.

On the basis of simple cut and split paradigm of Chazelle, a method to compute a convex decomposition of a non-convex polyhedron of arbitrary genus (handles) and shells (internal voids) is generalized [15]. The algorithm is extended to work for a certain class of no manifold polyhedral.

A set of topological decimation operations to its dual graph is applied to calculate a hierarchical segmentation of mesh triangles, which is used for 3D mesh approximate convex decomposition [16]. The method of segmentation incudes a) portraying a triangle mesh and its corresponding dual graph, b) contracting of an arc of the dual graph, followed by the marking of the end-points of the two triangles corresponding to the dual arc's as belonging to the same single cluster [17]. Segmentation can be of two types – direct segmentation and feature based segmentation. This particular model is important due to its adaptation to collision detection.

Inter-class variability enhancement and local-global ambiguity identification for each hand gesture is used in the automatic gesture recognition on Indian Sign Language (ISL) [18]. In ISL both hands can be used to represent an expression. Support Vector Machines (SVM) is used to classify each hand posture. In SVM data items are plotted as points in n-dimensional space, where n is the number of features. To classify dynamic gestures, Dynamic Time Warping (DTW) is used with the trajectory feature vector, and it achieves 86.3% recognition rate.

A pre-trained Google Net architecture representing a real-time sign language interpreter is described while developing the architecture with ILSVRC2012 dataset as training sets [19]. The idea is based on Convolution Neural Network (CNN). The main difference between Convolution Neural Network (CNN) and ordinary Neural Network is that CNN assumes that all the inputs are images, allowing researchers to encode certain properties into the architecture. The architecture is also trained with data sets from Surrey University and Massey University ASL data. This training set allows the architecture to apply transfer learning to this task. The system features a pipeline that captures video of signing a word as input through a designated application. Individual frames of the video are extracted and then they generate letter probabilities using a CNN.

Initially made use of Histogram of Oriented Gradients method extraction of feature, a feature extraction method uses two neural networks based recognition of 10 gestures [20]. The HOG [21] for an image describes the elementary property (such as color, shape and size) of an image. The objective of HOG method is object detection.

Data availability and analyzing their relative merits to determine features are essential for sign-language recognition (SLR) [22]. The objective is to continuous sign recognition and design a model that is independent of signer. The model should also be adapting to larger and noisy data sets. A statistical analysis based on four parameters: position, posture, orientation, and motion is used for a large vocabulary sign language interpreter with real-time continuous gesture recognition of sign language [23].

3. Data Set

The dataset for the proposed method has been centered on the popular and ubiquitous ASL (American Sign Language). It was formulated by the University of Surrey's Center for Vision, Speech and Signal. The images referred have been indicated in the figure (1). We were able to gather approximately more than 500 colored images however the algorithm is constrained to static gesture recognition.
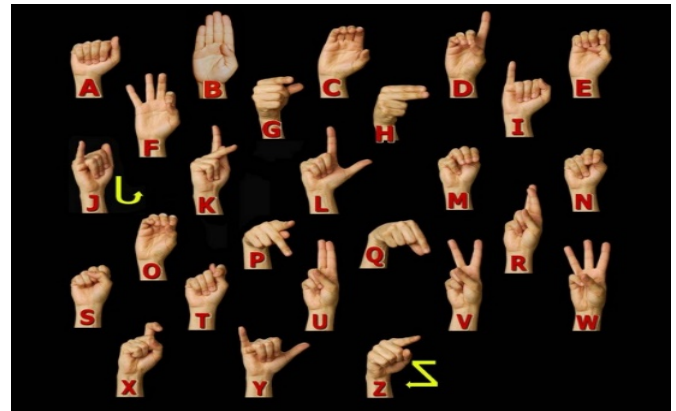
Figure 1: Different Gestures of ASL

4. Methodology

4.1. Pre-Processing

In the first step where we remove uneven parts of the image, dilate and erode and then generate a contour based on color edge detection. In the above set of figures, figure (2.a.I) is the initial figure, which is followed by figure (2.a.II) where we get the processed image. Using the processed image, we can find out the contour of the dark area using color edge detection technique [24] shown in the figure (2.a.III).

Once the contour is obtained then the rest of the area apart from the contour can be removed. We assumed that the hand will fit to (having minimal space between the hand and the frame border) the frame with the frame dimensions being constant. Once the background is removed, we need figure out the edges of the fingers and hands marked by black contrasting borders. Figure (2.a.IV) shows the finally obtained contour after preprocessing.

4.2. Approximation of Minimal Fitting Polygon

In digital image processing the approximation of arbitrary two-dimensional curves by polygonal figures is an imperative technique. In number applications, this technique is widely used

for a purpose to represent lines and boundaries by means of polygons that have a minimum number of vertices so that fit condition is satisfied. Using island finding algorithm, we will be able to separate the fingers by inner smaller polygons.

(I)   (II)

(III)   (IV)
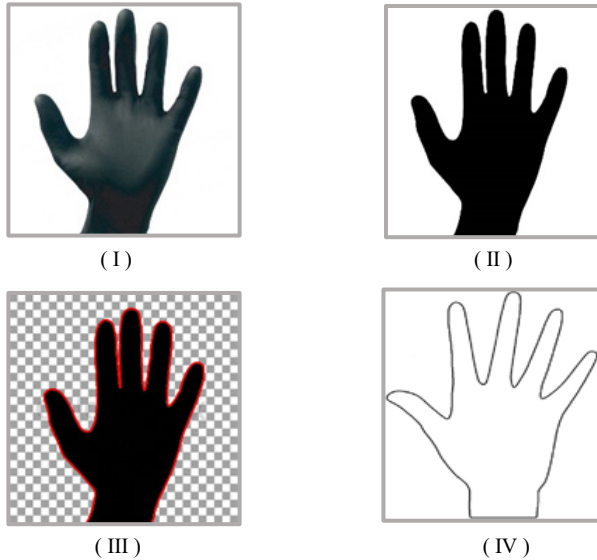
Figure 2 (a): Images Describing the Pre-Processing stage

To find the initial point we need to move by traversing alone the x-axis of the bottom frame and then we can find the first pixel that is colored black. The adjacent pixel can be located by a technique where we check the neighboring 8 pixels for a black pixel. The algorithm navigates through the contour of the figure from pixel to pixel, marking them based on their connectivity and the relative values of their neighbors. Connection between the pixels can be determined with the help of a 3x3 grid as shown in Figure (2.b). This grid can be either 4-connected or 8-connected.
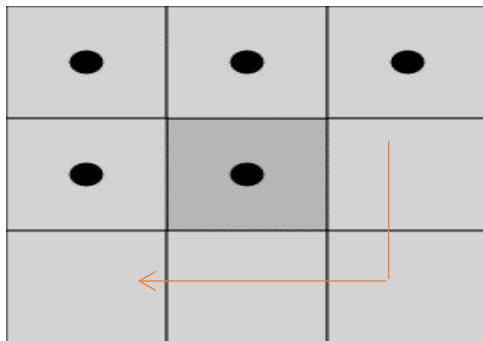
Figure 2 (b): 8 neighboring pixels of a pre-determined pixel

Now we need to find the 2D polygon by the following algorithm:

1. Now take a black pixel at the starting point and continue moving along the pixel.
2. Now joining the starting point pixel and adjacent point pixel (pivotal point) with a line to forms the base. Now we calculate the angle between the base and the next black point and store the calculated angle.
3. Now if calculated angle is not considerably larger or smaller than the previously calculated angle, then we need to eliminate the current pixel point, and take the next pixel point

lying in succession as the pivotal point. Else, if the angle is substantially more or less by a threshold value, then consider the current point and initialize the point as starting point. Go to step
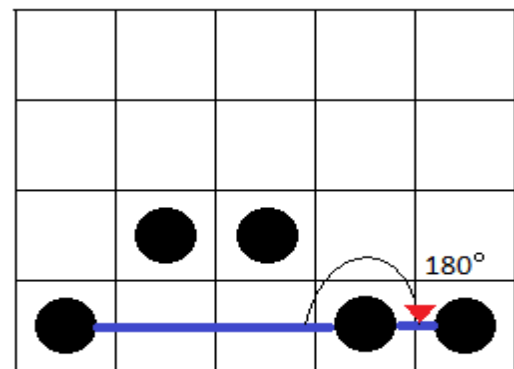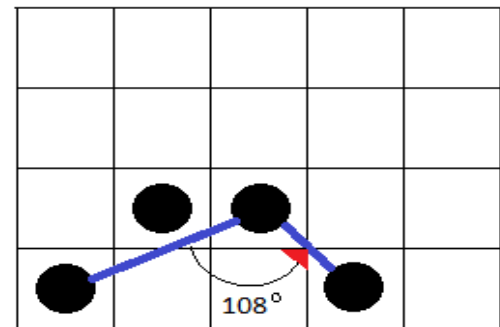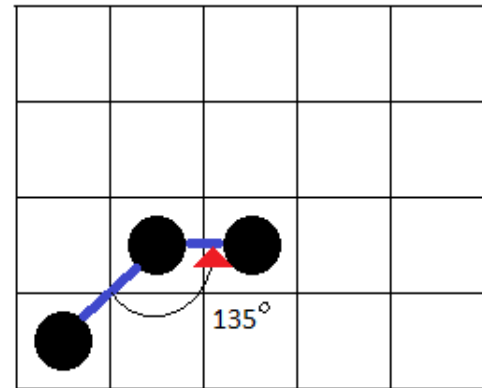
Figure 3 (c): Since 180° is considerably large, it is accepted

4.3. Island Finding Algorithm:

Here, we shall try to find out the minimal cyclic paths that enclose an area. A DFS traversal is done using the following constraints

1. Stack(S) ← Push(R); S is the stack and R is root node.
2. If S==null, end looping, else,

3.   If current node X present in Stack S, go to step 6.
4.   Else if X not present in Stack S and if it has child nodes then expand and explore child nodes if [Child(X)!=visited or Child(X)!=Parent(X)]. Else return to parent node.
5.   Repeat step 2.
6.   If substring[X, indexOfLastNode] or reverse (substring[X, indexOfLastNode]) is not already stored, then store (substring[X, indexOfLastNode] + X) as cyclic path C_n. Mark X as visited, POP(X) and PEEP(S). Go to step 5.
7.   If substring [X, indexOfLastNode] or reverse (substring [X, indexOfLastNode]) is already stored, then mark X as visited, POP(X) and PEEP(S). Go to step 5.

Once we find all the cyclic paths possible (C_1, C_2, C_3….C_N) then in a sequential order consider each cyclic paths, such that if elements of (C_n – lastElement) or elements of reverse (C_n – lastElement) is subset of C_m, then discard C_m. Continue until we have minimal unique cyclic paths possible. Then a given cyclic path forms either an island (finger or palm) or it's the outer contour of the figure.
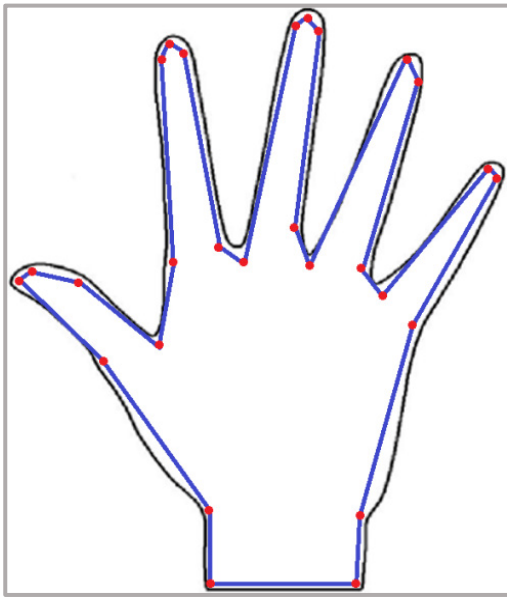
Figure 4: Illustration of Polygon Approximation of Polygon from the Shape of a Hand

In the above figure 4, the polygon with vertices of color red is sketched out from the curved outline of the image. A simple polygon P is formed when the vertices are joined by the blue lines.

The non-essential curves and bends are removed by the technique of approximation of minimal fitting polygon [25] and give us a figure that is bounded by line segments. The resultant image consists of a simple polygonal region in the plane bounded by a non-self-intersecting, closed, polygonal path. The polygonal path itself is part of the polygon; it is usually called the boundary.

4.4.   Decomposition into Convex Components

Now if we partition the complex models into pieces it will be easy to deal with complex models or many-sided figure. Numerous difficulties can be solved more proficiently when the input is convex. Convex components Decomposition into results in pieces that are easy to process and additional approximation

simples and eases computation, ensuring quicker and efficient output with manageable margins of inaccuracy.

Exact and approximate are two type of convex decomposition. Number of clusters in the model should be minimized so we do not use exact decomposition. We can get minimum number of clusters using Approximate convex decomposition ensuring that each cluster has a concavity lower that a predefined threshold.
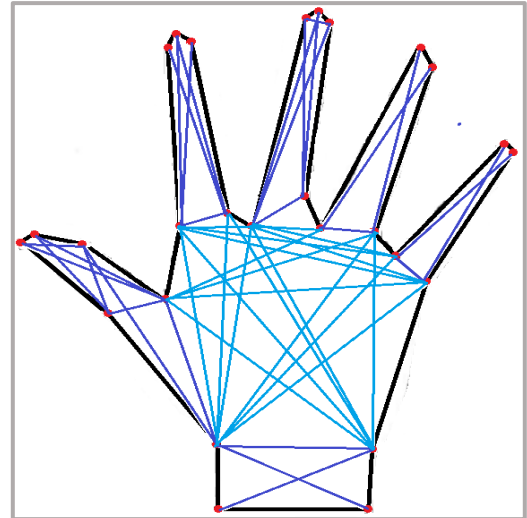
Figure 5: Polygon Decomposition into Non-Concave Sub graphs

The figure 5 given above shows the result of convex decomposition on the approximated polygonal segments obtained in the previous section.
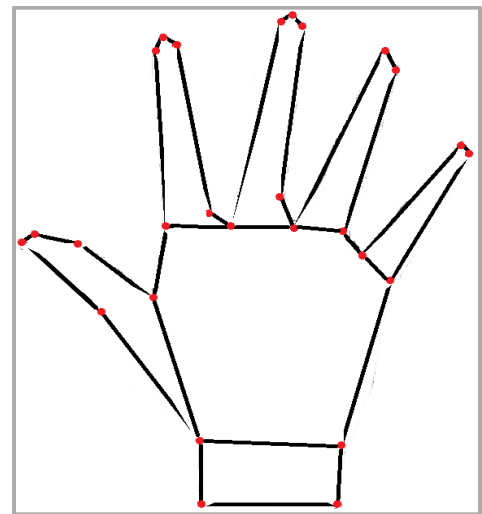
Figure 6: Resulting Convex Decomposition

In graph theory, where a complete subset of vertices of a graph G such that every two distinct vertices in the clique are adjacent is known as a clique. Subsequently, a maximum clique is one that has the largest possible number of vertices among other cliques. We then observe the maximum clique possible for any given point and then iteratively check for the points it connects. For given vertices V = {V_1, V_2, V_3…V_N},the we need to form two groups such that if $V_M \in$ V then, put VM into Checking array(C), and the vertices connecting V_M in Pointing array(P). Put P's last

element into C and iteratively check if C elements has common connections with vertices in P. Group those points to find if they share any common point to find the approximate convex partition [26]. The maximum clique in the above mentioned visibility graph G, which is also undirected, is a complete sub graph of G having the maximum number of vertices.

4.5. Detecting Wrist Point and Reference Line

Deriving the data necessary for feature extraction the wrist point and the reference line are needed. The data are calculated with respect to the reference line. As the image is always considered to have the picture of the gesture in upright position so the wrist point is calculated as the mid-point of the bottom most polygon.
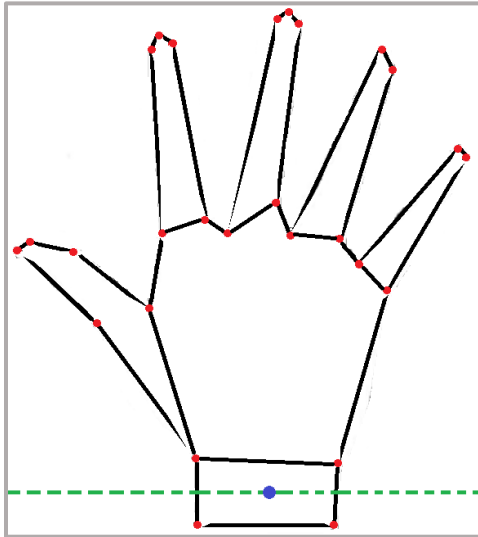
Figure 7: Deduction of Wrist Point and Reference Line

First identify the polygon associated with the wrist to derive the reference line. Then the mid-point of the polygon is calculated. The line that passes though the wrist point and is parallel to the bottom of the frame of reference is known as reference line. The green dotted line in figure 7 indicates the reference line. The bottom-most polygon is the polygon associated with the wrist, and the blue dot is the midpoint associated with the polygon.

4.6. Angle and Distance calculation with respect to Wrist Point

We now have the line of reference as well as the wrist point which are used to realize the different polygon segments in a given figure. To achieve this by observing the incrementing angles between each line joining the mid points of the polygonal segments and the line of reference. Our main objective here is to find unique feature vector to represent different hand gestures.

With reference to figure 8 by the convex decomposition process we begin by finding the mid-points of the segments. These points, marked as P1 through to P6, are then extended to join the wrist point marked as W. What is important to note here is that with each segment mid-point, the angle increases with the minimum being a possible 0° and the maximum being a possible 180°.This gradual expanding gap between the wrist point and each subsequent line signifies the change in finger positions and

helps us to differentiate between them. Another aspect of the line joining the wrist point and the mid points of the segments is its length.
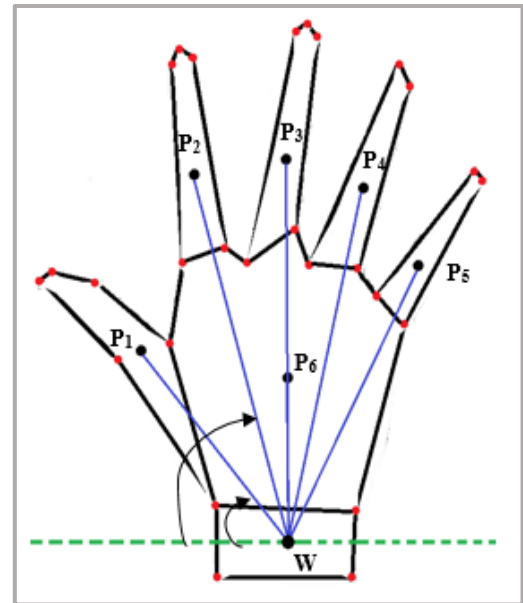
Figure 8: Connection of Wrist Point, described by P1, P2, with the Centroid (W) of the Decomposed Polygons

The property which helps us to decide whether a finger is in a folded or unfolded position is the distance between the wrist point and the mid-point of the corresponding polygonal segment of that finger. While the finger is in a folded position, the centroid of the segment gets located in a vertically lower level inside the polygon than it would if the finger was in an unfolded position. As a result, it becomes fairly easily distinguishable to recognize different signs, especially the ones that only differ by the positioning of a single finger.

Outcome of the above stated procedure produces the angle and distance of each mid-point of the different generated convex segments with respect to the reference line and wrist point. It shall be noted that the duality (angle and distance) needs to be converted to a singularity for feature extraction and forming a unique one-dimensional array representation of each distinguishable hand gesture.
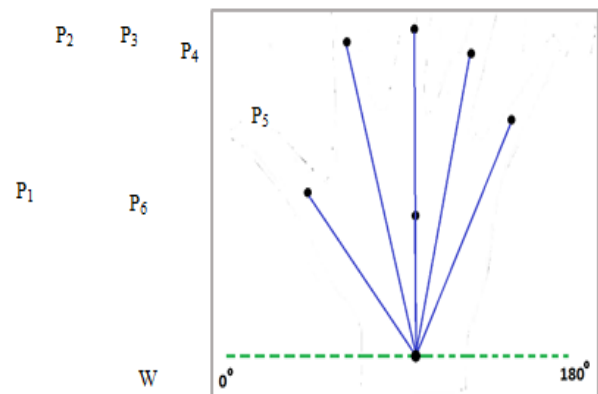
Figure 9: Unique Array of Data to Represent a Hand Gesture

pred - NumPy array

	0	1	2	3	4	5	6	7	8	9
0	0.866076	0.0158627	0.000789759	1.44649e-05	1.95762e-06	3.93198e-05	1.31903e-08	5.32135e-06	9.7464e-08	0.117211
1	0.979508	0.000120881	0.0179403	7.42718e-10	1.49179e-08	1.10229e-07	0.00242796	2.21401e-06	2.99634e-07	1.00516e-10
2	1.64009e-05	0.981988	2.73922e-10	3.00392e-07	6.03354e-06	0.0179857	2.21962e-06	8.16551e-07	8.16551e-07	2.02403e-09
3	1.41648e-08	0.930058	2.10224e-06	0.0170346	0.0463049	1.91699e-09	5.71448e-06	0.000312	1.55336e-05	0.00626668
4	0.00235374	0.047276	0.949565	2.90474e-07	0.000318543	0.000318543	2.14633e-06	4.31102e-05	0.000117186	5.83433e-06
5	5.84761e-06	1.44948e-08	0.951727	1.96165e-09	0.0473837	1.58955e-05	2.65481e-10	5.33233e-09	3.94009e-08	0.000867862
6	1.48066e-08	5.97342e-06	0.00240985	0.972203	0.000119979	1.62375e-05	8.08415e-07	0.000886535	0.00655066	0.0178065
7	0.000328505	2.21345e-06	0.00242734	0.979259	0.0179358	7.42529e-10	4.44583e-05	7.42529e-10	4.05407e-08	2.21345e-06
8	8.14315e-07	2.21354e-06	5.48681e-09	0.0179365	0.979298	0.00242744	6.01702e-06	2.9957e-07	4.05423e-08	0.000328518
9	1.06265e-07	1.43814e-08	2.13438e-06	0.0470129	0.944279	0.0063625	0.00234063	2.88857e-07	2.13438e-06	1.06265e-07
10	0.000892449	6.01328e-06	2.73002e-10	0.0179253	1.63458e-05	0.978689	0.00242593	4.44324e-05	4.05171e-08	4.05171e-08
11	0.00626461	0.000311897	4.22106e-05	1.41601e-08	0.000311897	0.929751	0.017029	0.0462896	5.2092e-09	1.0463e-07
12	0.000121066	2.2174e-06	3.00093e-07	0.0179678	2.73649e-10	7.43855e-10	0.981008	0.000894563	6.02752e-06	4.06131e-08
13	3.91773e-08	1.58052e-05	2.13901e-06	0.000116786	1.06495e-07	0.0471148	0.946325	5.81442e-06	0.00637629	4.29631e-05
14	7.421e-10	1.00432e-10	0.00242594	0.000892452	2.21217e-06	4.44326e-05	1.63458e-05	0.978693	2.99384e-07	0.0179254
15	2.98352e-07	5.46451e-09	0.0178636	5.46451e-09	5.99256e-06	0.000120364	0.000120364	0.975318	4.03775e-08	0.00657164
16	0.0178459	1.48394e-08	1.62734e-05	8.10204e-07	4.03377e-08	0.000888496	0.000326859	0.00656515	0.974354	2.20236e-06
17	0.0178169	2.71351e-10	0.00241126	2.00503e-09	4.02721e-08	5.45024e-09	0.00655448	0.000120049	0.972771	0.000326328
18	1.10398e-07	0.0179678	0.000894563	6.02752e-06	1.10398e-07	2.2174e-06	8.15736e-07	0.000121066	2.02201e-09	0.981007

pred - NumPy array

	0	1	2	3	4	5	6	7	8	9
0	True	False	False	False	False	False	False	False	False	False
1	True	False	False	False	False	False	False	False	False	False
2	False	True	False	False	False	False	False	False	False	False
3	False	True	False	False	False	False	False	False	False	False
4	False	False	True	False	False	False	False	False	False	False
5	False	False	True	False	False	False	False	False	False	False
6	False	False	False	True	False	False	False	False	False	False
7	False	False	False	True	False	False	False	False	False	False
8	False	False	False	False	True	False	False	False	False	False
9	False	False	False	False	True	False	False	False	False	False
10	False	False	False	False	False	True	False	False	False	False
11	False	False	False	False	False	True	False	False	False	False
12	False	False	False	False	False	False	True	False	False	False
13	False	False	False	False	False	False	True	False	False	False
14	False	False	False	False	False	False	False	True	False	False
15	False	False	False	False	False	False	False	True	False	False
16	False	False	False	False	False	False	False	False	True	False
17	False	False	False	False	False	False	False	False	True	False
18	False	False	False	False	False	False	False	False	False	True

pred - NumPy array

	0	1	2	3	4	5	6	7	8	9
0	0.951614	5.3317e-09	0.0473781	2.91101e-07	7.21567e-10	5.84692e-06	0.00086776	7.91294e-07	0.000117438	1.58936e-05
1	0.978299	1.48995e-08	0.0179182	0.000328183	5.48121e-09	1.10093e-07	0.000120732	1.63393e-05	0.000892094	0.00242496
2	1.44711e-08	0.950175	0.00235525	4.31379e-05	0.0473064	7.90098e-07	0.000117261	7.20476e-10	1.44711e-08	2.14771e-06
3	1.12529e-07	0.999945	6.14388e-06	2.26021e-06	4.53974e-05	1.52291e-08	8.31483e-07	2.78932e-10	2.06104e-09	4.13971e-08
4	2.02399e-09	4.45813e-05	0.98197	1.10506e-07	1.49554e-08	0.0179854	4.06529e-08	3.00387e-07	7.44584e-10	5.50178e-09
5	0.000852723	0.0465571	0.935125	0.000313699	2.11369e-06	0.0171274	5.7456e-06	1.56182e-05	7.77583e-07	1.05234e-07
6	1.34973e-05	2.47211e-07	4.96537e-06	0.808137	0.0402348	0.00200317	0.109369	2.47211e-07	1.82666e-06	0.0402348
7	5.44774e-09	1.62395e-05	0.000886645	0.972325	0.0178087	0.00241015	7.37271e-10	2.19777e-06	2.00411e-09	0.00655147
8	0.0449461	0.000302844	5.54679e-06	0.000823216	0.902766	0.00608279	7.50676e-07	1.50777e-05	0.0449461	0.00011141
9	0.00657231	0.000120376	8.11087e-07	2.20476e-06	0.975417	2.01048e-09	5.99317e-06	1.62911e-05	0.0178654	1.00096e-10
10	0.017809	1.48087e-08	5.97424e-06	0.00655155	0.000886655	0.972336	1.09422e-07	2.7123e-10	0.00241018	8.08525e-07
11	2.99303e-07	1.00405e-10	8.1359e-07	0.00089221	0.000328226	0.978427	6.01166e-06	0.0179205	0.00242528	1.10107e-07
12	0.000121022	2.21659e-06	4.45213e-05	0.000894235	1.10357e-07	5.49437e-09	0.980647	0.0179612	0.000328971	2.99982e-07
13	1.5536e-05	5.71536e-06	0.046312	2.10256e-06	7.7349e-07	0.000114796	0.930202	4.22311e-05	0.00626765	0.0170372
14	0.0178248	5.97954e-06	0.00655736	1.09519e-07	2.00591e-09	4.02898e-08	8.09243e-07	0.973199	0.00241232	9.98685e-11
15	3.0009e-07	1.49406e-08	0.0179676	1.00669e-10	0.000894556	0.000121065	3.0009e-07	0.981	2.73647e-10	1.63844e-05
16	0.000328985	0.017962	0.000894274	1.10362e-07	0.000121027	2.73561e-10	2.21668e-06	8.15472e-07	0.980691	4.06e-08
17	1.09744e-07	5.46381e-09	2.20426e-06	1.48522e-08	0.000327141	8.10901e-07	0.0178613	0.0065708	0.975193	4.42737e-05
18	8.16502e-07	2.73906e-10	2.21948e-06	6.03318e-06	0.0179847	2.02391e-09	1.63999e-05	1.63999e-05	4.45795e-05	0.981929

pred - NumPy array

	0	1	2	3	4	5	6	7	8	9
0	True	False	False	False	False	False	False	False	False	False
1	True	False	False	False	False	False	False	False	False	False
2	False	True	False	False	False	False	False	False	False	False
3	False	True	False	False	False	False	False	False	False	False
4	False	False	True	False	False	False	False	False	False	False
5	False	False	False	True	False	False	False	False	False	False
6	False	False	False	True	False	False	False	False	False	False
7	False	False	False	True	False	False	False	False	False	False
8	False	False	False	False	False	True	False	False	False	False
9	False	False	False	False	True	False	False	False	False	False
10	False	False	False	False	False	True	False	False	False	False
11	False	False	False	False	False	True	False	False	False	False
12	False	False	False	False	False	False	True	False	False	False
13	False	False	False	False	False	False	True	False	False	False
14	False	False	False	False	False	False	False	True	False	False
15	False	False	False	False	False	False	False	True	False	False
16	False	False	False	False	False	False	False	False	True	False
17	False	False	False	False	False	False	False	False	True	False
18	False	False	False	False	False	False	False	False	False	True

Figure 11: Graphical Explanation of Accuracy versus Number of examples.

To generate a feature vector, convert the duality (angle and distance) into a singularity is the main challenge. For a given angle and length use a preexisting formula, ***lsinθ*** or the sin component. The following table is a sample of generated feature vector for the data obtained from table 1.

Table 1: Calculating the Elements of Feature Vector

	Length(L)	sin(θ)	Lsin(θ)
P1	291.51	0.779	227.85
P2	451.79	0.967	436.79
P3	454.00	0.999	453.54
P4	195.01	1.000	195.01
P4	431.14	0.976	420.75
P5	366.03	0.895	327.59

There are cases where we have found that the number of data points are unequal. As a result, the maximum possible number of data points in a gesture from a given set of gestures is considered as the length of the feature vector array. In case of any gesture having less number of data points than the array size, the empty spaces are filled with 0. This is not analogous to padding with 0, because here it means that the figure has more fingers which are at an angle of 0° or 180°. The feature vector for any gesture is highly distinguished from another gesture.

Once ample number of feature vectors have been obtained, then we can proceed to create our feature file for training the machine. The matrix of features can be created using the feature vectors from the training set. We then train the machine with our matrix of features using Artificial Neural Network and compare the accuracy of each on a common test set.

Data obtained from ten images each of ten alphabets are used for training and testing. Learning is done using ANN classifier. For training and testing we have split the data in 8:2 ratio for each alphabet. Our training set consisted of 10 pictures each of 10 alpha-numeric gestures. Thus we have used 80 images in total to train the machine and further 20 to test the accuracy. Our training included the deployment of cross-validation technique to obtain more precise results. Cross-validation randomizes the data used for training and testing respectively.

5. Results

The aforementioned algorithm has been put to test and has yielded an average accuracy of 92.5%. In reference to the results generated by the code, the confusion matrix of a couple of outputs have been mentioned with resultant prediction values.

In the above stated figures, test 1 has revealed an accuracy of 100% stating that all images of the hand gesture has been successfully classified. The confusion matrix indicates true value for the hand gestures it has predicted for a given feature vector and false for the rest. In the second test the algorithm has missed classified two figures and hence it can be noticed in the confusion matrix where the algorithm has wrongly predicted true for hand gestures numbered 2 and 4 respectively. Additional tests have been conducted in order to obtain consolidate the rate of accuracy.

Table 2: A Comparison between the Proposed Frameworks with the Available CNN

	Proposed ANN	Equivalent CNN
Training Time	3 minutes for 100 sample	18 minutes for 100 samples
Prediction Time	1 millisecond	10 milliseconds
Feature Vector Length	10	255 X 255

The above figures strengthen the advantages of the proposed algorithm over existing CNN or Convolution Neural Network algorithms. Sample training shows that on an average 3 minutes are required for training our algorithm in comparison to CNN which takes roughly 18 minutes for the same number of images. As stated before, storage requirements are highly minimized due to the miniscule feature vector size and the number of images required to obtain adequate accuracy. Thus, for a considerably lesser number of images the algorithm generates acceptable accuracy in contrast to existing algorithms.

Thus the aforementioned framework is readily available for application on computation constricted environment in public sectors and also in absence of high data processing infrastructures.

An average accuracy of 92.5 % has been secured by our algorithm while observing each cycle of training and testing set obtained using cross-validation technique. In comparison to existing Convoluted Neural Network models (CNN), it requires substantially less data and produces acceptable accuracy. Further CNN is computationally expensive model of prediction in comparison to ANN model coupled with distinctive feature extraction.

6. Conclusion

Our technology and technique for gesture recognition has provided significant improvements over previously worked on methods in terms of accuracy and reliability. The disabilities experienced by individuals who are incapable of speaking are overcome with the help of this machine which eradicates the need of another individual for translation of the sign language to the commonly used English language. We have trained our system

with considerably less and manageable sized dataset consisting of various gestures of every letter in the English alphabet. There are some drawbacks, however, in the fact that the process of recognition of a gesture must be carried out with the assistance of a glove without which the creases hinder with the accurate depiction of the hand model and consequently results in imperfect polygon approximation. Also, the process must be carried out in front of a uni-coloured background to ensure that the outer contours of the hand are well defined. But with better and elevated image preprocessing techniques the need for gloves and background may be eradicated. We hope that our research will help the unfortunate ones become socially acceptable and can express themselves spontaneously and freely.

References

[1]   Y. Xu, Y. Dai, "Review of Hand Gesture Recognition and Application"; Contemporary Engineering Sciences, Vol. 10, 2017.

[2]   T. Messer, "Static hand gesture recognition." University of Fribourg, Switzerland (2009).

[3]   J. S. Kim, W. Jang, Z. Bien, "A dynamic gesture recognition system for the Korean sign language (KSL)." IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 26.2 (1996).

[4]   M. H. Kim, Y. H. Joo, J. B. Park, "Emotion Detection Algorithm Using Frontal Face Image"; ICCAS2005, June 2-5, KINTEX, Gyeonggi-Do, Korea.

[5]   D. A. Chowdhry, A. Hussain, Rehman, Z. U. R. Muhammad, "Smart Security System For Sensitive Area Using Face Recognition"; 2013 IEEE Conference on Sustainable Utilization and Development in Engineering and Technology (CSUDET).

[6]   C. A. Pickering, K. J. Burnham, M. J. Richardson, "A Research Study of Hand Gesture Recognition Technologies and Applications for Human Vehicle Interaction", 2007 3rd Institution of Engineering and Technology Conference on Automotive Electronics, 28-29 June 2007, Warwick, UK.

[7]   P. Pandey, V. Jain, "Hand Gesture Recognition for Sign Language Recognition: A Review"; International Journal of Science, Engineering and Technology Research (IJSETR), Volume- 4, Issue- 3, March 2015.

[8]   W. P. Clark, "The Indian Sign Language: With Brief Explanatory Notes of the Gestures Taught Deaf-mutes in Our Institutions for Their Instruction and a Description of Some of the Peculiar Laws, Customs, Myths, Superstitions, Ways of Living, Code of Peace and War Signals of Our Aborigines", Philadelphia, Pa.: LR Hamersly & Company, 1884.

[9]   W. C. Stokoe, "Sign language structure." (1978).

[10] V. Sumalatha, R. Santhi, "A Study on Hidden Markov Model (HMM)", International Journal of Advance Research in Computer Science and Management Studies, Volume 2, Issue 11, November 2014.

[11] K. Tang, T. Woo, "Algorithmic aspects of alternating sum of volumes. Part 2: Nonconvergence and its remedy",

[12] T. Starner, J. Weaver, A. Pentland, "Real-time american IEEE Transactions on Pattern Analysis and Machine Intelligence 20.12 (1998): 1371-1375.

[13] K. Selvakumar, "A Review on Polygonal Approximation Techniques of the Digital Planar Curves", International Journal of Pure and Applied Mathematics, Volume 117 No. 17 2017, 85-90 ISSN: 1311-8080 (printed version); ISSN: 1314-3395.

[14] Y. S. Kim, "Recognition of form features using convex decomposition", Computer-Aided Design 24.9 (1992): 461-476.

[15] J. M., Lien, N. M. Amato, "Approximate convex decomposition of polygons", Proceedings Of The Twentieth Annual Symposium On Computational Geometry, ACM, 2004.

[16] C.J. Bajaj, T. K. Dey, "Convex Decomposition of Polyhedra And Robustness", SIAM Journal on Computing 21.2 (1992): 339-364.

[17] M. Attene, B. Falcidieno, M. Spagnuolo, "Hierarchical Mesh Segmentation Based On Fitting Primitives", The Visual Computer. 22. 181-193. 10.1007/s00371-006-0375-x, (2006).

[18] K. Mamou, G. Faouzi, "A Simple And Efficient Approach For 3D Mesh Approximate Convex Decomposition", Image Processing (ICIP), 2009 16th IEEE International Conference on Image Processing, IEEE, 2009.

[19] J. Rekha, J. Bhattacharya, S. Majumder, "Shape, Texture And Local Movement Hand Gesture Features For Indian Sign Language Recognition", Trendz in Information Sciences and Computing (TISC), 2011 3rd International Conference on Trendz in Information Science and Computing, IEEE, 2011.

[20] M. R. Ahsan, M. I. Ibrahimy, O. O. Khalifa, "Electromygraphy (EMG) Signal Based Hand Gesture Recognition Using Artificial Neural Network (ANN)", 2011 4th International Conference on Mechatronics (ICOM), Kuala Lumpur, 2011, pp. 1-6. doi: 10.1109/ICOM.2011.5937135.

[21] Li, Yuqian, Su, Guangda, "Simplified histograms of oriented gradient features extraction algorithm for the hardware implementation", 2015 International Conference on Computers, Communications, and Systems (ICCCS) IEEE Xplorer, 2015.

[22] Z. Ren, J. Yuan, J. Meng, Z. Zhang, "Robust part-based hand gesture recognition using kinect sensor", IEEE transactions on multimedia 15.5 (2013): 1110- 1120.

[23] H. Cooper, B. Holt, R. Bowden, "Sign language recognition", Visual Analysis of Humans, Springer, London, 2011. 539-562.

[24] Plataniotis, K.N.; Venetsanopoulos, A.N.; "Color Edge Detection. In: Color Image Processing and Applications", Digital Signal Processing, Springer, Berlin, Heidelberg, 2000.

[25] E.K. Burke, R.S.R. Hellier, G. Kendall, G. Whitwell , "Complete And Robust No-Fit Polygon Generation For The Irregular Stock Cutting Problem", European Journal of Operational Research, 179 (2007) 27–49.

[26] P. Prosser, "Exact algorithms for maximum clique: A computational study." Algorithms 5.4 (2012): 545-587.

It Takes Two to Tango: Merging Science and Creativity to Support Continued Innovation in the IoT Domain

Helen Hasenfuss[1, *], Muftah Fraifer[1], Sameer Kharel[1], Asma Elmangoush[2], Alan Ryan[1], Walid Elgenaidi[3]

[1]CSIS-IDC, University of Limerick, V94 T9PX, Ireland

[2]The College of Industrial Technologies-CIT, Misurata, Libya

[3]Optical Fiber Sensors Research Center-OFSRC, University of Limerick, V94 T9PX, Ireland

A R T I C L E   I N F O

A B S T R A C T

The Internet of Things (IoT) is a new concept that has a great appeal for researchers, businesses and the ordinary tech user. It presents new possibilities of connection between devices and people and it stimulates our need and desire to interact, to exchange ideas and to communicate with the surrounding environment. This paper briefly explores aspects of the IoT that may be important for future developments and focuses on the impact of creative methodologies, such as user centered design (UCD) in a smart parking, IoT-based system prototype. It aspires to provide an alternative perspective for an improved user experience (UX) and aims to contribute to the discussion about the challenges, findings and perspectives when merging science and creativity to maintain continuous progress in the IoT domain.

1. Introduction

This paper is an extension of the article Look before you Leap presented in the conference IEEE 3rd international Forum on Research and Technologies for Society and Industry [1]. In that paper the complex system which was explored was a smart parking system through the application of the Internet of Things (IoT). It briefly discussed issues that relate to the IoT domain such as fundamental requirements, access technologies, security and privacy and data handling. The primary difference of approach in the development of this particular iteration of a smart parking system was the direct involvement of the users. The inclusion of users and/or the end users is a clearly defined methodology within design or human-computer interaction (HCI) disciplines, referred to as user centred design (UCD).

In order to preserve the quality and integrity of technological innovation it is important to combine methodologies from the science, technology, engineering and maths (STEM) disciplines and the Arts disciplines. A discussion of the methodological interplay between these diverse disciplines, as well as an attempt

to expand on the challenges that emerge in the application of UCD-based methodologies into industry driven projects, are the two primary objectives of this paper. The IoT is a network of connections between digital and physical data / objects and providing a seamless user experience for IoT products is becoming an integral aspect of product design. Improved user experience (UX) is not always straightforward, as UX designers need to be knowledgeable about both the visible and invisible layers of IoT (e.g. visible layers: user interface, physical form factor and hardware spec; invisible layers: enabling architecture and technology backend, end user targets, etc.). The smart parking system that is discussed aims to highlight the impact of the involvement of the user in the design process. As IoT projects do not deal with singular issues of society (e.g. interface design, ergonomics, computing hardware, production quantity, etc.), it is necessary to shift some focus into developing multidisciplinary experts that can communicate between diverse disciplines (e.g. product design & psychology) but that are also capable of translating between academic research and industry requirements. The growth and interest in the IoT is substantial as it is estimated that in 2020, 20.8 billion devices will be connected via the Internet, according to Gartner, Inc [1]. Approximately two billion people

*Helen Hasenfuss, University of Limerick, Helen.Hasenfuss@ul.ie

worldwide connect to the Internet in order to avail of the variety of services ranging from playing online games, social media, commerce, education, networking sites, entertainment and many more [2].

There is great potential for the concept of the IoT because it is expanding into and incorporating a variety of diverse research areas, see Figure 1. This expansion is advantageous because of the potential advancement of peripheral technologies, e.g., developing more efficient infrastructure to improve digital and physical resource management, energy sources for nanotechnology or advancing sensor technology.
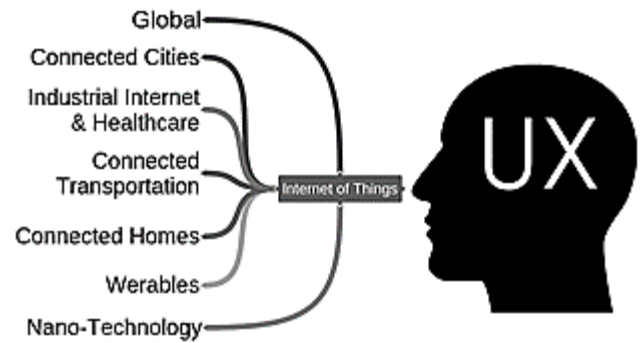
Section two, the related work, will summarise the key issues that were discussed in the related conference paper [1]. These relate to the challenges, requirements, and applications that can influence the accessibility of the IoT, especially from a UX perspective. The continuous evolution and refinement of the IoT is reflective of how technology in the computer industry developed. For example, the physical hardware of the first computers dictated the design, layout and user interaction style of the IoT. Once technology became established and grounded, researchers began to consider the user and how they interacted and engaged with the technology. A natural progression from this consideration was for the user to be directly involved in the design process itself. Since the importance and value of the user is now established, the application of UCD methodologies alongside technology design processes ensures that adaptable and sophisticated IoT products / tech can be developed without needing to apply a singular approach to design.

The user and their experience are key elements in the development of the smart parking system [1]. An analysis of why it is important to consider the user in future technological is discussed in section three. The challenges of implementing more creative methodologies such as the user centred design, agile method into industry-based projects is illustrated and the value of their contribution is explored.

In the fourth section, the smart parking system that was developed in the Interaction design centre (IDC) in the University of Limerick (UL) between 2013 and 2017 will be discussed. The approach as to how the implementation of the UCD method influenced the development of the technical and economic aspects of the project are highlighted.

The fifth section is a discussion that briefly considers elements that affect IoT based projects but also analyses whether the re-combination of STEM and Arts methodologies can make a difference in the approach to developing long term and lasting technological innovation. In an instant gratification society, endurance is becoming a rare characteristic even though it is still widely sought. The question of whether quantitative or qualitative research has the greater impact is explored from the perspective of science and artistic disciplines.

2. Related Work

It has become well known that technical challenges of IoT are diversified. In this section, we classify and identify some of these challenges.

Figure 1. The influence of the IoT on local to global landscapes

2.1 Security and Privacy

One of the major concerns about IoT is security, as heterogeneous objects related to multiple domains could be connected to the Internet and communicate with various application services. Users worry that unauthorized third parties may be able to get access and/or control over their appliances and misuse data from the personal or Industrial applications [3]. Potentially, sensitive data will be roaming via the global Internet. Furthermore, most integrated components within an IoT system are characterized by low power and computation capabilities, and therefore cannot implement complex security mechanisms. IoT holds the potential of connecting everything and everyone, while not everyone should have access to everything.

2.2 Access Technologies:

Access control is necessary as well as data authentication in new communication technologies have been introduced for IoT with the objective of simplifying the deployments and gaining wider network coverage without compromising cost or power consumption. New technical solutions have been implemented that present an IP-based backhaul where the end-devices are provided with direct connectivity to a base station implementation a long-range wireless technology such as SIGFOX (SIFOX 2018) or short-range wireless technology such as Zigbee [4]. The use of the long- or short-range wireless technology depends on the application domain and the operational requirements. Table 1 summarizes the main characteristics of some other IP Protocols widely used in IoT projects. Many of these technologies are used in a variety of domains including home automation, sensor applications, and smart grids. Interested readers are referred to Naito's research for further discussion and insights of listed technologies for IoT [5].

2.3 Compatibility and longevity

The existing IoT deployments are growing in many different directions, with many different technologies competing to become the standard. Many Standard organizations have adopted a Representational State Transfer (REST) based architecture in their IoT framework such as ETSI and OneM2M, while few others have used the SOAP such as IEEE1888 [1,4]. In addition, more data protocols such as CoAP, XMPP, and MQTT are developed to address the requirements of integrating resource-constrained devices and supporting ubiquitous access.

It can potentially cause complications and might involve the deployment of additional proxies and interworking software when connecting devices.

Table 1. IoT standard and Protocols

Protocol	Max Range	Max Data Rate (Kbps)	Life of Batteries	Max Num of Devices / base station	Topology
Sigfox	50Km	1	20 years	~1Mil	star
Neul	10Km	100	10-15	n/a	star
LoRaWAN	15Km	37.5	10	~10K	Tree
Z-Wave	100m	100	Few years	232	Mesh
Zigbee	70m	20/40/250	Few years	2-6K	star
Bluetooth low-Power (BLE)	100m	1M	10-15	7	star
6LoWPAN	200m	250K	Few months	100	Mesh

Figure 2. IoT fields

2.4 Fundamental requirement for IoT

IoT may also be a profitable venture for users in various businesses. It could potentially enable systems and smart manufacturing by connecting machines so that an industry can create networks along the entire life-time chain of a product or service [1, 6]. The Telecommunication Sector of the International Telecommunication Union (ITU-T) has specified high-level requirements of IoT in Y.2060 recommendation (ITU-T 2012).

2.5 Applications

IoT applications are now starting to integrate various domains such as personal, home, enterprise, utilities and environment, as illustrated by Figure 2.

Many services and applications have been developed, for example, in the medical domain with ubiquitous health care using sensors and body area networks to upload medical data to servers. Such home monitoring systems enable family members or community members (e.g. neighbourhood groups) to be notified about the condition of a person. This can potentially reduce the cost of health-care through pre-emptive action which can lead to an improvement in community awareness and care [7].

3. User experience challenges

IoT is growing every day and focus has been on technical aspects such as IoT infrastructure requirement, operating systems, security, power consumption, data, and storage. These technical aspects are relevant factors from an industrial perspective. However, involvement of users in any stage of development process has a positive impact on any product and improve the UX of the product. In an IoT ecosystem, different devices communicate with each other, possibly with different protocols. Different protocols and interfaces of devices have added complexity for users in its use in its ecosystem [8].

The tug of war between UX and user interfaces (UI) is an additional challenge to UX designers to provide seamless user experience without compromising security (Maiman 2015) [9]. Unification of interfaces to provide seamless experience is most challenging for UX designers in IoT. An interoperable design that can be easily adapted to the complexities of an IoT ecosystem is crucial to overcome this problem. Unification of interfaces across different appliances seems to be the biggest challenge for IoT user experience designers, as IoT products consist of different visible layers over a number of invisible layers, and designing a uniform experience requires much consideration (Janaway 2016, Staff 2016). While IoT UX can also be improved once IoT products are in market by getting feedback from users, changing physical design may not be straightforward due to its cost [9]. UX designers need to understand the different IoT invisible modules/layers that play a part in user interaction, like implementation of hardware, connectivity, data as IoT products are built on top of an IoT platform. Another challenge in interface design is that brands can have different requirements for the same software when used on different devices for example Android and iOS devices. Designing a UI for the IoT can be challenging since it, in many cases, is highly interconnected with other products, systems and services which affects the users' perceptions of their experiences.

The IoT can have a positive impact in the life of many citizens (users). It can be used to study human behaviour and understand how technologies are used in a daily life and this study can be useful for further IoT technology design. Involvement of the user via user centred design (UCD) in IoT design can't be ignored. The majority of IoT projects are still based in a lab setting, however ideally testing should be performed in real scenarios with real end users to find real problems [10].

3.1 Activity Centred Design (ACD) vs User Centred Design (UCD)

It is argued that the concept of UCD is not relevant in case of IoT. Cruickshank, L. and N. Trivedi suggest that the human should be seen as one of the smart objects in the IoT ecosystem and should fit into an activity centered design (ACD) approach. Cruickshank, L. and N. Trivedi argues that overall, the contribution of non-human elements is significantly higher than purely human elements in IoT systems. The discussion in academia relates to whether design should be able to keep up with emerging new agency rather than just try to satisfy users/customers [11].

In considering the ACD concept, we shouldn't forget that we are not trying to build robotic world but create robots for facilitating their (user/customer) daily life, maintaining privacy and without having tension between device and human. The results from Memedi's, research clearly shows that UCD plays important role in IoT. They found UCD as a helpful tool for identifying the requirement of the system user. For example, in a system developed for Parkinson's disease, patients it was found that 70% of users reacted positively to a IoT-based system prototype, (connected devices such as a wrist sensor, smartphone, bed sensor and an electronic dosing device). Positive impact of the system is due to including user in application/system development [12]. It was found that challenge of communication has been one of major component for hindering implementation of UX in the IoT domain. It was found that users are concerned about communication between device invisibly and users are unaware between information shared between it and different objects. It was also found users are interested in whether value is added to their life by connecting devices. Sometimes the complexity of IoT device configurations can also effect UX [13].

There are a variety of aspects that influence consumer acceptance of IoT systems and products. Factors that directly affect UX, range from tangible to nonphysical issues. For example, device memory and power: using many different IoT apps can be a drain on mobile devices or if it takes longer to complete a task using an IoT app than it would if it were done manually. Errors in IoT products can be detrimental to user experience, because the user can become irritated and generate more negative affect (see subsection 5.4). As a result, the user could simply switch to a similar product provided by an alternative company. Alternatively, qualities such as usefulness and trust have varying degrees of priority in relation to acceptance of IoT technology but are primarily dependant on the interaction and perception of the user.

Poor user experience for many IoT products is often due to ineffectual user-testing during the design process, leading to badly designed interfaces of applications and devices. For example, the smart heating system designed for supporting energy-saving lacked in all metrics used for its UX evaluation [14]. Users experienced difficulty accomplishing their tasks and the system also had inappropriate labelling [14]. Current trend of software development method has been developing rapidly, iteratively and incrementally and delivering modules of products to customer in short timeframes. The process is popularly known as "Agile".

3.2 How practical is Agile in conjunction with UCD?

The IDC in the University of Limerick, conducted interviews with 10 IT professionals working in software companies as UX experts,

designers and software developers from Ireland, Nepal and United State of America from August to October 2017. Every interviewee agreed that in some way they were following the Agile method and among different Agile methods "Scrum" is used in most cases. Agile is an umbrella term used for iterative and increment software development methodologies. Scrum is one of the sub-methods under Agile. In Scrum in every two to four weeks one complete module of the product is delivered to the customer which is known as a sprint.

From the above paragraphs it is clear that UCD should ideally be a mandatory concept in the development of user friendly and seamless UX in IoT devices. Brhel, M., et al., Butt, S. M. and S. M. Butt, Sohaib, O. and K. Khan have been writing about integrating the concept of UCD into the Agile development method [15-17]. A fundamental difference between agile development method and UCD is that the former focuses on the customer and the latter on the user. User and customer are almost the same, but every customer may not be an end user. For example, from a business perspective, everyone is a customer but, in our case, "user" refers to specific or targeted end users of a particular product, i.e. for a smart parking system based in UL – drivers who knew the campus grounds and administration were the appropriate users. To improve UX, suggestions have been made to use the Agile methodology to develop IoT devices in conjunction with UCD [18].

Regarding how often the software industry actually uses the concept of UCD in the development of new products, from the previously mentioned interviews of 10 IT specialists, we learned that incorporating UCD is not a straightforward process. The IT specialists agreed that implementation of UCD depends upon,

A) budget of the project,
B) requirement of customers and
C) that it is the responsible of the customer to know about their users. (In this case customer refers to an organization who enlists the services of a company that specialises in developing IoT products and systems.)

In [19], the paper proposes that usability testing is the least preferable testing method compared with unit, integration, system and acceptance testing as this testing are done by technical experts rather than involving end users. Mostly usability testing is conducted by software engineers rather than usability or UX experts. They indicate that customers give less priority to usability testing rather than acceptance testing, alpha and beta testing. Lack of usability is often attributed to the following aspects: 35% due to lack of time, 20% due to lack of training/knowledge, 15% lack of budget and rest for other reasons [19]. It state that *"companies generally do not give priority to usability and UX evaluation"* and including the user in software development is taken as a burden or delay in time and increased cost for relatively small benefit [19]. Different UCD tools have be implied in the Agile methodology for usability and UX. The usability tools vary slightly according to the discipline in which they are applied. For example, in requirements engineering, ethnographical observation, card sorting, personas, task scenarios, scenario and storyboards, and prototyping are the primary tools used to evaluate usability. In the design discipline: screen snapshots, product style guide, navigation maps are used. In the evaluation process, heuristic evaluation, cognitive

walkthrough, thinking aloud, and lab usability testing are implemented. Then, an evaluation by experts, well-designed usability tests and analysis of installed systems are processes used as defined by UCD or have been slightly modified according to project demand [20].

The process of developing a smart parking system by including the relevant users or stakeholder and the positive outcome of the project highlight that participatory- and user centred design methodologies are invaluable for the overall design process but in particular the early stages of project development. This postulation is supported through diverse academic literature [10, 12]. Consistency, intuitive design and good, logical mapping are just some of the cornerstones for the successful uptake of any design as illustrated by Don Norman in his book "The Design of Everyday Things" [21]. Both lab and field testing need to be performed on IoT devices with users. Lab testing is helpful to improve the design of products and field testing is helpful to improve the user experience. Different user experience evaluation techniques, such as contextual laddering, day reconstruction, experience sampling and UX curves, can be applied as per requirements.

3.3 Discussion of identified research gaps when applying UCD

One of the basic steps in using the UCD method is to know everything related to the existing problems of the system expected to be designed. In order to build up this knowledge base, initial research should be carried out on other systems that have the same purpose. Whilst research indicates that there is a need for smart parking system solutions, there were no previous studies published in the literature relating to the assessment of these systems that also take into account the views of users. As well as this research indicated that each solution had different technological requirements depending on which factors received the highest priority, e.g. cost effective, integration possibilities, interface, etc. Each prototype system has advantages and disadvantages in terms of the following criteria:

- cost
- reliability
- scalability
- accuracy
- communication type
- circuit complexity
- reliability
- method of operation
- ease of installation
- usability of the system

There is an inadequate research base to fully inform how to design and evaluate smart parking systems for users.

4. Challenges and findings from our perspective in the IoT domain

UCD is a broad term to describe a design process based on the integration of potential users and since this era is witnessing a huge technological revolution of future networks depends on the connecting different objects from watches, cars, to satellites. There is no question about it: technology has revolutionized every part of our life nowadays as the Internet is likely to have a dramatic impact on our daily lives as it evolves into an essential part of many systems such as smart transportation, and smart parking etc.

The dialectical question that exists among those interested in the IoT domain remains: do we really need to link all things to each other? Because the different answers to the same question have advantages and disadvantages to the benefits of that linkage. However, the stakeholder is the only one to be able to estimate the extent to which he can benefit from the possibility of linking different things together to achieve the desired goals. Enabling technology from the things around us to be able to interact with each other and with the user, has many advantages and disadvantages that require careful consideration before any system is designed to deliver its services in a meaningful manner. We see an urgent need to have a harmony between merging science and creativity to support continued innovation in the IoT domain.

Certainly, exploiting the things around them and looking for alternative solutions to improve the level of services that users need is a great challenge, especially with regard to longevity and sustainability. This will create solutions for many dilemmas and problems that people face in general in various aspects of life.

4.1 IoT based smart parking system

Fraifer et al., designed a smart system that helps to find parking spaces at the University of Limerick. The difference here is that they incorporated the user into the design process in an early process of user-centered design [22]. Three groups of users were used to explore the problem of finding parking on an enclosed but spread-out area and to design a series prototypes to resolve, or alleviate, the issues. These issues include,

- reducing pollution by reducing the amount of time drivers spend looking for parking spaces,
- finding spaces hidden by larger cars,
- real-time update of free parking spaces,
- discerning peak times for searching, parking and when spaces are freed up, etc.
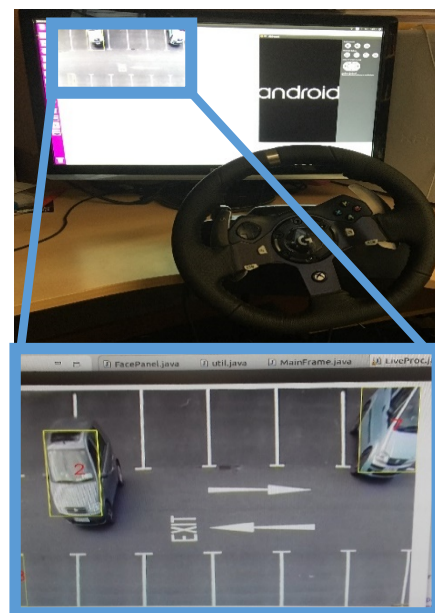
Figure 3. Illustrates how the CCTV footage was integrated in order create a search algorithm (left) and smart parking app (right of screen).

The basis of the design implements CCTV which is often installed in any parking lot for the purpose of securing and preventing theft making it a viable economic solution. The user interacts with the system through a specific application in which they can see which places are vacant and where they can book a parking space [22], see Figure 3. The study has presented a design development of an application of IoT technologies. This has provided some findings that challenges the traditional approach to designing IoT based systems, e.g. how to make abstract technologies connective and interactable. Such findings helped to identify key areas of future research and intervention. They also support the argument for the need of new ideas and solutions that are mindful of the driver's requirements and how to improve the potential IoT smart parking system. Below is analysis data and finding discussion from two iterations of UCD approach.

4.2 UCD methodology

The UCD methodology frames the phases throughout a design and development of the prototype on gaining a deep understanding of those who will be using the system.

- **Assigning the context of use:** 12 members of the UL community participated in the smart parking project. These users were on campus on a daily basis and regularly experienced issues in finding parking. They include students, staff and frequent visitors
- **Specify requirements**: The users were interviewed and asked to complete a questionnaire regarding the primary issues they experienced in their normal parking routine. Within this phase it is possible to discern the major issues that should be addressed throughout the project but also what the users envisioned could potentially alleviate these issues.
- **Create design solutions:** Within this phase the majority of design work takes places and the actual product comes to fruition. For the smart parking project an app was developed that would enable users to find and reserve parking spaces cutting down on the time and fuel spent cruising around the large campus in order to find a parking space.
- **Evaluate designs:** This last phase is the crucial difference a UCD approach can make to product development. The evaluation procedure can be applied often throughout the design process and does not always require the full number of users. For example, once a low fidelity paper prototype was developed, that sufficiently conveyed the idea of the app, it was possible to evaluate it with some of the users. After this evaluation, the requirements, solution and evaluation phases can be continuously re-applied, until the final product becomes more refined and of a high-quality finish.

4.3 Findings from designing an IoT based smart parking system at IDC

The data analysis was based on qualitative findings from a human-computer interaction perspective. From an user perspective, there is an inadequate research base to inform fully how to design and evaluate smart parking systems for users; therefore, this issue requires further investigation. A summary of the results and findings of research contributors from IDC are detailed in the following subsection 4.3.1 and 4.3.2.

4.3.1 First iteration of UCD

One of the most important benefits of this phase is the discovery of errors in the initial design of the system that would have led to failure if it had been left undetected. In addition, using different categories has main important benefits that has had beneficial results to redesign the prototype according to recommendations from users. The aim was to know the basic dilemma facing drivers while searching for a place to park. Other ideas relate to the development of solutions and ideas that may be part of the system.

4.3.2 Second iteration of UCD

The second iteration was used by a small group (focus group) consisting of 3 users. They were selected to perform a second session of UCD regarding the improvement of the system after the previous proposed modifications. This stage leads to the emergence of proposed system as a supposed final product. One of the findings in this stage and through the discussion with users, is that the system must include several things that contribute to solving the dilemma of finding a parking space. Below are the findings based on the data collection when applying the UCD approach. The initial model was an experiment in an indoor environment. The car park was designed using cardboard, where the camera was set up to face the simulated parking lot.

- The first idea was to draw circles with dark colours on every parking lot in order to facilitate the process of identifying whether the position is vacant or not. The system depended on computer vision, and it is easy to identify shapes by using the functions available in the OpenCV library.
- After presenting the model to the users, some of them made the comment that the system may not work in the event of snowfall as the circular shape is covered in winter by snow or autumn leaves. Users came up with the idea using actual cars, as that would be analogous to rectangles, using computer vision to detect if the parking lot is vacant.
- With regard to the designed mobile application, which is a fundamental element of the system, think-aloud sessions with users and two iterations of UCD enriched the appearance and usability of the model.
- The general tendency of participants was to make the application simple, but 9 users believed that the application would be more effective if it would show parking without activating reservation. They believed that using the application while driving may pose a risk to the driver. Enabling the driver to know the available vacant positions is in itself a good solution.
- To navigate directly to the vacant space to save time, fuel and mental effort, which is a basic goal of the system.

The primary concern of the 9 participants was that the system does not contain barriers to guarantee a parking space and gave the impression that the general system may take some time to function properly. Some users suggested activating the possibility of booking a specific space. In the cases where there is a fee for booking the space, the administration would be responsible for the space, should someone take it without a reservation. Developing scenarios often identifies important aspects of using a product in the real world that may not otherwise identified and considered in a lab setting, for example, the workflow being disrupted by a phone

call. Scenarios are useful throughout the design process, particularly when developing task descriptions for usability testing.

4.4 Findings from designing an IoT based smart parking system at IDC

The primary findings from the smart parking system at the IDC relate to the infrastructure which the system uses and how the users felt about using the system. The infrastructures (Wi-Fi and CCTV cameras) which this system used enabled a smooth integration and reliable service. Internet access is widespread throughout the campus and CCTV is used for campus security. In general, the users were happy with the system and the app that was developed because it was easy and quick to use and did not impede their searching process. Advantages of this system are the ability to find available parking spots, especially when they are not visible from a driving position, and that the number and location of free spaces are updated in real time. Table 2 summarizes the most important aspects that were defined by users in the 2nd round of evaluation [27].

Table 2. Findings from 2nd iteration of evaluating the smart parking system.

Source	Needs/Requirements/Features/ Solutions/ /Ideations	Findings
Student	Using Google Maps API library services to determine the destination accurately, especially in large parking areas. The possibility of seeing the position of the car in a real way by connecting the system to live broadcast, perhaps through live images.	Providing a precise positioning service such as Google Maps will be better for the user (user - 01 Second session) The availability of a service enables the user to see parking spaces in real form in the form of images that have a positive impact on the user when he/she arrives (user - 01)
Visitor	Create a database for users by asking the same user to enter his/her favourite times to take advantage of the vacant places for others if possible.	The availability of database users will help administrators to provide available vacant spaces smoothly, (user - 02)
Staff	The possibility of sending short messages or notifications when a user leaves a place before the end of the reservation. Provide information to the user about each position on the application, for example a counter in descending order to make the rest of the users aware of the nearest position of silent vacant	Enable the users to add the SMS of informing others that user leaves the pre-booked place before the end of the reservation. (user - 03) The availability of a digital counter in descending order showing the end of the booking represents a beautiful feature to remind the user, (user - 03)

5. Discussion

One challenge of technological innovation is its uptake into the public domain. The flexibility and adaptability inherent in the IoT concepts presents a myriad of new opportunities to find eloquent solutions as technology becomes more sophisticated. However, user integration into the technology design process is still a major

obstacle. After a company identifies the appropriate target user group, which in itself can be challenging, extracting the relevant user data can be a deterring factor for the majority of companies. Collecting user data is often time and resource consuming and it is necessary to translate the user's requirements and opinions into data that a company can use. As has been highlighted throughout this paper it is necessary to,

A) Engage the users from early on in the design and development process for any IoT potential products or services,
B) Adapt the existing business models to support cost-effective services, [1]

In addition to adapting existing models in industry, it is also necessary to consider the methodologies used in academic research. Whilst this concept is the primary focus in the discussion section, subsections 5.1 to 5.4 briefly summarise previously addressed issues [1] in order to provide context for the extended discussion on the merging of science and creativity to continuing progressing technological innovation. These issues are as follows:

- Business requirements
- Existing infrastructure
- Global connectivity and networking
- Emotion-driven design

5.1 Business requirements

The IoT refers to uniquely identifiable objects and their virtual representations in an Internet-like structure. Special requirements are needed in order to manage, secure and facilitate these embedded software assets. These systems are considered an industrial subset of the IoT, which may include billions of interconnected sensors, devices and systems, many of which will communicate without human involvement [1].

5.2 Existing infrastructure

The global and expanding nature of the IoT means that the viability of the network in its entirety is an important factor. For such a dynamic system to continuing growing and expanding it is necessary to consider how this viability can be maintained and is affected by the diverse number of variables of which it is comprised, i.e. the system's stability, the existing infrastructure and availability of resources (bandwidth, storage capacity, IP address space, etc.). Similar to any human endeavour having a good grounding or baseline is important, e.g. grammar enables a person to gain command of a language and to understand it. In relation to the IoT, since it is built upon and using the existing infrastructure of the Internet it is necessary to keep reviewing the efficiency of the technologies involved. For example, if an older infrastructure can no longer support new technologies because either conceptual or physical compatibility between technologies is no longer possible, too many resources may be required in order to maintain it. This action would shift the focus from development, i.e. looking forward, to maintenance, i.e. being tied to the present and past. It correlates with Mark Weiser's description that a "good tool is an invisible tool." [23] For example, becoming conscious of a failing infrastructure (glitches, connectivity issues, slow connection, limited development opportunities, etc) detracts from an actual IoT device or system. A good tool ensures that the user's focus is entirely on the task.

5.3 Global Connectivity

The IoT is a network of connected devices and a key issue of connectivity is compatibility. To accommodate the question of compatibility raises questions regarding the concept of standardisation. In a system that is inherently built to accommodate diversity, this approach appears to be contradictory. Even though standards can ease the element of interaction with technology (e.g. the classic floppy disk as the save icon), it is necessary to maintain a balance between system diversity and strength. For example, allowing systems and devices to be compatible with each other can create an IoT network that is stable in the event of infrastructure degradation or failure due to physical factors, (e.g. power outage, out of range, no connection, no storage, etc.) [24]. Allowing the method of connectivity to be diverse can ensure that the whole system is stronger against attack and more adaptable. The latter option is also more analogous to the method in which groups of people interact with each other, e.g. language dialects.

Considering the IoT development from human aspects, such as communication, growth, and development, the interdependent nature of this type of network or system becomes evident. The IoT aims to be integrated into an environment that is inhabited by social, living entities. As a result, logic and reason must interact with emotion and unpredictability, e.g. technical developments are directly influenced by user experiences and vice versa. The social and communicative aspects that are inherent in the IoT system creates spaces for large as well as small communities to work together in order to generate smaller networks that can eventually be integrated or expanded further. People can gain a sense of "socio-technical responsibility" (de Rosney 2014) and can break down knowledge barriers.

5.4 Emotional design

The social aspect mentioned in section 5.3 highlights an intriguing challenge particularly from a design perspective: the unpredictability of human nature. The ideas put forward in phenomenology illustrate how multifaceted a single experience can be. Aside from design considerations such as ergonomics, emotion and thereby experience generation and perception are critical elements in the design process. In the early years of computer development, technology defined user interaction styles. Nowadays the awareness of the importance of designing for an experience as well as for efficient technology is the progress of technological development that is embodied in the IoT. Previously a user had to adapt their behaviour according to the technology. As the focus in industry shifted from pure technology development to aesthetics and user experience it was possible to see an adaptation in the design process. Instead of primarily emphasising the engineering and scientific methodologies, it became necessary to consider socio- and design-based methodologies. A part of the user experience is the emotion involved in the interaction with technology [25]. Automated teller machines (ATMs) are often referenced as a user interface whereby clear mapping, ease of use, informative feedback and an aesthetic yet practical design are essential qualities. Studies illustrate that an interface embodying these characteristics ensures that users interact with more patience and tolerance and that they have a greater capacity for creative problem solving should an error occur. Overly-complicated and frustrating designs can evoke "negative affect". As a result, users tend not to engage long enough in order to find appropriate solutions thereby creating negative experiences [26]. It is important to note that negative affect can provide valuable insights regarding developing interfaces that may also need to be used in times of emergency, whereby stress, pressure, or panic can seriously hinder decision-making processes. Considering a user's emotions and experiences in the design process, provides a greater understanding of the effects of human emotional states which in turn can aid in the development of technology that behaves more intuitively.

5.5 Closing the gap between logic and creativity

Identifying the appropriate stake holders is a key element once a UCD approach is implemented. Targeting the right users will be most cost effective and provide valuable and accurate data. For example, testing a sophisticated gaming interface (e.g. joystick) on users who don't play computer games will not yield accurate useable data. Not engaging or interacting with the target group of end-users throughout the design process could be detrimental because it could result in a design or product that does not meet user requirements or is not commercially viable. A key component to UCD, aside from the involvement of the user, is the continuous evaluation process with the user, of the design itself. The repeated evaluation ensures that the primary aims, and requirements of the user are always up to date.

Design based methodologies are steps towards bridging the gap between scientific procedures and practice based artistic explorations. Merging science and creativity to support continued innovation in the IoT domain reflects the incorporation of new ideas that emerge through research into practical applications in the STEM disciplines. With respect to long term innovation, an argument can be made for a more active application of the STEAM approach as opposed to a purely STEM orientated procedure (the A in STEAM representing Arts). The generation of new ideas in research reflects in part the creative element of the human psyche [26].

The inclusion of the procedures evident in an artistic methodology has the potential to enrich those of the more logic and reason-based disciplines. It is evident in history that there was little differentiation made between the arts and sciences, however an intuitive method of discovering or validating certain truths is based primarily on deductive reasoning. This type of reasoning involves determining specific facts and narrowing the field of exploration, i.e. number of variables that affect the system, in order to ascertain specific and conclusive truths. An analogy can be drawn to top-down processing: a researcher may start with several questions but eventually ends with one singular focus. The STEM disciplines mainly embody this type of reasoning approach and has been invaluable in defining specific procedures that facilitate repeatability, consistency and transparency in good scientific research [26].

An issue with an approach that encourages an exclusive focus on singular elements of a system or product, is that the research scope can become stale, it can miss peripheral yet related subjects and can become too rigid and unforgiving. For these reasons it is important to consider the methodologies applied in the arts, which

in contrast to scientific or engineering disciplines are primarily based on inductive reasoning. Inductive reasoning embodies a bottom-up approach to research, i.e. it is the process whereby knowledge is acquired and constructed in order to generate questions that are further explored in the deductive process. An output of inductive inference are the probabilities of certain truths. For example, conclusions can be drawn logically from the knowledge acquired but it does not necessarily make accommodations if the knowledge base is inaccurate or incorrect. Therefore, the truths that emerge from inductive inference are a source of ideas and inspiration that can enrich the disciplines in which deductive reasoning is favoured [26].

The arts are primarily based on creativity - the ability to work with loose boundary conditions, the unknown or the unpredictable, to be able to see connections between diverse subject matter and to view errors as opportunities and not always as dead-ends. Similar to the sciences, a pure focus within the arts, e.g. dance, poetry, drawing, enabled discipline-specific methodologies to emerge (e.g. social art, practice-based methodologies). However, it is important to recognise that these methodologies in isolation are as self-limiting as those in the STEM disciplines.

The difference between logic and creative based disciplines is also reflected in the form of quantitative and qualitative output. Independently, each type of output has its value, but they also represent the different perspectives through which data can be interpreted. Therefore, it is not necessarily possible to declare that one form of output is better than the other, just as there is no one main research methodology that stands above others, but it is important for researchers to remain aware of these perspectives when using data and to apply a degree of objectivity.

In relation to long term technological innovation, a degree of time and energy must be invested into the early stages of technological development. In the generation of questions, it is important to draw conclusions that support a high probability of truth before the deductive approach is applied. Achieving this state indicates that the research must be thorough and accurate. The following example highlights the number of steps involved in taking an idea through the UCD process to create a high-fidelity prototype that is ready for industrial or public implementation. The even distribution of energy throughout these processes ensures that the end result is of good, reliable and repeatable quality.

- Idea generation through brainstorming with design team and ideally with relevant users
- Sorting of ideas into relevancy, possible pitfalls, new ideas, do's and don'ts
- Research one or two specific ideas and translate from idea into tangible concept
- Create a low fidelity prototype
- Run another brainstorming session with the prototype or run individual interviews
- Re-sort new contributions and research design and product elements further (e.g. material, platforms, hardware, etc.)
- Adjust and adapt the prototype to incorporate the feedback
- Run another evaluation session with the relevant users

- After the comments and feedback have been incorporated again it is possible to develop the prototype further.
- Create a high-fidelity prototype
- Implement the prototype and get real world data on interaction, glitches, successes, improvements, etc.
- Refine the prototype accordingly
- Run final tests with the prototype and evaluate the system via interviews with users
- At this point the prototype is thoroughly tested and can be moved on to the industrial phase whereby it could potentially be implemented beyond the test area.

The user centered design approach is one method by which an even distribution of energy and resources can be implemented into industry-based projects. Even though this process can initially take more time and resources, the results usually yield more accurate market research, public uptake and better quality of outputs, i.e. short-term versus long-term gain. The smart parking project described in this paper is an example of how important it is to consider the logical and creative aspects of the human psyche in the future development of technology.

The merging of logic and creativity advocates a flexible and adaptable approach to research. Whilst it is necessary to continue having singular-subject experts, to support technological innovation it is also necessary to train multidisciplinary experts. The latter type of expert and / or researcher develops the skill to communicate and translate between two or more diverse disciplines (e.g. biology, 3D printing and modelling). An advantage of this skill is the ability to alter perspective. It allows researchers to make connections between subjects or problems, etc, more quickly and they can bridge the gap between deductive and inductive inference.

6. Conclusion

The application and integration of design methodologies is a step towards integrating creative processes into scientific disciplines. Drawing out the skills in order to generate new ideas is vital in the enrichment of research scope. It is clear that the process of inductive inference (idea generation) and deductive inference (translating ideas into reality) is continuous and circular, as one phase feeds into the other. For this cyclical process to be successful it is necessary to find a balance in the application of creative and logical methodologies. In contrast focusing on either extreme (i.e. purely logic-based or creative-based methodologies) enables researchers to explore boundaries and limits of a specific discipline [26]. The peripheral topics that have been mentioned in this paper (e.g. business objectives, security, privacy, connection, etc) are not a conclusive list but demonstrate the range of interdependencies in a domain as versatile as the IoT. Even though the topics may not be directly involved, their influence can have interesting effects and outcomes on both the system in question and the related users. It is a constant interplay between variables that are known to the designer and variables that emerge as a result of implementation and real-life interaction. A chance always exists that designs may be used, not for what they were originally intended, but possibly for more creative or destructive purposes. It is another compelling reason for including users in

the product development process as early as possible because even though it may initially cost more with respect to resources and time, the long-term benefits are evident in the quality and efficiency of products produced.

The smart parking system detailed in this paper explores the relation between the involvement of the user and the development of the technology. The results from this project are congruent with research studies on system development and usability, to facilitate user acceptance and efficiency, accuracy, and satisfaction. Because the users were involved through all four phases of the design process, their needs, suggestions, and preferences were continuously incorporated into the design and evaluation of an IoT based smart parking prototype. The UCD methodology transforms users into active contributors to the design process and has provided improved UX in each iteration. It enables the final product to be tailored directly around users' needs rather than providing superfluous functions that disrupt the design.

Acknowledgments

References

[1] M. Fraifer, S. Kharel, H. Hasenfuss, A. Elmangoush, A. Ryan, W. Elgenaidi, and M. Fernström, "Look before you leap: exploring the challenges of technology and user experience in the internet of things" In Research and Technologies for Society and Industry (RTSI), 2017 IEEE 3rd International Forum on, 1-6, IEEE, Modena, Italy, 2017. https://doi.org/10.1109/RTSI.2017.8065920

[2] D. Miorandi, S. Sicari, F. De Pellegrini, and I. Chlamtac, "Internet of things: Vision, applications and research challenges" Ad hoc networks, 10(7), 1497-1516, 2012, https://doi.org/10.1016/j.adhoc.2012.02.016

[3] A. Grau, "Can you trust your fridge?" IEEE Spectrum, 52(3), 50-56, 2015. https://doi.org/10.1109/MSPEC.2015.7049440

[4] M. Tariq, Z. Zhou, J. Wu, M. Macuha, and T. Sato, "Smart grid standards for home and building automation" in 2012 IEEE International Conference on Power System Technology (POWERCON), pp. 1–6, Auckland, New Zealand, 2012. https://doi.org/10.1109/PowerCon.2012.6401448

[5] K. Naito, "A Survey on the Internet-of-Things: Standards, Challenges and Future Prospects" J. Inf. Process., vol. 25, pp. 23–31, 2017. https://doi.org/10.2197/ipsjjip.25.23

[6] D. Uckelmann, M. Harrison, F. Michahelles, "An Architectural Approach Towards the Future Internet of Things" In: Uckelmann D., Harrison M., Michahelles F. (eds) Architecting the Internet of Things. Springer, Berlin, Heidelberg, 2011. https://doi.org/10.1007/978-3-642-19157-2_1

[7] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions" Future generation computer systems, 29(7), 1645-1660, 2013 https://doi.org/10.1016/j.future.2013.01.010

[8] M. Dixit, J. Kumar, and R. Kumar, "Internet of things and its challenges" in 2015 International Conference on Green Computing and Internet of Things (ICGCIoT), 810-814, IEEE, Noida, India, 2015.https://doi.org/10.1109/ICGCIoT.2015.7380574.

[9] J. Bergman, T. Olsson, I. Johansson, and K. Rassmus-Gröhn, "An exploratory study on how Internet of Things developing companies handle User Experience Requirements" in International Working Conference on Requirements Engineering: Foundation for Software Quality, 20-36,

Springer, Cham, Utrecht, The Netherlands, 2018. https://doi.org/10.1007/978-3-319-77243-1_2

[10] M. Nati, A. Gluhak, H. Abangar, and W. Headley, "Smartcampus: A user-centric testbed for internet of things experimentation" In Wireless Personal Multimedia Communications (WPMC), 16th International Symposium on, 1-6, IEEE, Atlantic City, NJ, USA 2013. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6618632

[11] L. Cruickshank, and N. Trivedi, "Beyond Human-Centred Design: Supporting a New Materiality in the Internet of Things, or How to Design When a Toaster is One of Your Users" The Design Journal, 20(5), 561-576, 2017. https://doi.org/10.1080/14606925.2017.1349381

[12] M. Memedi, G. Tshering, M. Fogelberg, I. Jusufi, E. Kolkowska, and G. Klein, "An Interface for IoT: Feeding Back Health-Related Data to Parkinson's Disease Patients" Journal of Sensor and Actuator Networks, 7(1),14, 2018 https://doi.org/10.3390/jsan7010014

[13] J. Bergman, and I. Johansson, "The user experience perspective of Internet of Things development," Master thesis, Lund University, 2017.

[14] S. Wall, and F. Healy, Usability testing of smarter heating controls. A report to the Department for Energy and Climate Change. Amberlight. DECC, London, 2013https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/266220/usability_testing_smarter_heating_controls.pdf

[15] M. Brhel, H. Meth, A. Maedche, and K. Werder, "Exploring principles of user-centered agile software development: A literature review" Information and Software Technology, 61, 163-181, 2015. https://doi.org/10.1016/j.infsof.2015.01.004

[16] S.M. Butt, and S.M. Butt, "Usability Evaluation Method for Agile Software Development" International Journal of Software Engineering and Computer Systems, 1(1), 29-40, 2015. http://dx.doi.org/10.15282/ijsecs.1.2015.3.0003

[17] O. Sohaib, and K. Khan, "Integrating usability engineering and agile software development: A literature review" In Computer design and applications (ICCDA), 2010 international conference on (Vol. 2, pp. V2-32). IEEE, Qinhuangdao, China, 2010. https://doi.org/10.1109/ICCDA.2010.5540916

[18] X. Larrucea, A. Combelles, J. Favaro, and K. Taneja, K., "Software engineering for the internet of things" IEEE Software, (1), 24-28, 2017. https://doi.org/10.1109/MS.2017.28

[19] M.K Larusdottir, E.R. Bjarnadottir , J. Gulliksen, "The Focus on Usability in Testing Practices in Industry. In: Forbrig P., Paternó F., Mark Pejtersen A." Human-Computer Interaction, vol 332. Springer, Berlin, Heidelberg https://doi.org/10.1007/978-3-642-15231-3_11

[20] D.A.Magües, J.W. Castro, and S.T. Acuna, "HCI usability techniques in agile development" In Automatica (ICA-ACCA), IEEE International Conference,pp. 1-7, IEEE, Curico, Chile, 2016. https://doi.org/ 10.1109/ICA-ACCA.2016.7778513

[21] D. Norman, The design of everyday things: Revised and expanded edition, Constellation, 2013.

[22] M. Fraifer , H. Hasenfuss, M. Fernström, "Taking Away the Green Screen— A Brief Discussion of a Multidisciplinary Approach to IoT via, Smart Parking and Interaction Design" In Smart Trends in Systems, Security and Sustainability, 33-45, Springer, Singapore, 2018. https://doi.org/10.1007/978-981-10-6916-1_4

[23] M.Weiser, "The world is not a desktop", interactions, 1(1), 7-8, 1994. https://doi.org/ 10.1145/174800.174801

[24] J.O Kephart, and.M. Chess, "The vision of autonomic computing" Computer, (1), 41-50, 2003. https://doi.org/ 10.1109/MC.2003.1160055

[25] D. A. Norman, Emotional Design: Why We Love (or Hate) Everyday Things, Basic Civitas Books, 2004.

[26] H. Hasenfuss, "A design exploration of an agent template for multiagent systems (MAS) for shape shifting tangible user interfaces," PhD Thesis, University of Limerick, 2018.

[27] M. Fraifer, "A design exploration of an IoT based smart parking system with stakeholders using User-Centred Design" unpublished PhD Thesis, University of Limerick, 2018

The Visualization of Cattle Movement Data in The State of Pará in 2016 Through Networks of Animal Transit Graphs and Guides

Samuel Carvalho de Aragão[1,*], Agnaldo Reis Pontes[1], Luis Manuel Borges Gouveia[2], Samuel Franco Lopes[3], Pier Kenji Rauschkolb Katsuda[4], Anirene Galvão Tavares Pereira[4], Márcio Teixeira Oliveira[5], Jefferson Pinto de Oliveira[6], Rita do Socorro Brito Coroa[6], Gilson Ferreira Araújo[6], Marcio Merêncio Panza de Siqueira[6]

[1]*Federal Institute of Education, Science and Technology of Pará – IFPA, 68627-100, Brazil.*

[2]*Fernando Pessoa University – UFP, 04249-004, Portugal.*

[3]*Faculty of Amazônia – UNAMA – 68010-200, Brazil.*

[4]*São Paulo State University (UNESP)- FMVA – 16050-680. Brazil.*

[5] *Federal Institute of Education, Science and Technology of Mato Grosso do Sul – IFMS, 79641-162. Brazil.*

[6] *Pará State Agricultural Defense Agency – ADEPARÁ, 66613-115, Brazil.*

A R T I C L E   I N F O

A B S T R A C T

Animal movement is inherent in the marketing between the rural productive units, establishing space-time connections between them. The relational nature of such information is kept in the Animal Transit Guides (GTA), a mandatory issuance in Brazil. When evaluating such set of information, this work aimed at characterizing the bovine movement network in the state of Pará through the application of concepts from Social Network and Network Theory analysis. Where the nodes and edges that structure the network are respectively the rural properties and the number of cattle moved between them. The results obtained characterize the bovine movement in the state, which show distinct patterns such as the movement of a great number of animals rambling and a clear seasonality due to commercialization purposes. The evaluation of the network components' modularity also shows the subdivision of the state in four commercial clusters, where the central measures of its components define municipalities of economic and epidemiological interest. The results described herein help in the support of health surveillance and control measures, as well as the adoption of economic strategies and policies.

1. Introduction

Social Network Analysis (SNA) and the Theory of Networks are methodological concepts that describe the interactions between individuals in a group, as well as the collective behavior of a group [1]. Although widely adopted in the social sciences, psychology, and anthropology, SNA and the Theory of Networks have only recently been added to the analytical arsenal used to evaluate the nature and extent of relationships in the rural productive chain [2]. These methodologies allow researchers to evaluate the structure of the flow of animals between different productive units, and the impact that this has on the spread of a wide range of infectious agents [3–6]. SNA and the Theory of Networks also allow the extraction of characteristics of particular economic interest to the livestock industry, such as commercial patterns and commercial agglomerations [7–9].

The study of networks initially describes how the entities that teh networks are composed of are connected to each other, where the unit of interest is called the node and the connections between these are called edges [10]. Unlike in the human sciences, where the unit is the individual, in veterinary medicine the focus is on the collective unit. Municipalities and rural properties are represented by nodes and the number of animals moved between these are represented by the edges [11-12]. From an economic point of view, animal movement between productive units represents an

important aspect of livestock commercialization, as it establishes extensive spatial connections between different units [9]. This movement also favors the occurrence and spread of diseases, as movement of animals is of central importance to the dynamics of diseases and occurrence of outbreaks [13,14]. In most industrialized countries, measures are adopted to identify, record, and trace animal movements as a prerequisite for the effective adoption of disease surveillance and control measures [15].

In Brazil, the Animal Transit Guide (GTA), instituted by Ordinance 22 on January 13th, 1995 [16] and updated in 2012 by Normative Instruction No. 18, dated July 18th, 2006 [17], is the official document mandatory for the intra and interstate movement of animals, regardless of the purpose of the movement. It is an important tool for sanitary control, and provides support for the prevention of diseases that affect domestic species of production [18]. Given the relational nature of the information contained in the transit guides, these are readily converted into the attributes that make up a network, where nodes represent the properties of origin and destination, and their edges represent the animal movement that occurs between them [19].

In the state of Pará, the issuance, control, and analysis of the GTAs, as well as adoption of the necessary measures to maintain the health of the livestock are the responsibility of the Agency of Agricultural Defense of the State of Pará (ADEPARÁ). ADEPARÁ is the government body responsible for animal health in the state of Pará. The agency monitors the fifth largest national cattle herd in Brazil, which comprises more than 19 million heads [20]. This monitoring is done according to norms established by the Ministry of Agriculture, Livestock and Supply (MAPA) [21].

The State of Pará, which contains approximately 9% of cattle in Brazil, and experienced a 43% increase in the number of cattle between 2004 and 2013, performed better than the states that include the largest national herds, such as Mato Grosso, Minas Gerais, Goiás and Mato Grosso do Sul [22]. Given the importance of cattle breeding not only at the state level, but at the national level, it is of prime importance to characterize animal movements among the structural components of the production chain. Such information will enable the implementation of effective measures for sanitary, market, and managerial control [23]. Based on the information described above, this work describes the main characteristics of cattle movement in the state of Pará. We applied network analysis techniques to the relational, temporal, and spatial components of the dataset formed by the GTAs issued during the year 2016 in the state of Pará.

2. Materials and Methods

2.1 Study area

The state of Pará is located in the Northern Region of Brazil, and covers an area of more than 1,247,000 km^2. The state of Pará has estimated population of over 8 million inhabitants, and in 2016 included more than 110,000 rural properties [24].

2.2 Animal Transit Guides

The information used to compose the cattle handling network was made available by ADEPARÁ, and includes the compilation of all GTAs issued during the year 2016 in the state of Pará. Each GTA issued has a unique identifier associated with the date, party, and issuing regional unit. Information relevant to the characterization of the movement carried out includes the type of transport, the species transported, the purpose of the transport, the number of animals transported according to sex and age group (only for cattle and buffalo), municipalities and federal units of origin and destination, as well as information identifying the properties or establishments from where and to which the animals were moved. The information comprised by the CPF/CNPJ (Physical Person Registration/National Registration 0f Legal Entities), name of the owner, and name of the property or establishment of origin or destination, is assigned a unique identifier established by ADEPARÁ. Spatial information, i.e. the latitude and longitude of the municipalities of origin and destination were added to the dataset.

2.3 Cattle Handling Network

The network structure is derived from the relational information contained in each GTA issued, where the nodes represent the productive units, and the edges represent the cattle movement between them. The identification of each node corresponds to that adopted by ADEPARÁ and its attributes include the name, federative unit, and geographical position of the municipality to which it belongs. The same strategy was adopted to identify the edges, keeping the unique identifier designated by ADEPARÁ to each GTA issued. The attributes include the date of issue, the temporal component of the network, the purpose and type of transport, the number of cattle transported according to age and sex, and the identification of the origin and destination nodes. The movement described by each edge carries implicit direction information, starting at the origin and ending at the destination. Therefore, node A connected to node B, represented by A-B, is different from the connection B-A, where the distance between them is defined as the lowest number of directed edges necessary to connect them.

The network properties and characteristics were evaluated, including their diameter, modularity, heavy grades of input and output, centrality of intermediation, proximity centrality, PageRank, and clustering coefficient. The network diameter is defined as the path with the greatest number of edges between two nodes A and B of the network [25]. Modularity is a measure of the quality of the clustering of a given set of nodes [26]. The number of edges arriving at a given node and the number of edges that start from this are respectively the degrees of entry and exit [25]. The weight or multiplier of these degrees was defined as the number of animals moved between two nodes [27].

The number of shortest paths among all nodes of the network passing through node A is defined as intermediation centrality of node A [28]. The inverse of the sum of all distances from node A to all other nodes of the network defines the centrality of its proximity [29]. The PageRank of node A corresponds to the probability of arriving at this node starting from a random network node B [30]. The clustering coefficient indicates the degree to which the network nodes tend to cluster [31].

The descriptive analysis, as well as conversion of the information made available by ADEPARÁ, in the set of nodes and edges that make up the network were performed using development codes in R statitistical software [32], with RStudio 2.11.1+ [33]. The analysis and extraction of the network characteristics described above were performed using the R package 'iGraph' [34] and the open access software Gephi 0.9.1 [35].

Table 1: Number and percentage of cattle and GTAs issued, according to the purpose of cattle movement

PURPOSES	Cattle (n)	Cattle (%)	GTA (n)	GTA (%)
SLAUGHTER	3515074	31.63	210438	48.54
FATTEN	6803942	61.23	193063	44.53
SPORTS	3267	0.03	139	0.03
EXPORT	174604	1.57	7733	1.78
EXHIBITION	2103	0.02	190	0.04
AUCTION	131214	1.18	5115	1.18
RESEARCH	1	~0	1	~0
Quarantine	243677	2.19	8105	1.87
Reproduction	182671	1.64	5819	1.34
Return of Agglomeration	55234	0.5	2939	0.68

3. Results and Discussion

With 135 regional management units and 1,841 users registered in the system, the State Agency of Agricultural and Livestock Defense of the State of Pará issued 433,545 GTAs for cattle between January 1st and December 31st, 2016. The animals transported had origins or destinations in one of the 82.010 registered properties, which were distributed in 929 municipalities of 24 federative units. In Pará, there are 78.362 properties distributed in 142 municipalities, with more than 32% of the state's properties concentrated in only 10 municipalities.

Over 90% of GTAs involved road transport as a means for the movement of cattle (Table 1). This number corresponds to the distribution of the number of animals transported by GTAs, with a median of 25 cattle, which corresponds to the maximum vehicle load capacity. Road transport is an important factor in determining strategies for epidemiological control, given the possibility of rapid spread of diseases by movements over large territories [9]. In spite of the number of animals transported by road, a significant number of cattle (over 3 million) were moved on foot (Table 2), with an approximate average of 126 animals per GTA issued. In this type of movement, the animals are moved in large numbers over short distances, remaining in a region close to the origin, a factor that contributes to the local spread of diseases [36].

Table 2: Number and percentage of cattle and GTAs according to the type of transport

Type of Transportation	Cattle (n)	Cattle (%)	GTA (n)	GTA (%)
ON FOOT	3299532	29.69	26138	6.03
RAILWAY	1059	0.01	32	0.01
MARITIME / FLUVIAL	329155	2.96	13648	3.15
ROAD	7482269	67.33	393727	90.82

In Pará, beef cattle farming is very important, as shown in Table 1, and 31% of cattle moved in the state are destined for slaughter, and 61% of cattle are moved for fattening. The low index of animals moved for breeding (1.64%) is related to two marketing moments, after rearing and after fattening. The number of cattle moved for quarantine purposes (2.19%) is related to the export of live cattle, with 1.57% of the cattle moved for export purposes (Table 1). This information is also supported by the

temporal evaluation of the GTAs issued (Figure 1 and Figure 2), which indicates an increase in the number of cattle destined for quarantine during the months of May and June, which precedes the increase of GTAs issued for export during the months of June and July, which is end of the rainy season, and the period that cattle gain the most weight.

Figure 1 shows a clear seasonality of cattle breeding in Pará, with an increase in the number of cattle destined for fattening during the months of April and October, and a progressive increase in the number of cattle slaughtered between May and August. Cattle moved for fattening in October are slaughtered in early May, and those moved for fattening in April are slaughtered at the end of the rainy season in mid-August. Cattle moved for exhibitions and sports are concentrated in the months of July and August (Fig. 1 and Fig. 2), which are critical time points for the monitoring of these cattle in order to prevent the spread of diseases [2].
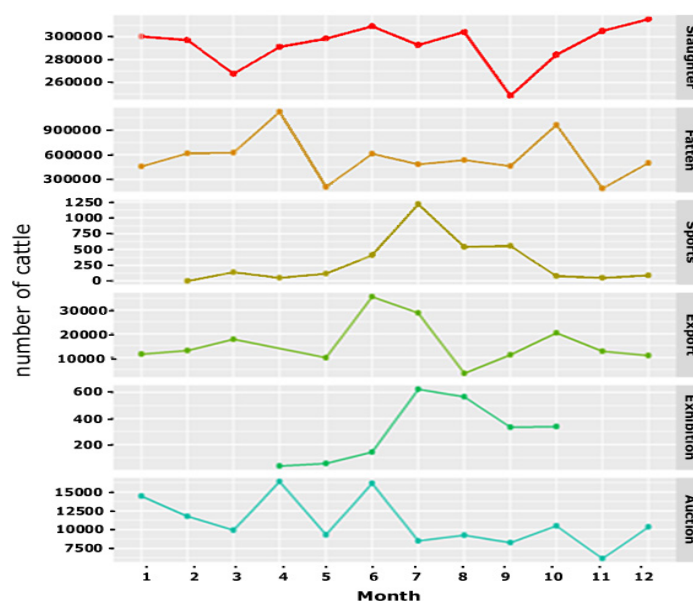
Figure 1 - Number of cattle moved during the year 2016 according to the purpose of the GTA.

Table 3: Number and percentage of cattle moved according to age and sex.

Age (months) and Sex	Bov (n)	Bov (%)
0 - 12 (males)	1496779	13.47
0 - 12 (females)	378254	3.40
13 - 24 (males)	1510723	13.60
13 - 24 (females)	705692	6.35
25 - 36 (males)	1428094	12.85
25 - 36 (females)	801877	7.22
> 36 (males)	1812238	16.31
> 36 (females)	2978358	26.80
Total (males)	6247834	56.23
Total (females)	4864181	43.77
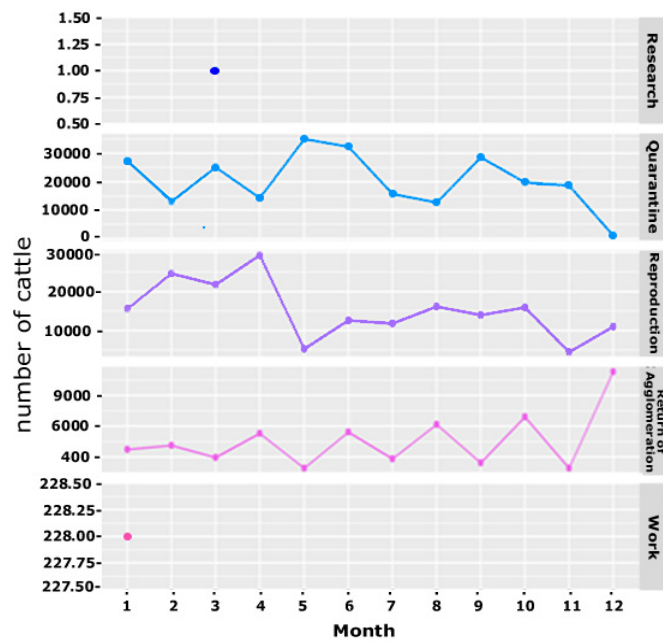
Figure 2 - Number of cattle moved during the year 2016 according to the purpose of the GTA.
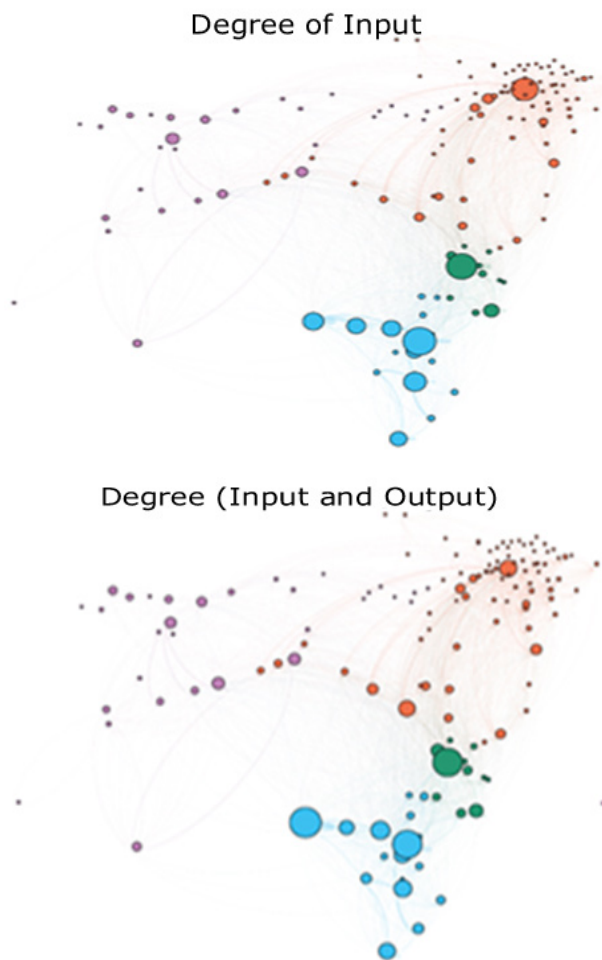
Figure 3 - Properties of the cattle movement network in the state of Pará according to the modularity class to which the municipality belongs.

Figure 4 - Properties of the cattle movement network in the state of Pará according to the modularity class to which the municipality belongs

The commercialization of male cattle in Pará is independent of age. The high number of females of reproductive age transported for slaughter (Table 3) and export purposes (Table 1), has a direct impact on the number of cattle in the state, reducing the population growth rate, and keeping the number of cattle stable. In disease surveillance programs, categories of epidemiological importance are those composed of animals susceptible to disease [37]. Therefore, determining the sample group of interest reduces the number of samples necessary for statistical confirmation of the establishment of disease in the population, as well as its spread, reducing the time required for detection and response by the responsible health agency [23].

The diameter of the net of cattle movement obtained is equal to 5, which, in other words, indirectly represents the number of municipalities through which an animal must pass to cover the largest geodesic distance of the net. Using the latitude and longitude information of each municipality, the average distance between the municipalities is approximately 220 km, with a magnitude of approximately 8.736 km, which is the sum of all distances. These results highlight the vast territory of the state.

4. Conclusion

Although the analyses and metrics evaluated were used in the state of Pará, they can be applied and expanded to other states, as well as to other production species. The quantitative description of the cattle movement network proposed in this study allowed us to reveal factors that shape the commercialization and displacement patterns of cattle in Pará. This data can contribute to the development of effective strategies to control and monitor the diseases that occur in cattle.

Conflict of Interest

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] B. Martínez-López, A. M. Perez, and J. M. Sánchez-Vizcaíno, "Social Network Analysis. Review of General Concepts and Use in Preventive Veterinary Medicine," *Transbound. Emerg. Dis.*, vol. 56, no. 4, pp. 109–120, May 2009.

[2] J. L. Hardstaff, B. Häsler, and J. R. Rushton, "Livestock trade networks for guiding animal health surveillance," *BMC Vet. Res.*, vol. 11, no. 1, p. 82, 2015.

[3] F. Natale, L. Savini, A. Giovannini, P. Calistri, L. Candeloro, and G. Fiore, "Evaluation of risk and vulnerability using a Disease Flow Centrality measure in dynamic cattle trade networks," *Prev. Vet. Med.*, vol. 98, no. 2, pp. 111–118, 2011.

[4] I. Z. Kiss, D. M. Green, and R. R. Kao, "The network of sheep movements within Great Britain: network properties and their implications for infectious disease spread," *J. R. Soc. Interface*, vol. 3, no. 10, pp. 669–677, 2006.

[5] H. H. K. Lentz *et al.*, "Disease Spread through Animal Movements: A Static and Temporal Network Analysis of Pig Trade in Germany," *PLoS One*, vol. 11, no. 5, pp. 1–32, 2016.

[6] B. Shirouchi *et al.*, "Fatty acid profiles and adipogenic gene expression of various fat depots in Japanese Black and Holstein steers," *Meat Sci.*, vol. 96, no. 1, pp. 157–164, 2014.

[7] E. E. Gorsich *et al.*, "Mapping U.S. cattle shipment networks: Spatial and temporal patterns of trade communities from 2009 to 2011," *Prev. Vet. Med.*, vol. 134, no. Supplement C, pp. 82–91, 2016.

[8] M. C. Vernon and M. J. Keeling, "Representing the UK{\textquoteright}s cattle herd as static and dynamic networks," *Proc. R. Soc. London B Biol. Sci.*, vol. 276, no. 1656, pp. 469–476, 2009.

[9] P. Bajardi, A. Barrat, F. Natale, L. Savini, and V. Colizza, "Dynamical Patterns of Cattle Trade Movements," *PLoS One*, vol. 6, no. 5, pp. 1–19, 2011.

[10] C. Dubé, C. Ribble, D. Kelton, and B. McNab, "A Review of Network Analysis Terminology and its Application to Foot-and-Mouth Disease Modelling and Policy Development," *Transbound. Emerg. Dis.*, vol. 56, no. 3, pp. 73–85, Apr. 2009.

[11] S. Schärrer *et al.*, "Evaluation of farm-level parameters derived from animal movements for use in risk-based surveillance programmes of cattle in Switzerland," *BMC Vet. Res.*, vol. 11, no. 1, p. 149, Jul. 2015.

[12] F. Kschischang, "An Introduction to Network Coding," 2012.

[13] S. E. Robinson, M. G. Everett, and R. M. Christley, "Recent network evolution increases the potential for large epidemics in the British cattle population," *Journal of the Royal Society Interface*, vol. 4, no. 15. London, pp. 669–674, Aug-2007.

[14] A. Ortiz-Pelaez, D. U. Pfeiffer, R. J. Soares-Magalhães, and F. J. Guitian, "Use of social network analysis to characterize the pattern of animal movements in the initial phases of the 2001 foot and mouth disease (FMD) epidemic in the UK," *Prev. Vet. Med.*, vol. 76, no. 1, pp. 40–55, 2006.

[15] P. Motta *et al.*, "Implications of the cattle trade network in Cameroon for regional disease prevention and control," *Sci. Rep.*, vol. 7, p. 43932, Mar. 2017.

[16] Brasil, "Portaria no 22, de 13 de janeiro de 1995," 1995.

[17] Brasil, "Decreto n° 5.741, de 03 de março de 2006," 2006.

[18] OIE OM de SA, "Código sanitário para os animais terrestres," 2015.

[19] M. Amaku *et al.*, "Infectious disease surveillance in animal movement networks: An approach based on the friendship paradox," *Prev. Vet. Med.*, vol. 121, no. 3, pp. 306–313, 2015.

[20] ADEPARA, "Agência Estadual de Defesa Agropecuária do Estado do Pará," 2015.

[21] IBGE, "Instituto Brasileiro de Geografia e Estatística."

[22] FAPESPA, "BOLETIM AGROPECUÁRIO DO ESTADO DO PARÁ," 2015.

[23] K. L. VanderWaal *et al.*, "Network analysis of cattle movements in Uruguay: Quantifying heterogeneity for risk-based disease surveillance and control," *Prev. Vet. Med.*, vol. 123, no. Supplement C, pp. 12–22, 2016.

[24] IBGE, "Sistema IBGE de Recuperacao automatica–SIDRA." 2015.

[25] T. Opsahl, F. Agneessens, and J. Skvoretz, "Node centrality in weighted networks: Generalizing degree and shortest paths," *Soc. Networks*, vol. 32, no. 3, pp. 245–251, 2010.

[26] M. E. J. Newman, "Modularity and community structure in networks," *Proc. Natl. Acad. Sci.*, vol. 103, no. 23, pp. 8577–8582, 2006.

[27] R. R. Kao, L. Danon, D. M. Green, and I. Z. Kiss, "Demographic structure and pathogen dynamics on the network of livestock movements in Great Britain," *Proc. R. Soc. London B Biol. Sci.*, vol. 273, no. 1597, pp. 1999–2007, 2006.

[28] L. C. Freeman, "A Set of Measures of Centrality Based on Betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, 1977.

[29] G. Sabidussi, "The centrality index of a graph," *Psychometrika*, vol. 31, no. 4, pp. 581–603, 1966.

[30] M. Rosvall and C. T. Bergstrom, "Maps of random walks on complex networks reveal community structure," *Proc. Natl. Acad. Sci.*, vol. 105, no. 4, pp. 1118–1123, 2008.

[31] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, p. 440, Jun. 1998.

[32] R. Ihaka and R. Gentleman, "R: A Language for Data Analysis and Graphics," *J. Comput. Graph. Stat.*, vol. 5, no. 3, pp. 299–314, 1996.

[33] J. S. Racine, "RStudio: A Platform-Independent IDE for R and Sweave," *J. Appl. Econom.*, vol. 27, no. 1, pp. 167–172, 2012.

[34] G. Csardi and T. Nepusz, "The igraph software package for complex network research," 2006.

[35] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: An Open Source Software for Exploring and Manipulating Networks Visualization and Exploration of Large Graphs," 2017.

[36] C. Ensoy, C. Faes, S. Welby, Y. Van der Stede, and M. Aerts, "Exploring cattle movements in Belgium," *Prev. Vet. Med.*, vol. 116, no. 1, pp. 89–101, 2014.

[37] L. J. Hoinville *et al.*, "Proposed terms and concepts for describing and evaluating animal-health surveillance systems," *Prev. Vet. Med.*, vol. 112, no. 1, pp. 1–12, 2013.

Virtual Output Queues Architecture for High Throughput Data Center Nodes

Angelos Kyriakos[1,2], Ioannis Patronas[1,2], Georgios Tzimas[1], Vasileios Kitsakis[1], Dionysios Reisis[*1,2]

[1]National and Kapodistrian University of Athens, Electronics Lab, Physics Dpt, GR-15784, Zografos Greece

[2]Institute for Communication and Computers (ICCS), National Technical University of Athens, Greece

A R T I C L E   I N F O

A B S T R A C T

The latest design approach for Data Centers (DCs) follows the direction of exploiting optical switching to connect Top-of-Rack (ToR) switches that serve thousands of data storing and computing devices. A ToR's usual function is the Virtual Output Queues (VOQs), which is the prevalent solution for the head-of-line blocking problem of the DC switches. An effective VOQs architecture improves the DC's performance by reducing the frames communication latency and it is efficient with respect to the implementation cost. The current paper introduces a VOQs architecture for the ToRs of DCs that function with Time Division Multiple Access (TDMA). The proposed VOQ architecture contains a bounded number of queues at each input port supporting the active destinations and forwarding the input Ethernet frames to a shared memory. An efficient mechanism of low latency grants each queue to an active destination. The VOQs constitutes a module of a ToR development, which is based on a commercially available Ethernet switch and two FPGA Xilinx boards, the Virtex VC707 and the Xilinx NetFPGA. The VOQs architecture's implementation and validation took place on the NetFPGA board.

1. Introduction

Data centers are comprised of a large number of Servers running Virtual Machines (VMs) and storage resources, which are installed in racks and communicate via the local data center network. The data centers performance depends on the available computing and data storing capacity, the architecture and the features as well as the performance of the underlying network and the Top-of-Rack (ToR) switches connecting the servers to the data center. A key factor in improving the performance of the ToR switches is the solution of the head-of-line blocking issue that is most often settled by embedding Virtual Output Queues architectures [1]. The performance of the networks depends on their interconnection scheme, which usually adhered to the multi-layer approach, and they were based on the Fat Tree or the folded Clos architectural schemes [2, 3, 4]. These approaches nevertheless, are not efficiently scalable and also, in the cases of data centers with a large number of nodes, lead to the use of a considerable number of switches, cables and transceivers, which increase power consumption.

In an effort to overcome these deficiencies researchers and engineers have introduced data center interconnections including an optical circuit switching as well as an electrical packet switching networks [5, 6, 7]. A notable design is the all optical data center proposed by the Nephele project [8]. The Nephele design adopts the Time Division Multiple Access (TDMA) mode of operation in the optical data center network. Consequently, the transmissions are completed within fixed time segments, namely the slots; each slot is assigned for sending a TDMA frame on a specific path that connects a transmitter node to a receiver node. The Nephele data center network is a Software Defined Network (SDN) and all the arrangements regarding its operation are dictated by a central data center controller. The controller is responsible for generating the TDMA Schedule, which defines which nodes communicate during each time-slot [9]. The first version of the scalable, high capacity Nephele network is able to accommodate up to 1600 Top-of-Rack (ToR) switches and each ToR uses 20 links to connect to the data center optical network.

The overall system topology of the data center network is depicted in Figure 1. The network includes I (I ≤ 20) parallel planes, each consisting of I (I ≤ 20) unidirectional optical rings.

*Dionysios Reisis, +30 210 727 6708/6720 & dreisis@phys.uoa.gr
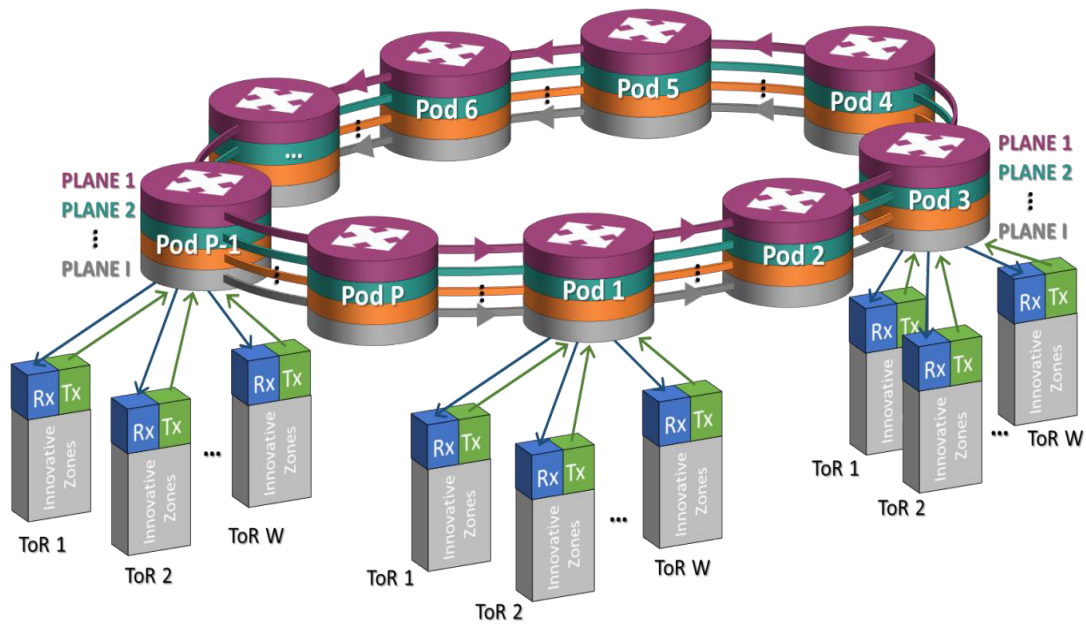
Figure 1: Nephele Data Center Network Architecture
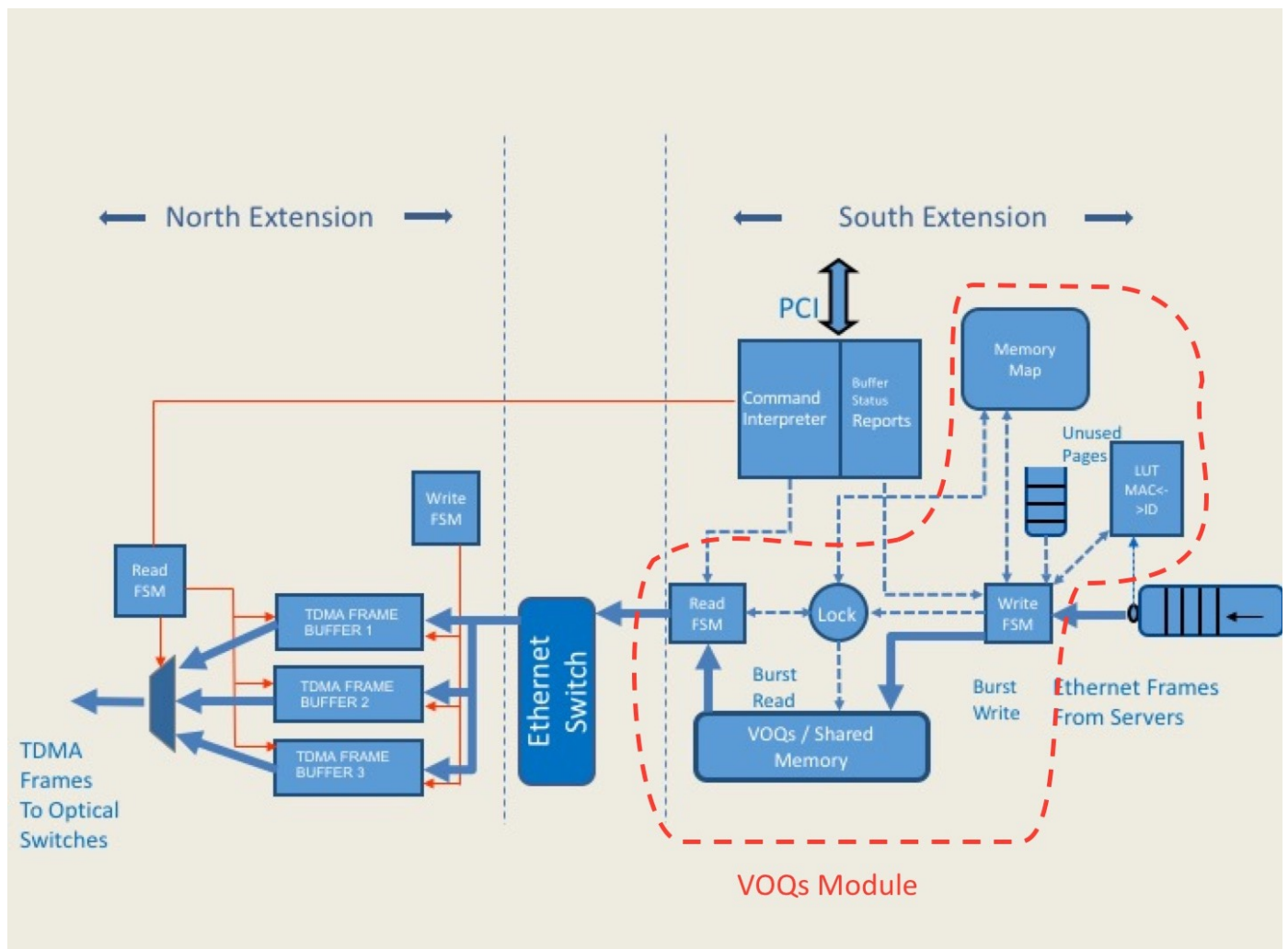
Figure 2: The Nephele Top-of-Rack (ToR) Switch Architecture Overview

The rings interconnect P (P ≤ 20) Points of Delivery (PODs). A POD comprises of I Wavelength Selective Switches (WSS) to connect the I rings, and is connected to W (W ≤ 80) ToRs, through W pod-switches, one for each ToR switch. Each ToR switch has I north ports, such that the ith north port is directed to the ith POD of each plane (each port is connected to a different POD switch). The south ports of the ToR switch connect the servers, though network interface cards (NICs), with the data center network. The performance of the ToR switch contributes significantly to the operation of the entire data center and it depends on the utilization of its resources as well as on the efficiency of the algorithms and the techniques that it employs. Among the techniques that are critical with respect to the ToR's performance is the handling of the Virtual Output Queues (VOQs). VOQs is an attractive technique for overcoming the head-of-line blocking cases [10].

This paper presents an efficient VOQ organization regarding the resource utilization and the latency needed to assign the incoming Ethernet Frames to the queues of matching destinations. The VOQ architecture introduced in this paper is advantageous due to the following: first, it is efficient with respect to the required implementation area, because it reduces the resources needed to a single shared buffer per output port. This buffer stores all the queues of data that this output port will transmit. Second, the architecture is efficient with respect to the utilization of the shared buffer's bandwidth; this is because it maximizes the throughput utilization of the buffer's interface by utilizing for storing and reading a paging organization, with each page containing a large number of Ethernet frames. Third, the proposed VOQ architecture is scalable, which is an advantage considering the scalability of the entire data center.

The proposed technique achieves the aforementioned goals based on the following ideas. The receiving ethernet Frames with the same destination are collected at the input of the switch into pages of frames. This operation is accomplished by using small sized queues positioned at each input Ethernet port. In the proposed design the number of these small sized queues at the input is bounded by the sum of the connections that are: a) serviced by each input Ethernet port and b) active during a small window of time. The latency is minimized with respect to the time required to associate each input Ethernet Frame to one of the queues. This is accomplished by employing a mechanism that maps each small size queue to one of the active destinations each time an Ethernet Frame arrives at the ToR.

The motivation for designing the proposed VOQs architecture came by the requirements of the ToR included in the Nephele project but it can serve any network, that receives an input of Ethernet frames and particularly those networks which operate under TDMA scheme, are software defined and their nodes may have to overcome the head-of-line blocking. The prototype Nephele ToR switch includes a commercially available Ethernet switch (Mellanox SX1024 [11]) and two Xilinx boards: one Virtex VC707 and one NetFPGA SUME [12]. The

implementation and the validation of the VOQs architecture took place on the NetFPGA board.

The paper is organized in five sections. Section II briefly highlights the architecture of the Nephele ToR switch. Section III introduces the architecture and the organization of the VOQs. Section IV presents the details of the FPGA implementation and finally, Section V concludes the paper.

2. The Architecture of the Top-of-Rack (ToR) Switch

The ToR design is a switch and its ports are divided in two sets: a) the south ports, which are 16 10GEthernet ports connecting the ToR with the servers b) the corresponding 16 10Gbps north ports that are connected to the optical data center network. The ToR switch consists of three fundamental blocks. The first is an Ethernet 16×16 switch having all ports as 10GEthernet [11]. The second is the North Extension. It is implemented on an FPGA and its role is: a) the formation of TDMA frames that consist of Ethernet frames and b) to implement the interface of the ToR to the network's optical (POD) switches by using its north ports. The third block is the South Extension. This FPGA based block connects the servers to the ToR. It has increased complexity and its functionality includes: a) the execution of the scheduling commands, b) to be responsible for the communication of the ToR to the data center's control plane, c) to implement the VOQs design and d) to control all the functions of the ToR.

Figure 2 presents the ToR switch's architecture as well as the functional blocks dedicated to the upstream traffic. In the part of the South Extension the figure shows the *LUT MAC-ID* that it assigns a tag to each incoming Ethernet frame. These 11 bits tags will be used within the ToR for addressing the Ethernet frames and saving on the required resources for address bits with respect to the bits required for the MAC addresses of the destinations of the incoming frames. The next action is to forward the Ethernet frames to the *VOQs/Shared Memory* block. This block stores the Ethernet frames in pages. Each page includes a large number of Ethernet frames and its length matches the length of a TDMA frame (also called Nephele frame). All the pages that belong to a destination are arranged in a linked list. The pointers required for keeping the information of each destination's linked are managed by the *Memory Map* block.

The *Command Interpreter* block (Figure 2) is responsible for the translation of the SDN controller commands: it provides to this ToR the destination ToR, which has to receive data in the upcoming TDMA slot. The ToR complies to this command and it retrieves the first page with Ethernet frames that belongs to the linked list associated to the commanded destination and sends this page to the *Ethernet switch*. There is a *Lock* mechanism (Figure 2) that grants either the storing operation of the input Ethernet frames to the shared memory or the reading operation from that memory of the TDMA frames. In more detail, the *Lock* mechanism divides the time into small time windows T_L. Each T_L is dedicated for either writing to the shared buffer or reading from

it. Hence, when the ToR reads from the shared buffer it will continue buffering in the small size queues the incoming traffic from the servers. The length of the T_L is computed at design time to balance: first, the throughput of the shared buffer, which requires long burst transactions for improved performance and second, the need of the ToR operation for writing/reading to/from the shared buffer at close time instances.

The role of the *Ethernet switch* in the upstream direction, is to forward the Ethernet frames to the North Extension and particularly to the buffer of the corresponding destination's north port. In that buffer the Ethernet frames formulate the final TDMA/Nephele frame, to which are also added first, the preamble and second, a word required for each device synchronization. The *Command Interpreter* follows the schedule received from the control plane servers and it specifies (the red control signal of Figure 2) the slot that the ToR will transmit that TDMA frame. For the downstream direction, the Nephele design mandates the Ethernet switch to just forward the frame from the north input port to the corresponding south port. That is, the design complexity of the ToR is mostly related to the upstream path.

The communication of the ToR switch with the control plane is accomplished through the PCI Express interconnection. The PCI Express interface in the proposed architecture is implemented by the use of the Xilinx IP Core for PCIe and RIFFA (Reusable Integration Framework for FPGA Accelerators) [13]. The RIFFA framework consists of an API (Application Programming Interface) and a driver/kernel module for the host PC and IP core for the FPGA, all of which are open-source. The module provided

by RIFFA for the FPGA is designed as an extension to the Xilinx core, which handles the physical layer of the PCIe interface.

3. Virtual Output Queues

The proposed VOQ design improves the required hardware resources based on the following concept. During a narrow time window T_B, the ToR switch receives Ethernet frames at its south ports for various destinations in the data center network, which we define as *active destinations*. We consider that for all practical purposes, the number of the active destinations, during T_B, has an upper limit, which can be an outcome of statistical measurements of the network traffic patterns. The active destinations' upper limit is significantly smaller compared to the number of all the possible destinations in the data center. Hence, letting a queue to keep all the incoming Ethernet frames during T_B that have the same active destination and prepare in this queue a burst to be written to the shared buffer, leads to an architecture that includes a set of queues with cardinal number equal to that of the active destinations, while it still keeps the high throughput at the shared buffer.

Considering the above, the VOQs architecture is comprised of: first, the Shared Memory (buffer), second the Memory Map depicted in Figure 2 and third, the VOQs controller. The detailed architecture of the VOQs controller is shown in Figure 3: it is a design of the VOQs controller that includes four (4) active destinations and the corresponding queues, based on a hypothesis that the application asks for four active destination and as shown in Figure 3 there is one queue to support each active destination. In order to define the length of the time window T_B we consider the following facts..
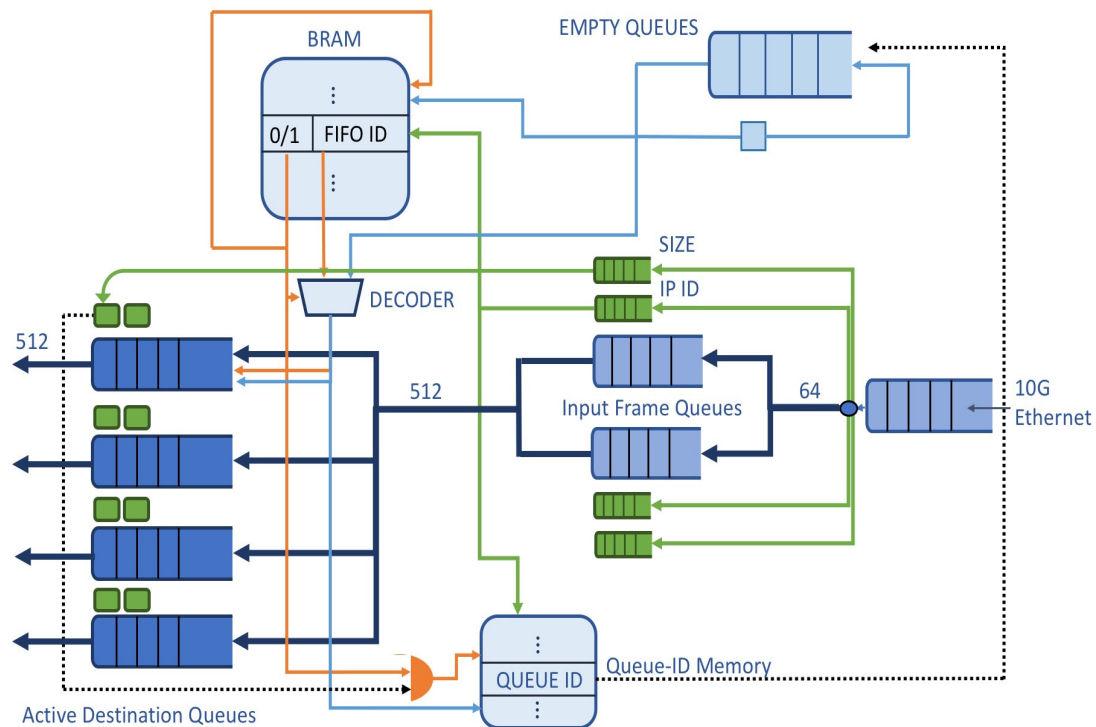
Figure 3: VOQs Controller Architecture Overview

The design of the shared buffer employs a Dynamic Random Access Memory (DRAM) that can reach the considerable throughput of 80 Gbps at its interface; this performance will be feasible if the entire VOQs architecture can operate with burst transactions for reading and writing from/to the shared buffer, thus exploiting the DRAM interface, which requires a minimum burst time t_{mb} depending on the DRAM specifications. The performance of the DRAM organization degrades significantly when the size of the burst size decreases. We note here that, this performance degradation cannot be expressed (defined) as a function of the burst size, e.g. proportional. Therefore, reading and writing from/to a page in the shared buffer (in the linked list assigned to a destination) must be performed in bursts and each burst has to consist of multiple Ethernet frames, in the order of Kbytes. Therefore, we need an architecture of queues able to gather into a single queue all the incoming Ethernet frames that have the same destination; in that queue, the controller will formulate a burst of these Ethernet frames. Finally, it will operate in burst mode to store these frames into the page of the linked list of that destination, which is kept in

We consider the time window T_B and the number of queues for active destinations k to be calculated by the following reasoning. At a clock cycle T_0, given that are available k queues storing Ethernet frames of k different IP, there will be Ethernet frames arriving to at most all of these queues and at the clock cycle T_b that at least one of these queues has completed a burst, and this queue can write the burst to the buffer. Therefore, this queue can formulate another burst either for the IP that it was supporting up to T_b or the queue can be reassigned by the controller to serve another IP. Thus, in this scenario, the worst case is that we have to keep the k queues serving their IPs for as long as no queue has completed a burst: assuming that each queue receives an Ethernet frame in a round robin fashion T_B is at most equal to k × t_{mb}.

According to the above, the efficiency of the VOQs architecture is defined as the maximization of the utilization of the available resources and the DRAM buffer throughput. For this purpose, the design has to: a) include k queues for preparing the bursts, so that each queue prepares a burst that will be stored in an active destination linked list of pages; b) minimize latency and c) minimize the number of the k queues along with their size. The su

cceeding paragraphs describe how we achieve the above goals and they describe in detail the operations of the VOQs Controller as well as its functional blocks and components.

The ToR switch is connected with 10G Ethernet to the servers through its south ports. First, the Ethernet Frames that arrive from the servers at the rate of 10G are buffered in the port queue of the 10G Ethernet module and then are forwarded and buffered to the two *Input Frame Queues* (Figure 3) in the following way: we start counting the incoming frames and depending on the arrival sequence the odd numbered incoming Ethernet Frames are stored in the first *Input Frame Queue* (the upper queue on Figure 3) and the even Ethernet Frames to the second queue. This dual queue architecture gives us the necessary time in order to perform in

real-time the two following operations on the Ethernet Frames: while we store a frame in one of the *Input Frame Queues*, we calculate its size and extract its destination's IP, which then are stored to two queues of significantly lesser size, the *IP ID* queue and the *SIZE* queue, which are positioned close to each *Input Frame Queue* in the design of Figure 3.

LUT Memory Map

**Shared Memory Level
Pages of Ethernet Frames**
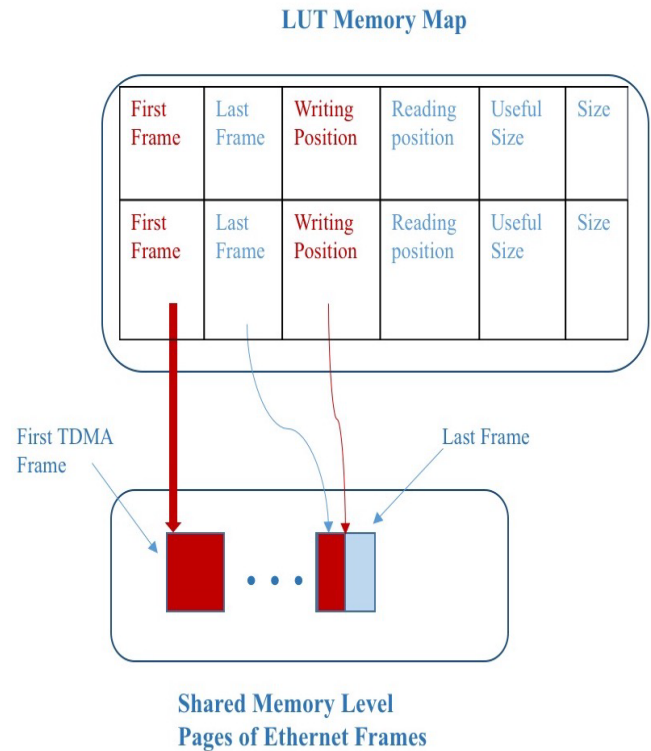
Each frame's IP stored in the *Input Frame Queues* is passed as input (address) to a LUT, named *BRAM* in Figure 3. The LUT will specify (will give as output) the id of an *Active Destination Queue* (on Figure 3 we shown an example design with four queues): in the specified *Active Destination Queue* we will buffer all the Ethernet Frames with the current active destination IP, in order to form a burst that it will be stored into the linked list of pages of that destination in the DRAM buffer. Apart the id of the *Active Destination Queue* in that *BRAM* location is also stored a flag (0/1). When the flag is equal to "1", it specifies the case in which the *Active Destination Queue* id (stored in the LUT) is granted to the active destination IP. Alternatively, the case when the flag equals to "0" indicates that the frame's destination IP is not yet served by any of the *Active Destination Queues* and hence, the controller has to assign an *Active Destination Queue* to this IP. Now, we consider the case of an Ethernet frame arriving at the ToR and its IP address does not correspond to any of the *Active Destination Queues*. If we have correctly calculated (during the design of the ToR) the minimum required number of the *Active Destination Queues* that it is sufficient to serve the application demands, the VOQs controller will have an empty *Active Destination Queue* available for assignment to a newly arrived

Ethernet frame that requests an *Active Destination Queue* to buffer the following frames with the same IP destination. All the id (numbers) of the unused *Active Destination Queues* are buffered in the queue named *Empty Queues* in Figure 3. At the same clock cycle that we read from the *BRAM* the id of the *Active Destination Queue* that serves the frame's IP along with the "1/0" (assigned to a queue or not) flag, we also read the first empty queue id from the *Empty Queues*. The multiplexer shown at Figure 3 bellow the BRAM is controlled by the flag in order to select: a) the *BRAM* output when the flag equals "1" and b) the *Empty Queues* output if the flag is "0". In the first case where we will use the *BRAM* output, the empty queue id that was just extracted from *Empty Queues* will be returned back in the *Empty Queues,* since it was not used. The above design minimizes the latency for the assignment of an active queue to the new destination.

We have to mention that in order to exploit the high throughput of the DRAM interface, we have to write the Ethernet Frames in the shared buffer as a burst of contiguous words of a significant length (512 bits in the example implementation of the proposed architecture). We note here that, in a writing burst of Ethernet frames the last 512-bit word might not be completely filled with Ethernet frames payload and for completing the burst we add 0xFF as padding. The simple padding provides the advantage of simplifying the control and it reduces the latency at the cost of the dummy data overhead in many pages in the shared buffer. This padding overhead becomes larger for small Ethernet frames and it is reduced significantly in the case of full Ethernet frames. Note here that, when it's time to transmit a TDMA frame the shared memory will provide us with a page: we must be informed regarding the exact number of the useful data in this page in order to remove the padding. For this purpose, we store in the header of each page the useful size along with the actual page size, which is the overall sum of the useful size and the size of the padding stored in the shared buffer.

A small size dual port memory shown in Figure 3. as *Queue-ID Memory*, stores the IP that it is currently served by each *Active Destination Queue*. Each address *X* of the *Queue-ID Memory* corresponds to the *Active Destination Queue* with id *X*. The data at that address *X* of the *Queue-ID Memory* is the destination's IP that is accommodated by this *Active Destination Queue*. When it is the first time that an Ethernet Frame is stored in an empty *Active Destination Queue* the id of this queue is used as the address to the *Queue-ID Memory*, and in that address, we store the frame's IP. During the whole time that this *Active Destination Queue* serves the IP, the *Queue-ID Memory* keeps the IP in that address. Only when an *Active Destination Queue* is left with all its data forwarded to the shared buffer, we will: first, erase the contents of the served destination in the *BRAM* by acquiring the address (IP) from the *Queue-ID Memory* and second, write the queue id to the *Empty Queues* to refresh the *Active Destination Queues* that are vacant and they can be granted to another destination IP. Consequently, the location in the *Queue-ID Memory* will be overwritten by the new IP, which will be served by the corresponding *Active Destination Queue*.

The proposed design minimizes the time required to perform all the previously mentioned operations with respect to clock cycles. The architecture can achieve the time minimization due to the parallelization of the operations and as a result, the VOQ architecture diminishes the latency of each stage. Consequently, the *Active Destination Queues* can be as many as the application dictates as upper bound. Moreover, the length of each queue doesn't need to grow beyond the size of the burst that it is specified by the DRAM controller for reaching its maximum throughput.

The block called *Memory Map* stores all the information related to each linked list in the shared buffer associated to each destination IP. The memory map entries are shown in Figure 4, 5 in two working examples. Each entry of the *Memory Map* block has the following pointers: one at the address of the first page of the list noting from what page we are currently reading data to transmit; one to the last page, required to inform the VOQs that this is the page, which currently stores all the Ethernet frames for the associated destination; one for the "next to write" address of the last page (writing position in Figure 4), one for the "next to read" address of the first page (reading position in Figure 5). Moreover, the *Memory Map* entry provides the exact number of useful data in the page: this information is used to compute the total volume of data of the Ethernet frames with or without out the padding.
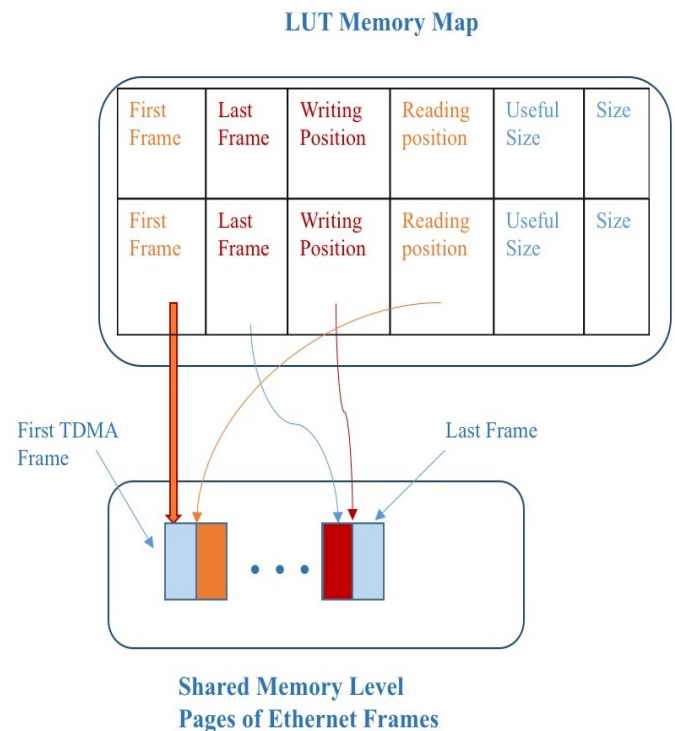
Figure 5: Memory Map Organization Concurrent Write & Read Operations

The pointers at each *Memory Map* block location are refreshed during each burst write/read transaction. Thus, at the beginning of a write/read operation to/from the DRAM buffer we know the exact number of the data (bytes) that will be written/read. We operate the linked list of pages as a queue since we always transmit the head page of the list. The memory map architecture is able to concurrently write and read from the same linked list of pages as shown in Figure 5.

A noteworthy advantage of the novel VOQ technique is the scalability of the architecture, which can be easily configured to accommodate various numbers of Active Destination Queues, the size of the DRAM shared buffer and the size of the *Memory Map* block. The pointers and the size of the linked list of pages for each destination are stored in block rams (BRAM) in the FPGA. The required size of the BRAMs is proportional to: first, the DRAM memory size, and second the number of destinations in the data center network. In the case that the size of the mapping information is relatively large and it constrains the designer of implementing the Memory Map on the FPGA internal BRAM memory, the proposed architecture of the Memory Map block can be implemented on an external Static Random Access Memory (SRAM).

4. FPGA Implementation Details

We have realized an example VOQs design with four (4) active destinations (k = 4 is adequate for most applications in accordance with our T_B and k calculations). The development of the example implementation was made on the NetFPGA SUME board using the Xilinx Vivado development tool. The design includes 3 Intellectual Property (IP) hardware Cores from Xilinx: a) 10GbE Subsystem, which includes the MAC and the 10GbE PCS/PMA b) Integrated Block for PCI Express c) Memory Interface Generator (MIG) for the shared DRAM buffer. The NetFPGA board receives the scheduling commands from the host desktop PC, which is running Linux and communicates with the data center's controller, which runs on a different PC in the same local network.

Table 1: FPGA Resources

Slice LUTs	4194
Slice Registers	2415
Slices	1888
Logic LUTs	2639
Memory LUTs	1285
Flip flop pairs LUTs	4848
Block RAMs	62

The resources occupied in the NetFPGA SUME for the VOQs Controller are presented in the Table 1, reported by the Vivado tool. The input small sized queues are all performing at 156 Mhz clock and use 64 bits word length, in order to comply with the 10G Ethernet physical layer standard. The Active

Destination Queues and memories alongside of them in our implementation are performing at 200 MHz with 512-word length.

5. Conclusion

The current paper presented a VOQs architecture, which is efficient with respect to latency and the hardware resources and it supports a ToR switch that is adaptable to any data center network operating under the TDMA scheme. The most noteworthy novelty of the proposed VOQs architecture is the efficient use of a single large shared buffer, the performance of which is fully exploited. The VOQ organization is based on the notion of *Active Destination Queues* that leads to maximize the utilization of the shared buffer and reduces significantly the required number of the *Active Destination Queues* to the number of the connections that are active during a narrow time window. Moreover, the management/control of the *Active Destination Queues* is efficient due to the minimum latency that induces to the operation of the ToR switch. Furthermore, the proposed architecture is scalable with respect to the number (k) of the *Active Destination Queues,* the scale of the data center network (number of destinations), the shared buffer size and the Ethernet protocol (Ethernet type/Frame size).

Conflict of Interest

The authors declare no conflict of interest.

References

[1] A. Kyriakos, I. Patronas, G. Tzimas, V. Kitsakis and D. Reisis, "Realizing virtual output queues in high throughput data center nodes," 2017 Panhellenic Conference on Electronics and Telecommunications (PACET), Xanthi, 2017, pp. 1-4. doi: 10.1109/PACET.2017.8259971

[2] M. Al-Fares, A. Loukissas, and A. Vahdat, "A scalable, commodity data center network architecture," SIGCOMM Comput. Commun. Rev., vol. 38, no. 4, pp. 63–74, Aug. 2008. [Online]. Available: http://doi.acm.org/10.1145/1402946.1402967

[3] A. Greenberg, J. R. Hamilton, N. Jain, S. Kandula, C. Kim, P. Lahiri, D. A. Maltz, P. Patel, and S. Sengupta, "Vl2: A scalable and flexible data center network," in Proceedings of the ACM SIGCOMM 2009 Conference on Data Communication, ser. SIGCOMM '09. New York, NY, USA: ACM, 2009, pp. 51–62. [Online]. Available: http://doi.acm.org/10.1145/1592568.1592576

[4] N. Farrington, E. Rubow, and A. Vahdat, "Data center switch architecture in the age of merchant silicon", in Proceedings of the 2009 17th IEEE Symposium on High Performance Interconnects, ser. HOTI '09, 2009, pp. 93–102.

[5] H. H. Bazzaz, M. Tewari, G. Wang, G. Porter, T. S. E. Ng, D. G. Andersen, M. Kaminsky, M. A. Kozuch, and A. Vahdat, "Switching the optical divide: Fundamental challenges for hybrid electrical/optical datacenter networks," in Proceedings of the 2Nd ACM Symposium on Cloud Computing, ser. SOCC '11. New York, NY, USA: ACM, 2011, pp. 30:1–30:8. [Online]. Available: http://doi.acm.org/10.1145/2038916.2038946

[6] N. Farrington, G. Porter, S. Radhakrishnan, H. H. Bazzaz, V. Subramanya, Y. Fainman, G. Papen, and A. Vahdat, "Helios: A hybrid electrical/optical switch architecture for modular data centers," in Proceedings of the ACM SIGCOMM 2010 Conference, ser. SIGCOMM '10. New York, NY, USA: ACM, 2010, pp. 339–350. [Online]. Available: http://doi.acm.org/10.1145/1851182.1851223

[7] K. Tokas, C. Spatharakis, I. Kanakis, N. Iliadis, P. Bakopoulos, H. Avramopoulos, I. Patronas, N. Liakopoulos, and D. Reisis, "A scalable optically-switched datacenter network with multicasting," in 2016 European Conference on Networks and Communications (EuCNC), June 2016, pp. 265–270.

[8] P. Bakopoulos, K. Christodoulopoulos, G. Landi, M. Aziz, E. Zahavi, D. Gallico, R. Pitwon, K. Tokas, I. Patronas, M. Capitani, C. Spatharakis, K. Yiannopoulos, K. Wang, K. Kontodimas, I. Lazarou, P. Wieder, D. Reisis, E.

Varvarigos, M. Biancani, H. Avramopoulos, "NEPHELE: an end-to-end scalable and dynamically reconfigurable optical architecture for application-aware SDN cloud datacenters", IEEE Communications Magazine, 2018

[9] K. Christodoulopoulos, K. Kontodimas, K. Yiannopoulos, E. Varvarigos, "Bandwidth Allocation in the NEPHELE Hybrid Optical Interconnect", 2016 18th International Conference on Transparent Optical Networks (ICTON), July 2016.

[10] Pedro Yébenes, German Maglione-Mathey, Jesus Escudero Sahuquillo, Pedro J. García, Francisco J. Quiles, "Modeling a switch architecture with virtual output queues and virtual channels in HPC-systems simulators", 2016 International Conference on High Performance Computing & Simulation (HPCS), Innsbruck, Austria, July 2016.

[11] Mellanox Technologies, "SX1024: The Ideal Multi-Purpose Top-of-Rack Switch", White Paper, May 2013. Available: https://www.mellanox.com/pdf/whitepapers/SX1024-The-Ideal-Multipurpose-TOR-Switch.pdf

[12] N. Zilberman, Y. Audzevich, G. A. Covington, and A. W. Moore, "Netfpga sume: Toward 100 gbps as research commodity," IEEE Micro, vol. 34, no. 5, pp. 32–41, Sept 2014.

[13] Matthew Jacobsen, Dustin Richmond, Matthew Hogains, and Ryan Kastner. 2015 RIFFA 2.1: A Reusable Integration Framework for FPGA Accelerators. ACM Trans. Reconfigurable Technol. Syst. 8, 4, Article 22 (September 2015), 23 pages. Available: http://dx.doi.org/10.1145/2815631

An Electroencephalogram Analysis Method to Detect Preference Patterns Using Gray Association Degrees and Support Vector Machines

Shin-ichi Ito[*], Momoyo Ito, Minoru Fukumi

Department of Science and Technology, Faculty of Science and Technology, Tokushima University, 770-8506, Japan

A B S T R A C T

This paper introduces an electroencephalogram (EEG) analysis method to detect preferences for particular sounds. Our study aims to create novel brain–computer interfaces (BMIs) to control human mental (NBMICM), which are used to detect human mental conditions i.e., preferences, thinking, and consciousness, choose stimuli to control these mental conditions, and evaluate these choices. It is important to detect the preferences on stimuli. If the stimuli related to the preference can be detected, the NBCIMC can provide stimuli to the user based on their emotions by detecting their favorite stimuli. The proposed method adopted EEG recording technique, extraction techniques of EEG features and detection methods of preferences. EEG recording employs a simple electroencephalograph, for which the measurement position is the left frontal lobe (Fp1) of the brain. We assume that the differences of the EEG activities on the patterns of preference are expressed in the association between the changes of the power spectra on each frequency band of the EEG. To calculate the association, we employ the gray theory model. The EEG feature is extracted by calculating the gray association degree, then, the preferences are detect using a support vector machine (SVM). Experiments are conducted to test the effectiveness of this method, which is validated by a mean accuracy rate >88% on the favorite sound detection. These results suggest that the detection of subject's favorite sounds becomes easy when the EEG signals are analyzed while the gray associate degrees are used as the EEG feature and the SVM is used as the classifier.

1. Introduction

This paper is an extension of a paper presented at the 2018 International Conference on Electronic, Information, and Communication (ICEIC), 2018 [1]. This paper discusses the detection of human mental state conditions, proposes a method for detecting patterns of human preferences, and contains additional discussions on comparative methods.

Recently, electroencephalogram (EEG)-based interfaces (brain–computer interfaces (BCI) and/or BMI) have gained considerable research interest in many fields. This study aims to create new BMIs to control human mental (NBMICM), which are used to detect human mental conditions e.g., preferences, thinking, consciousness, choose stimuli for controlling these mental conditions, and evaluate these choices. It is important to detect the preferences on stimuli. If the stimuli related to the preference can

be detected, the NBCIMC can provide stimuli to the user based on their emotions by detecting their favorite stimuli.

There are any approaches to analyze and detect the human preferences in previous studies. Sawata et. al. proposed the method to detect the individual favorite music by calculating the audio features of EEG signals during listening to the music [2]. In [3], the author analyzed the differences of the relationship between the ethnic groups and music preferences by detecting the preference on music using EEG signals and their analyses. In [4], the authors analyzed the relationship between the haptic preference and the Gamma EEG to detect the features of the EEG signals. However, these previous studies employed the electroencephalograph that had many electrodes to record the EEG and it may be difficult to use the interface using the EEG on daily basis [5]. This paper used a simple device that has a single dry-type electrode. The sensing position is on the left lobe. The prefrontal cortex activities are changes when a human mental state is changed [6,7], therefore, EEG activities in the prefrontal pole are variable. Furthermore, the

EEGs of the frontal cortex activities vary from person to person [8,9]. We propose a method using single-point sensing to analyze EEGs and mitigate the adverse effects of such individual differences.

To analyze EEG activities, there are numerous approaches [10]. In [3], [4], the authors calculated the power spectrum of EEG signals to extract the features. In [11], [12], the authors employed principal component analysis for extracting EEG features [11,12]. In [13], the authors used independent component analysis to extract the features. Then, In [13,14], the researchers employed *k*-nearest neighbor (*k*NN) to detect and/or classify the EEG features. Blankertz et. al. used linear discriminant analysis (LDA) as classifier [15]. In [12], the researcher employ artificial neural networks (ANN) to classify the EEG signals. In [16], the author used self-organizing map (SOM) to recognize the human emotions. In [2], [11], [17] the authors employed support vector machine (SVM) to analyze human preference and EEG signals, respectively.

Here, we assumed that relationships among the power spectra of the frequency bands become unique when creating the preferences on stimuli and different on each preference. This paper employed the gray associate degree calculation technique [14,18] to compute the relationships. Also, the gray associate degree calculation technique can reduce the noise signals in the EEG signals because of one of statistical processing models. The SVM classifier was employed to detect the preferences because the previous studies had good results to detect the preference using the SVM. We were able to detect the subject's preference patterns by analyzing the EEG activities during listening to certain sounds. If the stimuli are sounds and the favorite sounds are detected by analyzing the EEG signals, the NBCIMC can give user healing sounds to be comfortable blanket by detecting the favorite sounds. Finally, we conducted experiments using real EEG data for testing the effectiveness of the proposed method.

2. Proposed method

The proposed method adopted EEG recording technique, extraction techniques of EEG features and detection methods of preferences, as illustrated in Figure 1.

MindTune (MT) was used to record the EEG signals. This device has a dry-type sensor and a few electrodes installed in a pair of headphones. For the EEG recordings, the left ear and Fp1 in the international 10-20 system are reference and exploring electrodes, respectively (Figure 2). The timetable of each EEG recording was 15 seconds (no listening) and 15 seconds (listening to sound) as a set, as shown in Figure 3. The recorded EEG signals were transformed into the power spectra using a fast Fourier transform. Then, the power spectra on each frequency band were based on rhythm of brain activities. This paper divided into four frequency bands; theta, low-alpha, high-alpha and beta. The range of frequency bands are 4–7Hz, 8–9Hz, 10–12Hz and 13–24Hz, respectively. After the EEG recording, the subject completed an easy questionnaire on the preference evaluation of the sounds. The preference patterns were based on the responses to the questionnaire, of which the criterion and the preference patterns (indicated in parentheses) were based on whether the subject liked
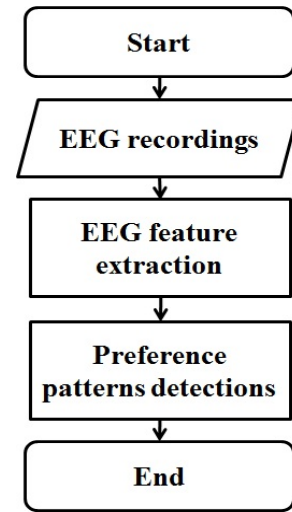
Figure 1: Procedure of the proposed method.
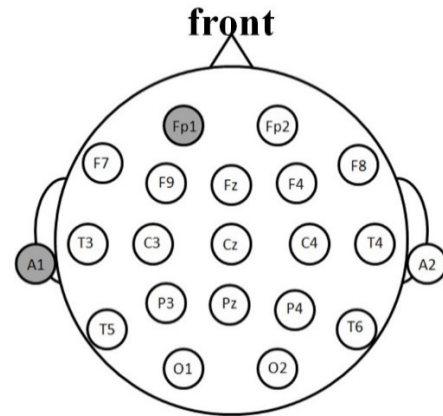
Figure 2: International 10-20 system. Reference electrode is at left earlobe (A1) and exploring electrode is at left prefrontal pole (Fp1).

Figure 3: Timetable of each EEG recording.

(FavoriteSound), disliked (DislikeSound), or felt neutral toward, i.e., neither liked nor disliked (Other), a particular sound.

The gray associate degree was calculated to extract the EEG features based on the time variations of the power spectra on each frequency band of the EEG because we assumed that the differences of the EEG activities on the preference patterns are expressed in the association between the changes of the power spectra. The eigenvectors were calculated to extract the gray association degree from the gray relationship coefficients. We regarded the eigenvector of the greatest eigenvalue as the EEG feature. The matrices of the discrete time series data of the power spectra of the EEG frequency bands (EEG) and the gray relational coefficient (GRC) are defined as follows:
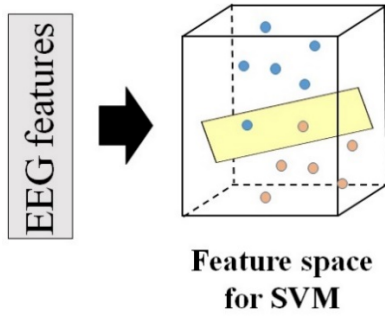
Figure 4: Classifier for detecting preference patterns using the SVM.

$$EEG = \begin{bmatrix} EEG_\theta(1) & EEG_\theta(2) & \cdots & EEG_\theta(T) \\ EEG_{low\alpha}(1) & EEG_{low\alpha}(2) & \cdots & EEG_{low\alpha}(T) \\ EEG_{high\alpha}(1) & EEG_{high\alpha}(2) & \cdots & EEG_{high\alpha}(T) \\ EEG_\beta(1) & EEG_\beta(2) & \cdots & EEG_\beta(T) \end{bmatrix} \tag{1}$$

$$GRC = \begin{bmatrix} \Gamma_{11} & \cdots & \Gamma_{14} \\ \vdots & \ddots & \vdots \\ \Gamma_{41} & \cdots & \Gamma_{44} \end{bmatrix} \tag{2}$$

subject to

$$\Gamma_{ij} = \frac{1}{T}\sum_{k=1}^{T} R_{ij}(k).$$

where EEG and T denote the power spectra and the maximum time, respectively. R is the gray relational coefficient. i and j are the frequency bands. R is defined as

$$R_{ij} = \frac{\Delta_{min.} + \upsilon\Delta_{max.}}{\Delta_{ij}(k) + \upsilon\Delta_{max.}} \tag{3}$$

subject to

$$\begin{cases} \Delta_{min.} = \min_{\forall j \in i} . \min_{\forall k} . = \left\| EEG_i(k) - EEG_j(k) \right\|, \\ \Delta_{max.} = \max_{\forall j \in i} . \max_{\forall k} . = \left\| EEG_i(k) - EEG_j(k) \right\| \end{cases}$$

where υ and Δ_{ij} indicate the distinguishing coefficient ($\upsilon \in [0,1]$) and $\left\| EEG_i(k) - EEG_j(k) \right\|$, respectively. $EEG_i(k)$ and $EEG_j(k)$ denote the reference and comparison matrices, respectively.

In preference detection, the kernel SVM classifier (shown in Figure 4) using the RBF kernel function was employed to detect the preference patterns because the SVM classifier was employed to detect and/or classify the EEG features in previous studies [2,11,17]. The previous studies could obtain good results using the SVM classifier. Then, this paper employed a one-versus-rest strategy for detecting the preferences. An N-fold cross-validation method was used for testing. Finally, this paper calculated the mean accuracy rate for the preference detection ($MeanAc$) to confirm the effectiveness of the proposed method:

$$MeanAc = \frac{1}{N}\sum CrrectNum / SampleNum \tag{4}$$

where $CorrectNum$ is the total number of correct answers $SampleNum$ means the total sample number.

3. Experiments

The subjects were four males (average age: 22.5 years) and one female (age: 22 years). The EEG device was positioned on the forehead of each subject. The EEG was recorded more than once in the laboratory with ongoing background noise. The subjects listened to 15 kinds of the sounds in an experiment. The sounds comprised window bells sound, helicopter noise, fire engine , grade crossing, scotch tape, cicada buzz, bush warbler buzz, mosquito, roar of waves, soda water, unwrapping the paper, fireworks, train noise, frictional noise of styrene foams and drill noise, respectively. The experiments were repeated five times for each subject, producing a total of 375 sounds that were played. Thus, T and υ for the EEG feature extraction were 15 and 0.5, respectively, while C and γ for the kernel SVM were 2^{-5} and 2^{-12}, respectively. Then, the number of N for cross validation was 5.

Table 1 shows the experimental results for the means and standard deviations of the accuracy rates of the detections of FavoriteSound (Favorite), DislikeSound (Dislike), and Other. The comparative methods, PCA, and the power spectra were used to extract the feature. The nearest neighbor (1NN) method was used as classifier to detect the preferences. The EEG features of the comparative methods were shared by the proposed method. Gray, PCA, and Freq were the gray association degrees, which were the results of the PCA and the power spectra of the EEG (Equation 1), respectively, as the EEG features. SVM and kNN were the classifiers used by SVM and 1NN, respectively, for detecting the preference patterns.

For all methods (Gray + kNN, PCA + SVM, PCA + kNN, Freq + SVM, and Freq + kNN) other than our proposed method, the mean detection accuracies of FavoriteSound were the highest with accuracies of 88.27, 76.53, 88, 77.34, 87.47 and 82.57, respectively. Using our proposed method (Gray + SVM), the mean detection accuracies of FavoriteSound, DislikeSound, and Other were the highest with accuracies of 88.27, 52.27, and 59.6, respectively, as compared with the comparative methods (Gray + kNN, PCA + SVM, PCA + kNN, Freq + SVM, and Freq + kNN). Moreover, the standard deviations of the detection accuracies of

Table 1: Means and standard deviations of the accuracy rates of detections of FavoriteSound, DislikeSound, and Other.

(%)	Favorite	Dislike	Other
Gray + SVM (Proposed method)	**88.27 ± 0.01**	**52.27 ± 0.01**	**59.6 ± 0.01**
Gray + kNN	**76.53 ± 0.01**	50.67 ± 0.02	49.64 ± 0.01
PCA + SVM	**88.0 ± 0.01**	50.67 ± 0.01	59.47 ± 0.01
PCA + kNN	**77.34 ± 0.01**	47.46 ± 0.02	50.93 ± 0.01
Freq + SVM	**87.47 ± 0.02**	48.8 ± 0.01	52.27 ± 0.01
Freq + kNN	**82.57 ± 0.03**	47.74 ± 0.02	51.2 ± 0.01

Favorite, Dislike, and Other were 0.03, 0.02, and 0.01, respectively. The mean detection accuracies of Dislike and Other were 60% or less.

4. Discussions

Using all methods with accuracies of 76.53% or more, we confirmed that the mean detection accuracies of FavoriteSound were the highest. These results suggest that the favorite stimuli are detected by analyzing the left frontal cortex activities because the activities of the left frontal cortex on favorite stimuli become unique. Bajoulvand et. al. showed that frontal cortex activities had a close relation to human preference because the sensing position was frontal lobe and folk music preference was analyzed based on EEG analysis [3]. Therefore, this result shows the validity of our consideration.

We confirmed that the mean detection accuracies of all preferences obtained by the proposed method were higher than those obtained by the comparative methods. These results suggest that the relationships among EEG frequency bands become unique when creating the preferences on stimuli and they are different on each preference. Therefore, the detection of a subject's favorite sounds becomes easy when the EEG signals are analyzed while the gray associate degrees are used as the EEG feature and the SVM is used as the classifier. Sawata et. al. showed that favorite music was detected using SVM with mean accuracies of 83.6% or more, although 12 channels were used to record the EEG signals [2]. Therefore, this result shows the validity of our consideration.

The standard deviations of the detection accuracies for all methods were 0.03 or less, which suggest that the EEG signals and the standard deviations of the detection accuracies became high when the distributions of the extracted EEG features and the distributions related to each preference pattern were stable and the left frontal cortex activities related to the preference were widely varied.

The means of the detection accuracies of Dislike and Other were 60% or less, which were substantially lower than for FavoriteSound. These results suggest that detection is difficult for negative stimuli, such as DislikeSound, and borderline cases, such as Other, when the left frontal pole is the sensing position for the EEG analysis. If responses to negative stimuli and/or borderline case are detected, the sensing positions must be changed.

5. Conclusions

This paper introduced an EEG analysis method to detect human preferences. The proposed method adopted EEG recording technique, extraction techniques of EEG features and detection methods of preferences. In EEG measurement, the sensing point was Fp1 (the left frontal lobe). The gray association degree was used to extract the EEG feature and an SVM was used to detect the preferences of humans for particular sounds. In order to show the effectiveness of the proposed method, experiments were conducted with real EEG data. Using all comparative methods with accuracies of 76.53% or more, the mean detection accuracies on FavoriteSound were confirmed to be the highest while the proposed method produced higher accuracies for FavoriteSound, DislikeSound, and Other. These results suggest that the detection of subject's favorite sounds becomes easy when the EEG signals are analyzed while the gray associate degrees are used as the EEG

feature and the SVM is used as the classifier. However, we confirmed that the detection accuracies of Dislike and Other were 60% or less, which were substantially lower than the results for the detections of FavoriteSound. These results suggest that the detection of negative stimuli, such as DislikeSound, and borderline cases, such as Other, are difficult when the sensing position for the EEG analysis is the left frontal pole.

For future research, we will change the sensing positions to detect other preferences, such as Dislike and Other.

Acknowledgment

References

[1] S. Ito, M. Ito, M. Fukumi, "An Electroencephalogram Analysis Method to Detect Preference Using Gray Association Degree" in Proc. of the 17th International Conference on Electronics, Information, and Communication (ICEIC) 2018, Honolulu, 2018. http://dx.doi.org/10.23919/ELINFOCOM.2018.8330622

[2] R. Sawata, T. Ogawa, M. Haseyama, "Human-centered favorite music estimation: EEG-based extraction of audio features reflecting individual preference" Proc. of 2015 IEEE International Conference on Digital Signal Processing, 818-822, 2015.

[3] A. Bajoulvand, R. Z. Marandi, M. R. Daliri, S. H. Sabzpoushan, "Analysis of folk music preference of people from different ethnic groups using kernel-based methods on EEG signals" Applied Mathematics and Computation, **307**, 62-70, 2017.

[4] W.Park, D.-H. Kim, S.-P. Kim, J.-H. Lee, L. Kim, "Gamma EEG Correlates of Haptic Preferences for a Dial Interface" IEEE Access, **6**, 22324-22331, 2018.

[5] Y. Wang, X. Gao, B. Hong, C. Jia, S. Gao, "Brain–Computer Interfaces Based on Visual Evoked Potentials" IEEE Eng Med Biol Mag., **27**(5), 64-71, 2008.

[6] R. J. Davison, "Anterior Cerebral Asymmetry and the Nature of Emotion" Brain Cog., **20**(1), 125–151, 1995.

[7] R. J. Davison, "Anterior Electrophysiological Asymmetries, Emotion, and Depression: Conceptual and Methodological Conundrums" Psychophysiology, **35**(5), 607–614, 1998.

[8] J. B. Allen, "Issues and Assumptions on the Road from Raw Signals to Metrics of Frontal EEG Asymmetry in Emotion" Biol Psychol., **67**(1–2), 183–218, 2004. http://dx.doi.org/10.1016/j.biopsycho.2004.03.007

[9] J. A. Coan, J. B. Allen, P. E. Mcknight, "A Capability Model of Individual Differences in Frontal EEG Asymmetry" Biol Psychol., **72**(2), 198–207, 2006. http://dx.doi.org/10.1016/j.biopsycho.2005.10.003

[10] F. Lotte, M. Congedo, F. Lecuyer, B. Arnaldi, "A Review of Classification Algorithms for EEG-based Brain-Computer Interfaces" J. Neural Eng., **4**(2), R1–R13, 2007. http://dx.doi.org/10.1088/1741-2560/4/2/R01

[11] H. Lee, S. Choi, "PCA+HMM+SVM for EEG Pattern Classification" Proc. 7th Int. Symp. on Signal Processing and its Application, **1**, 541–544, 2003.

[12] T. Hoya, G. Hori, H. Bakardjian, S. Nishimura, T. Suzuki, Y. Miyawaki, A. Funase, J. Cao, "Classification of Single-trial EEG Signals by a Combined Principal + Independent Component Analysis and Probabilistic Neural Network Approach, Proc. ICA2003, **197**, 197–202, 2003.

[13] J. F. Borisoff, S. G. Mason, A. Bashashati, G. E. Birch, "Brain-Computer Interface Design for Asynchronous Control Applications: Improvements to the LF-ASD Asynchronous Brain Switch" IEEE Trans. Biomed. Eng., **51**(6), 985–992, 2004. http://dx.doi.org/10.1109/TBME.2004.827078

[14] S. Ito, M. Ito, K. Sato, S. Fujisawa, M. Fukumi, "Preference Classification Method Using EEG Analysis Based on Gray Theory and Personality Analysis" OJCSIT, **4**(3), 276–280, 2014.

[15] B. Blankertz, G. Curio, K. R. Muller, "Classifying Single Trial EEG: Towards Brain–Computer Interfacing" Adv. Neural Inf. Process. Syst., NIPS 01, **14**, 157–164, 2002.

[16] R. Khorsrowabadi, C. Q. Hiok, A. Watab, K. A. Kai, "EEG-based Emotion Recognition Using Self-Organizing Map for Boundary Detection" Proc. of ICPR, 4242–4245, 2010.

[17] T. Felzer, B. Freisieben, "Analyzing EEG Signals Using the Probability Estimating Guarded Neural Classifier" IEEE Trans. Neural Syst. Rehabil. Eng., **11**(4), 361–371, 2003.

[18] K.-L. Wen, Gray Systems: Modeling and Prediction, Alibris, Amazon, 2004.

Designing Experiments: 3 Level Full Factorial Design and Variation of Processing Parameters Methods for Polymer Colors

Jamal Al Sadi*

Department of Mechanical Engineering, (HCT-ADMC). Shakhbout street, Abu Dhabi, UAE

A R T I C L E   I N F O

A B S T R A C T

In this work, we investigate the effects of variation of processing parameters on the quality of dispersion of polycarbonate compound. In order to achieve appropriate pigments dispersion, we performed compounding process parameters optimizations, by investigating three processing parameters, temperature, screw speed, and feed rate. We utilized experimental design for the optimization of process parameters based upon three levels full factorial response surface methodology was utilized. The experimental designs, statistical and numerical optimization were performed using design expert software. Statistical equation was developed to understand individual parameters interactions on the values of color. The model was established as statistically significant based on diagnostic tests performed. Our analysis of variance (ANOVA) illustrates that the parameters of color (dL*, da* and db*) are affected by the three investigated parameters. The process parameters required to attain color values in a minimum desirable deviation dE* of 0.8 were found through optimization to be equal to 245.26 ℃, 741.27 rpm, and 24.72 kg/hr. Furthermore, we also demonstrate variations of the processing variables while other parameters remained constant (General Trends). Both strategies generated process parameters that were statistically significant.

1. Introduction

Because color has a vital function for the production of polycarbonate pigments, materials need to be extruded with good dispersion properties and uniform particle sizes. In order to yield plastic with a commercialisable color, adding pigments to it is usually required. However, a great challenge is likely to be faced to attain the required color from the first attempt. Several variables affect the color properties of polymers compounding during their extrusion steps, including temperature, screw speed, feed rate, residence time, and screw configuration. Several researchers investigated the effects of such process variables on yielded color during polymers compounding [1- 2]. Being constituted of chemical species, it is likely that the pigments will take part in chemical reactions depending on process conditions. Thus, the correct selection of the right variables is vital to attain the color requirements. Furthermore, the time-temperature relationship can also affect the polymer characterestics. The required pigments dispersion and good uniformity can be attained by decreasing the viscosity of the resin and increasing the mixing time [3].

Spectrophotometers can serve as important measures to control the quality, quantify color, and numerically compare variations in colors [3].

Allowable tolerance limits in particular terms of dL*, da*, db* or dE* are usually chosed by the client; however, for the polycarbonate grade-3 under this study, limits were equal or less ≤ 1.0 for dE or ≤ 0.6 for dL*, da*, db* [4]. The deviation in L*, a*, b* is represented as "dE*, where

$$dE^* = [(dL^{*\,2}) + (da^{*\,2}) + (db^{*\,2})] \qquad (1)$$

Instead of using absolute values of color, color differences concerning target values regarding dL*, da* and db* are used. The total change in color, dE* is used to represent the color difference in the CIELAB color space [5].

Design of Experiments (DOE) is a planned approach that allows an experimenter to plan the experiments and determine cause-and-effect relationships. DOE is extensively used in numerous areas of science because it reduces the number of experiment that need to be performed.

*Jamal Alsadi, (HCT-ADMC), Email: jalsadi@hct.ac.ae

For optimal dispersion of these pigments, the optimization of extrusion process parameters is required. Researchers designed experiments to evaluate the effect of process parameters on colour properties of a compounded polycarbonate grade. A regression model was generated. Several factors were contributed to color mismatch. Such factors need to be studied to understand their effect on output colour [6, 7, 8].

Many experimental designs have been recognized as useful techniques to optimize process variables. A modified general factorial DOE has been employed for investigating the effect of changes in compounding process variables on gloss and surface appearance of a PVC sheet. [9] .Different types of RSM designs are available, including a factorial design, central composite design (CCD), Box-Behnken design, and D-optimal design. [10] The execution of a DOE involving the Box-Behnken design (BBD) has been reported to determine a relationship between processing parameters and viscosity variation for a wood-plastic compound [11].

The BBD, being a combined array design, requires fewer runs than Taguchi's crossed array designs and allows estimation for significant interactions. [12] It is the most efficient design in terms of runs and requires only three levels of each factor in order to generate a quadratic model. [12,13] To estimate curvature, other designs require either five levels of each factor such as in a central composite design (CCD) or even more experimental runs such as in a three level factorial design.

Analysis-of-variance (ANOVA) is essential to validate the significance and fitness of the model; it explains, whether the developed quadratic model is meaningful. It investigated the bearing of process parameters and interaction of these parameters. The robustness of RSM designs is ensured by considering the propagation of error (POE). POEs, a measure of the standard deviation of the transmitted variability in the output response, are caused by fluctuations in significant controllable process variables during experimentation assuming uncontrollable factors (noise) to be zero. [14].

In this study, an experimental investigation of the processing parameters was conducted using DOE. DOE was utilized to determine the optimum number of experiments to be run so that sufficient data was available for analysis. The designs were prepared for three processing parameters: temperature, speed, and feed rate. The effect of the processing parameters on output response parameters was studied. Experimentation for various grades was carried out to observe the effect of controlled variation of different processing parameters on the colour attributes of compounded plastics. The results were analyzed to determine an optimum set of processing parameters in order to ensure minimum wastages and timely delivery of orders.

Statistical Design of Experiments can be used to study the color responses to variation in these processing parameters with the help of methods such as the Response Surface Methodology (RSM). In this approach the first step is to properly design experiments in order to evaluate model parameters efficiently after performing experiments. Second step is to develop a second order polynomial for the responses [14].

$$y = \beta_o + \sum_{i=1}^{k} \beta_{ij} x_i + \sum_{i=1}^{k} \beta_{ij} x_i^2 + \sum \sum_{i<j=2}^{k} \beta_{ij} x_i x_j + \varepsilon$$

(2)

Where y is the predicted response, β_0 is a constant, β_i is the ith linear coefficient, β_{ii} is the ith quadratic coefficient, β_{ij} is the ith interaction coefficient, x_i is the independent variable, k is number of factors and ε is error. Coefficients of the model predicted through regression of the obtained experimental data. Details of parameter estimations for the model done by these authors are reported elsewhere [13]. RSM is a collection of statistical and mathematical techniques useful for developing, improving and optimizing process. The Three-level full factorial design is one of the most powerful and efficient experimental design among other response surface designs (central composite, Doehlert matrix, and Box Behnken designs). The ultimate aim of the present study is to employ the 3 level full factorial design to optimize the processing parameters to have a minimum deviation in color properties (dE* < 0.8)

Focus was extended to study the variations of independent of processing parameters. Parameters were usedare such as temp, speed, and feed rate were used to affect the dependent responses for consistent output color (L*,a*,b*,dE*). The procedure of controlling the variations of two Processing parameters and keeping the third parameter constant (general trends), focused on the variation of the optimal federates parameters to achieve a minimum desirable deviation dE* of 0.3.

Design optimization of the two procedures to precisely determine cause and effect relationships. Both designs yielded models that were statistically significant and optimal color were found.

2. Experimental Set up

Experimentation was carried out for the investigated material at Industrial Plant. A blend of two polycarbonate resins was used along with four different pigments, the color formulation of these grades in parts per hundred (PPH) is presented in Table 1. The melt flow index (MFI) for Resin 1 was 25 gm/10min, and that of Resin 2 was 6.5 gm/10min.

Table 1: Color formulation used for investigated grade

S.No	Type	PPH
1	Resin 1	33
2	Resin 2	67
3	Pigment A	0.20
4	Pigment B	0.05
5	Pigment C	0.0004
6	Pigment D	0.0016
7	Pigemnt E	0.0710

3. Design of Experiments

The Design of Experiments (DOE) containing the 27 experimental runs was used to implement a Three-Level Full Factorial Design and a DOE containing 9 experimental runs was used to implement variations of the processing parameters while other parameters are constant (General Trends) are shown in Table 2 and Table 3 respectively .

Table 2: Design level in actual and coded unit

Parameters	Units	3 Levels		
		-1	0	+1
Temperature	°C	230	255	280
Speed	rpm	700	750	800
Feed rate	kg/h	20	25	30

Table 3: Processing Parameters Variables

Speed RPM	BZ1 (°C)	BZ2 (°C)	BZ3 (°C)	BZ4 (°C)	BZ5 (°C)	BZ6 (°C)	BZ7 (°C)	BZ8 (°C)	BZ9 (°C)	DZ1 (°C)	F. R Kg/hr
750	70	195	230	230	230	230	230	230	230	230	25
750	70	195	255	255	255	255	255	255	255	255	25
750	70	195	280	280	280	280	280	280	280	280	25
700	70	195	255	255	255	255	255	255	255	255	25
750	70	195	255	255	255	255	255	255	255	255	25
800	70	195	255	255	255	255	255	255	255	255	25
750	70	195	255	255	255	255	255	255	255	255	20
750	70	195	255	255	255	255	255	255	255	255	25
750	70	195	255	255	255	255	255	255	255	255	30

3.1 3 Level Full Factorial Design

The processing was carried out on a twin- screw extruder of 25.5 mm diameter, with ratios of L/D=37 and D_o/D_i=1.55. The materials were extruded in an intermeshing , ZSK26- Coperion - Germany ,27 kW twin co rotating screw extruder (TSE). The three process parameters, the temperature of the heating zones, feed rate to the extruder, and the screw speed were considered in the experimental design, and the levels used are shown in Table 2. Parameters were varied on 27 different treatments with additional five center points, the total of runs are (32 treatments) for 3 level full factorial response method to study their effects on color. The additional five centre points were added to estimate the experimental error and for the detection of nonlinearity in the responses [13].

Extruded melt was quenched in cold water, air dried and then pelletized. Using injection molding, the pellets from each run were molded into three rectangular chips (3x2x0.1") after which their values (CIE L*, a*, b*) were investigated by utilizing a spectrophotometer (CE 7000A, X rite- Inc. USA). Target color output for these values were L*= 70.04, a*= 3.41, and b*= 18.09. Statistical analysis of data was performed using The Design Expert Software (Version 8, Stat-Ease Inc. USA) to quantify and relate the effects of variables at a confidence interval of 95%. In order to attain zero deviation from target color, numerical optimization of the data was performed.

3.2 Variations of the Processing Parameters (General Trends)

Because color is directly related to the process parameters involved, herein, we performed a control study to investigate the effects of operating variables (temperature, speed, and feed rate) on color. Three processing parameters were controlled individually at three different stages, while fixing all other parameters (GT). Based on our observed strong correlations between the processing variables and the color generated, we conclude the following recommendations: Flow rate was 20 kg/hr, 25 kg/hr, and 30 kg/hr, at a speed of 750 rpm and a temperature of 255 °C [15-17] .

 The selected processing temperatures were 230°C, 255°C and 280°C with a speed and flow rate fixed at the middle values (750 rpm and 25 kg/hr, respectively). A similar procedure was followed for both speed and flow rate. The selected speeds were 700, 750 and 800 rpm and the selected flow rates were 20, 25, and 30 kg/hr. The following tables show the experimental processing conditions. The general trends (GT) experimental design is shown in Table 3.

Assuming that the aforementioned variables were utilized, in this work we suggest optimized process parmeters to attain plastic grade color consistency.

4. Results

Using analysis of variance (ANOVA), correlations between variables were investigated and processing parameters were optimized to generate resulting color properties. The Design Expert software was used to the effects of the operating process parmeters on dL*, da*, db*. We performed sequential F-tests, by utilizing an initial linear model, modified by subsequently adding suitable linear or quadratic terms [20]. The F-statistic was considered for each type of model, and the highest order model with significant terms was chosen. Analysis of all tristimulus values was based on the same process. Based on the F statistics ≤ 0.05 and probability values ≤ 0.1, only significant terms were added. Our ANOVA sequential model sum of squares results for dL*, da*, db* are given in Table 4.

The highest order model with significant terms (*Prob > F* is less than 0.05) are the 2F models, and are applicable to describe the dL*, da* and db* responses. R^2 values (Table 4) provide confirmation and suggest that a variability of 78% in dL*, da* and db* is decreased to 74 %. The remainng variability is unexplained and can be ascribed to noise. The "Predicted R-Squared" and the "Adjacent R-Squared" are in rational agreement. A signal to noise ratio larger than 4 is usually desirable since it suggests that the model can be utilized for navigation into the design space, and can be quantified by "Adequate Precision". The "Lack of the fit" test was utilized to compare between the residual and the pure error, and resembled a p-value > 0.05, which is insignificant.

Table 4: ANOVA results for dL*, da* and db* response

Response	Significant Terms	R^2	Predicted R^2	Adjacent R^2	Adequate Precision
dL*	A, B, C , AB, AC, BC	0.78	0.38	0.55	9.70
da*	B, C, BC	0.75	0.24	0.39	8.53
db*	B,C	0.75	0.28	0.30	8.61

To generate predicted responses for the dL*, da* db*, several models based on linear regression were produced as given in Table 5. The effects of temperature, speed, and feed rate operating variables and their interactions are represented by the polynomial equations. The effect of these variables on the responses correlate with the coefficient values. The differnce between predicted and actual values is in the 0.2 limit suggesting a strong agreement between these 2 values.

RSM curves were used to investigate interactions and optimize process parameters variables. The contour graph (Figure 1a) illustrates the first order relation between temperature and speed, at the feed rate of 24.71kg/hr for dL*, and resembles that various temperature and speed combinations can satisfy the required objective. Moreover, the contour graph (Figure 1b) illustrates the relation between temperature and the feed rate of 741.2 rpm. Finally, the contour graph (Figure 1c) illustrates the relation between speed and feed rate at temperature of 245.2 °C. The global optimal value (dL* = 0.0) is realized at 245.2 °C, 741.2 rpm and 24.71 kg/hr as demonstrated in the graphs. The maximum and minmum dL* values at 95% confidence interval are both equal to 0.11.

Table 5: Regression Model for dL*, da* and db*

Response	Regression model
dL*	+63.86390 - 0.19647 * Temperature -0.065085 * Speed - 0.99472 * Feed Rate +1.84353E^{-004} * Temperature * Speed +1.96624E^{-003} * Temperature * Feed Rate + 6.39611E^{-004} * Speed * Feed Rate
da*	+14.59778 - 0.018496 * Speed - 0.47296 * Feed Rate + 5.98224E^{-004} * Speed * Feed Rate
db*	+4.08697 - 4.78866E^{-003} * Speed - 0.029746 * Feed Rate

A= Temperature,   B= Speed,   C = Feed rate

Table 6: dL*, db* and da* Actual and Predicted Values

Color	dL*		da*		db*	
Run NOs	Actual Value	Pred. Value	Actual Value	Pred. Value	Actual Value	Pred. value
1	-0.52	-0.5	0.61	0.57	0.31	0.14
2	0.16	0.093	0.23	0.12	-0.17	-0.25
3	-0.59	-0.57	-0.3	-0.087	-0.76	-0.34
4	0.11	-0.2	0.29	0.003	-0.22	-0.4
5	0.01	-0.24	-0.16	-0.059	-0.16	-0.49
7	-0.65	-0.36	-0.29	0.12	-0.77	-0.25
8	0.14	-0.029	0.34	0.3	-0.11	0.008
9	-0.53	-0.47	0.027	0.024	-0.38	-0.16
10	0.087	-0.18	0.14	0.03	-0.18	-0.4
11	-0.21	-0.087	0.077	-0.087	-0.56	-0.34
12	-0.44	-0.54	0.65	0.24	0.37	-0.099
13	0.25	0.41	0.27	0.24	-0.14	-0.099
14	-0.55	-0.23	-0.15	-0.059	-0.55	-0.49
15	-0.71	-0.2	-0.05	0.031	-0.93	-0.4
16	-0.53	-0.4	-0.12	-0.031	-0.84	-0.64
18	-0.34	-0.26	-0.22	0.024	-0.35	-0.16
19	0.037	-0.065	0.29	0.24	-0.22	-0.099
20	-0.48	-0.49	0.65	0.3	0.37	0.008
21	-0.2	-0.33	-0.017	-0.087	-0.33	-0.34
22	0.43	0.43	0.32	0.3	0.013	0.008
23	0.027	-0.13	0.24	0.12	-0.24	-0.25
24	0.087	0.2	0.24	0.57	-0.16	0.14
27	-0.04	-0.13	0.05	0.12	-0.31	-0.25
28	-0.15	-0.13	0.12	0.12	-0.013	-0.25
29	-0.19	-0.13	0.093	0.12	-0.083	-0.25
30	-0.02	-0.13	-0.087	-0.087	-0.02	-0.25
31	-0.1	-0.13	-0.09	0.12	-0.06	-0.25
32	-0.033	-0.15	0.13	-0.031	-0.32	-0.64

Figure 2 shows the interaction between speed and feed rate for da* at 245.2 °C, resembling the first order relationship and suggest that several feed rate and speed combinations can satisfy the objective. The optimum value of da* = 0.15 is realized at 741.2 rpm and 24.7 kg/hr. The maximum and minimum allowable values at 95% confidence interval for da* are 0.23 and 0.07 respectively.
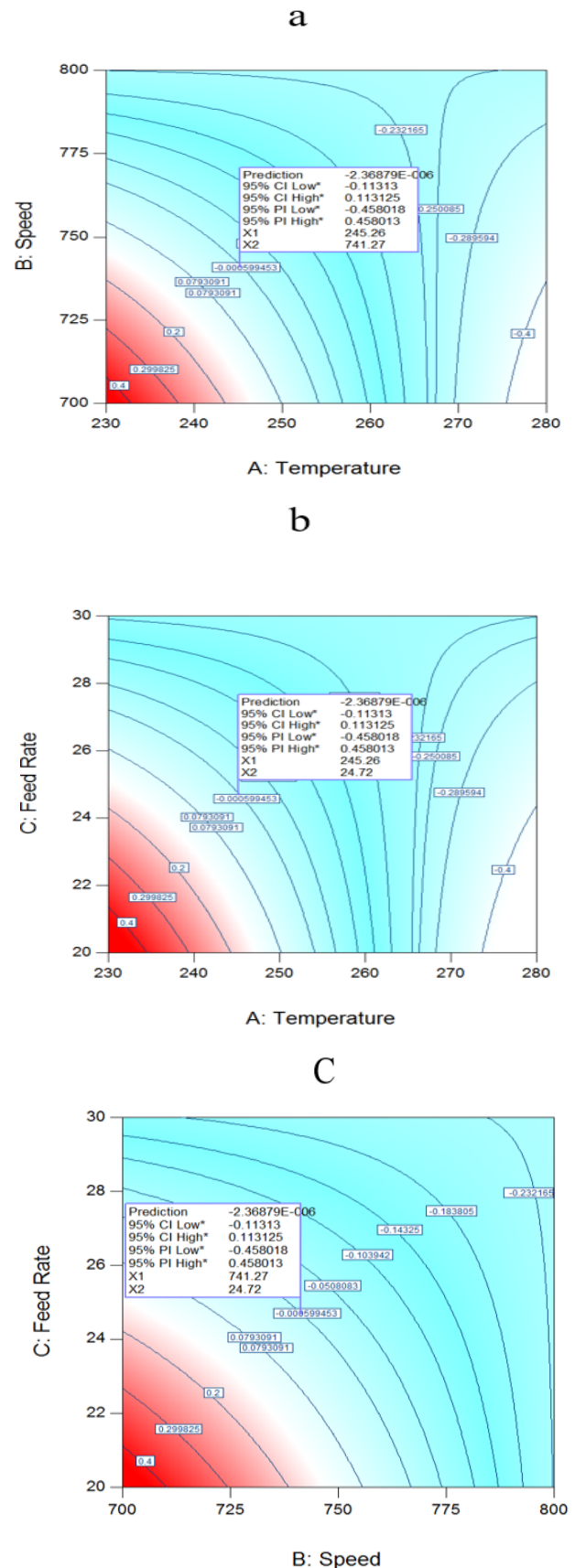
Figure 1: (a) Interaction between temp and speed at 24.7 kg/hr for dL* (b) Temp and feed rate at 741.2 rpm for dL* (c) Speed and feed rate at 245.2 °C for dL*.
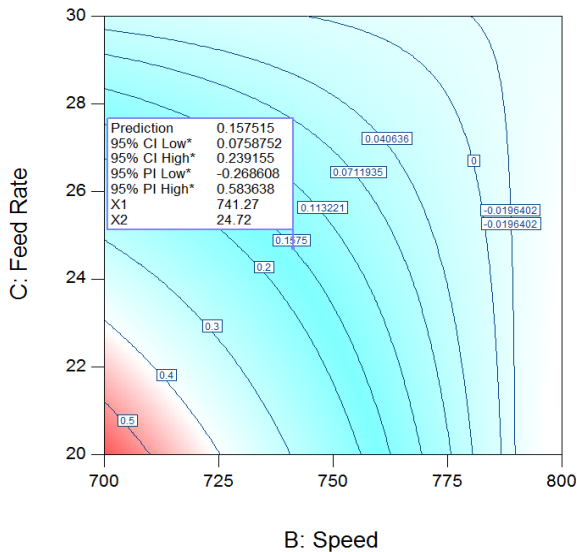
Figure. 2: Speed and feed rate interaction at temperature of 245.2. ºC for da *
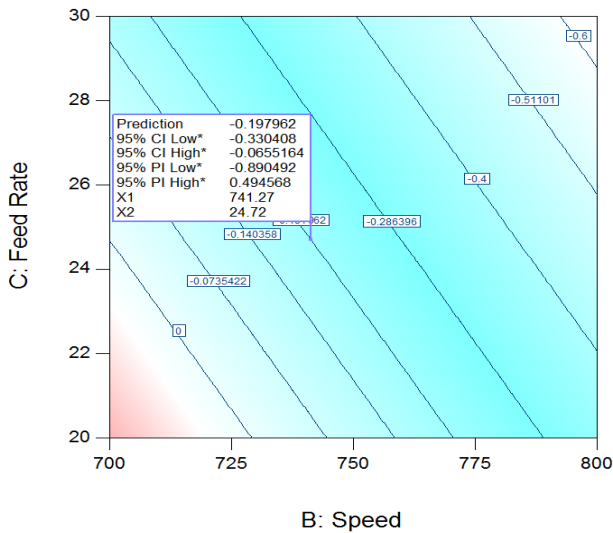
Figure 3: Speed and feed rate contour plot for d b* at 245.2 ºC
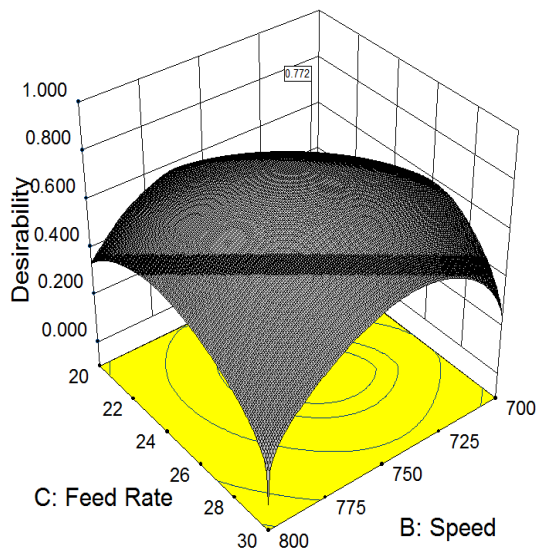
Figure 4: Desirability Graph at temperature of 245.2 ºC
(A= Temp; B= Speed, C= Feed rate)

The interaction between speed and feed rate for db* at 245.2 ºC is shown in Figure 3. The linear behavior includes only speed and feed rate as significant model terms as compared to dL* and da*. Interestingly, at lower values of speed and feed rate around 20 kg/hr and as well as at higher values of speed and feed rate around 24 kg/hr, db* approaches zero, while temperature stays at 245.2 ºC. At the global optimum, the predicted value of db* is equal to -0.19 and the allowable maximum and minimum values at 95% confidence interval are -0.05 and -0.33 respectively.

An important effect on the color responses was exerted by temperature (A), speed (B) and feed rate (C). The optimal settings of the three parameters regarding all responses were realised by utilizing a decision making method based on multiple criteria, and a total desirability function "d" [19]. A desirability function is quantifies the quality and presents a convenient responses comparison method to select the optimal settings; Figure 4 shows a 3D plot of the global desirability D, maintaining a feed rate of 24.7 kg/hr, and indicates that the maximum combined desirability of 77% is attained at 245.2ºC, 741.2 rpm, and 24.7 kg/hr.

5. Effect of variation feed rate on color values

A plot of feed rate variation from 20 kg/hr to 30 kg/hr with a fixed speed at 750 kg/hr and a fixed temperature at 255 ºC is shown in Figure 5; The difference in color decreases initially as the feed rate increases up to 30 kg/hr.[20-22]

A response contours plot of speed versus feed rate at a temperature of 245.2 ºC is shown in Figure 6, and resembles the presence of region that is feasible to achieve target values. The plot shows a region between the da*=0.30 and db*=0.20 contours which illustrates the temperature and speed operation conditions at which the mean responses (dL*, da*, db* and SME) target are met at a fixed feed rate of 24.44 kg/hr.

The results suggest that the optimal tristimulus values of dL* = 0.0, da * = 0.15, db* = - 0.19 are attained at 245.2 ºC, 741.2 rpm, and 22.7 kg/hr. The total minimum deviation in tristimulus values (using equation 1) is equal to 0.25, which is reasonably acceptable as compared with the maximum allowable deviation (dE* = 0.8).
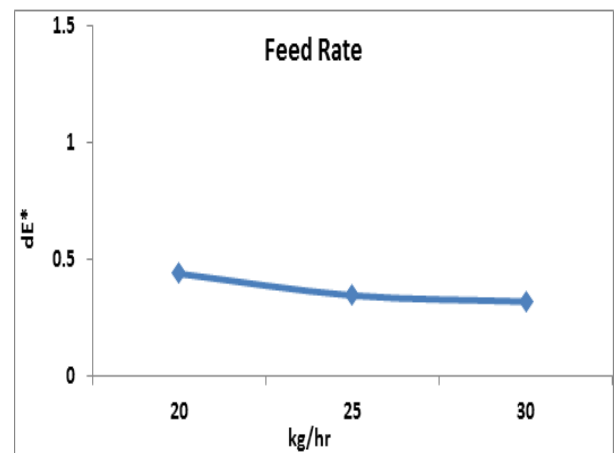
Figure 5: Effect of the variation of feed rate at fixed a fixed temperature of 255 °C, and a fixed speed of 750 rpm.
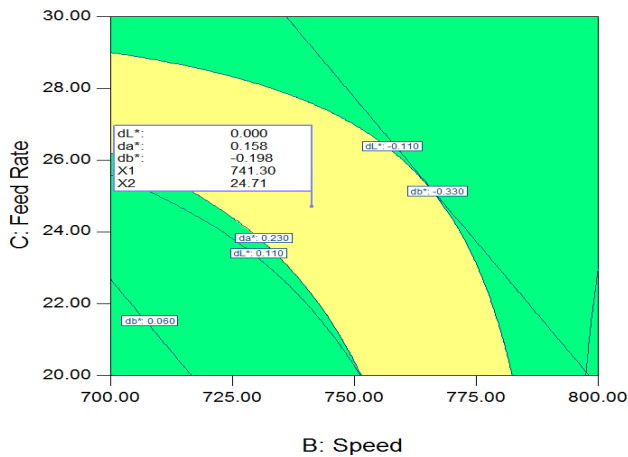
Figure. 6: Speed and feed rate overlay plot at a temperature of 245.2 °C.
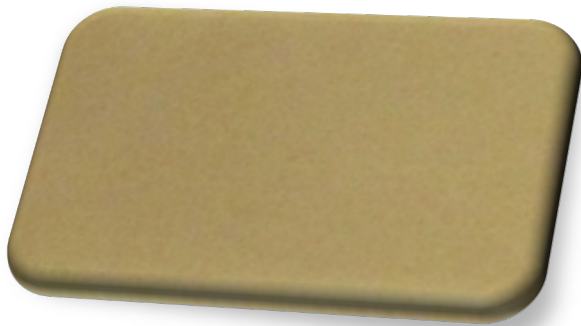
Figure 7 : Rectangular color chip (3x2x0.1")

Using the lab injection molding, the extruded pellets were molded into a rectangular chip (3x2x0.1") after which the color values (CIE L*, a*, b*) were examined by using a spectrophotometer (CE 7000A, X rite- Inc. USA). The image of the specimen is shown in Figure 7.
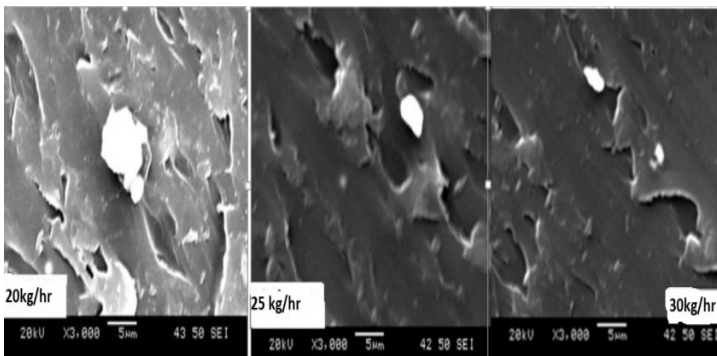
Figure 8: Scanning Electron Microscopy images at feed rates of (a) 20 kg/hr, (b) 25 kg/hr, and (c) 30 kg/hr, at a fixed temperature of 255 °C, and a fixed speed of 750 rpm.

Scanning Electron Microscopy (SEM) images were probed using a Joel 5500 LV at a 20 kV voltage, at different feed rates, and are shown in Figure 8. The images illustrate that primary particles and agglomerates were present in the compounded grades. The images show that larger agglomerated pigments could be observed at a lower feed rate as compared with higher feed rates [20-22].

Conclusions

An experimental design approach which relied on three levels full-factorial surface method was utilized to study the effects of operational process variables, and suggested appropriate predictive models for dL*, da*, and db*. We used a regression model to calculate tristimulus values and validated our work by confiming that a good agreement exists with experimental results. At target values of 245.2 °C, 741.2 rpm, and 24.7 kg/hr, tristimulus values were close, with a minimum total deviation (dE*) of 0.25 at a 95% confidence interval. An experimental design approach, which relied on the interaction between tristimulus color values and processing parameters showed that the minimum color difference throughout the experiment was equal to 0.34 for dE*. As the feed rates increased from 20 kg/hr to 30 kg/hr, the color difference values (dE*) substantially decreased. Finally, an optimum set of processing parameters for the grades of the polycarbonate can be yielded through the utilization of the optimized parameters, and hence reduce colors mismatch so that wastages can be reduced.

References

[1] S. P. Rwei, Distributive Mixing in a Single-Screw Extruder – Evaluation in the Flow Direction, Polymer Engineering and Science, (2001).

[2] A.Y. Wong, Screw Configuration Effects on The Color Mixing Characteristics Of Polymer In Single-Screw Extrusion, The University of Hong Kong, China Tinhua Liu, Sichuan Union University, China (1998).

[3] D. I. Meade, "Introduction to colorant selection and Application Technology," in Coloring of Plastics, Fundamentals, 2nd ed., edited by Robert Charvat (2004).

[4] ASTM D 2244-93. Standard Test Method for Calculation of Color Differences from Instrinentally Measured Color Coordinates. 222-215

[5] Richard Abrams, Plastic additives, and compounding (2001)

[6] Rauwendaal, C., Polymer Mixing, A Self-Study Guide. Hanser Publishers, Munich., 1998.

[7] Rwei, S.P., Distributive Mixing in a Single-Screw Extruder – Evaluation in the Flow Direction. Polymer Engineering and Science, 2001

[8] Mulholland, B.M., Effect of Additives on the Color & Appearance of Plastics. SPE ANTEC. Ticona, Ticona Engineering Polymers ,2007.

[9] Effertz, K., Understanding the Effects of a Compounding Process on the Production of Co-Extruded Vinyl Sheet through the Utilization of Design of Experiments (part II). ANTEC-Conference Proceedings: p. 3574-3578 ,2004

[10] http://itl.nist.gov/div898/handbook. [Online] available accessed in 2013.

[11] Bender, T.M., Characterization of Apparent Viscosity with respect to a PVC-Wood Fiber Extrusion Process. ANTEC-Conference Proceedings, 2002.

[12] Borror, C.N., Montgomery, D. C. and Myers, R. H., Evaluation of Statistical Designs for Experiments Involving Noise Variables. Journal of Quality Technology. Vol 34(1): p. 54-70 , 2002.

[13] Anderson, M.J., and Whitcomb, P. J., RSM Simplified: Optimizing Processes using Response Surface Methods for Design of Experiments. New York, NY: CRC Press. , 2005.

[14] Montgomery, Douglas C., Design and Analysis of Experiments, 6th Ed., John Wiley and Sons, New York,2005.

[15] J. AlSadi1, S. Ahmad1, U. Saeed1, G. Rizvi1, D. Ross2, R. Clarke2, J. Price2. (2012) Execution of 3 level full factorial design to evaluate the process parameters: polymer color properties, Annual Technical conference of the Society of Plastics Engineers (ANTEC), Orlando, Florida, USA, pp. 1-5, 2012.

[16] Saeed.U. AlSadi, .J, Ahmed. S, Rizvi. G., Ross. D, 2013, "Neural Network: a potential approach for error reduction in color values of polycarbonate" Journal Advance polymer technology. 33 (2), Summer 2013, DOI 10.1002/Adv. 21402

[17] J. AlSadi, M. Rabbani, S. Ahmed, G. Rizvi, R. Clarke and D. Ross, (2011) Effect of Processing Parameters on Colour During Compounding, Annual Technical Conference of the Society of Plastics Engineers (ANTEC), Boston, USA, pp. 1-4, 2011

[18] Y. K. Chang, M. K. Bustos, H.Lara, Braz. J. Chem. Eng. vol. 15 n. 4 São Paulo Dec. 1998

[19] M. A. Islam, V. Sakkas, T. A. Albani, J. Hazardous Materials, (170) 2009.

[20] J. Alsadi, U. Saeed, S. Ahmad, G. Rizvi and D. Ross , Processing issues of color mismatch: Rheological characterization of polycarbonate blends, Polymer Engineering & Science, Volume 55, Issue 9, September 2015, Pages: 1994–2001, Version of Record online : 8 DEC 2014, DOI: 10.1002/pen.24041

[21] AlSadi, J., Ahmad, S. , Saeed, U. , Rizvi, G. , Ross, D. , Clarke, R. and Price, J., Effects of Processing Parameters on Color Mismatch During Compounding. Annual Technical conference of the Society of Plastics Engineers (ANTEC), Orlando, Florida, USA ,pp 1-5 , 2012.

[22] J, Alsadi . Colour Mismatch In Compounding Of Plastics: Processing Issues and Rheological Effects (Doctor of Philosophy in Mechanical Engineering), The Faculty of Engineering and Applied Science University of Ontario Institute of Technology-UOIT). Ontario, Canada(2015).

FPGA Implementation of Ultra-High Speed and Configurable Architecture of Direct/Inverse Discrete Wavelet Packet Transform Using Shared Parallel FIR Filters

Mouhamad Chehaitly[1,*], Mohamed Tabaa[2], Fabrice Monteiro[1], Juliana Srour[3], Abbas Dandache[1]

[1]Université de Lorraine, LGIPM, Metz, France

[2]Multidisciplinary Research and Innovation Laboratory (LPRI), EMSI Casablanca, Morocco

[3]Lebanese University, Beirut, Lebanon

A B S T R A C T

This work presents new pipeline-parallel, generic and configurable parallel hardware architectures for the Direct/Inverse Wavelet Packet Transform (DWPT/IDWPT) independent of any specific family of wavelets, implemented in FPGA technology using a parallel architecture of direct FIR filter. We propose in the following paper, new P-parallel structures for the DWPT and IDWPT transforms based on the Mallat binary tree algorithm. Therewith, we developed a P-parallel/modified direct FIR filter architecture using pipeline-parallelize and hardware resource sharing, which provides not only ultra-high speed data processing but also a limited amount of hardware as resources are shared between filters and the bidirectional wavelet packet transformation. This model follows two important strategies: I) a powerful structure pipeline/P-parallel using strict data management and interleaving, II) sharing hardware at different levels in the transformation and between the two DWPT/IDWPT transformations. These architectures are modeled in VHDL at RTL modeling level. They are generic and fully configurable: at synthesis and post-synthesis.

The simulation results show an acceleration of data processing to an approximate value of P-Parallel multiplied by frequency with lower used resources. Furthermore, the impact of tree depth and filters order on throughput is very light due to the linearize architecture of our model. The synthesis was achieved using the Intel Quartus Prime Lite Edition software and targeting the Intel Altera Cyclone FPGA technology.

1. Introduction

1.1. Sate of art

Wavelet transform are considered the rightful heir to Fourier transform, since they are the best signal analysis tool in time and frequency domain.

From a historical point of view, the founder of the wavelets world was J. Morlet in 1980 who developed a new transform called "Wavelet Transform (WT)" to solve the limitations of the different versions of the Fourier transform on signal analysis passage from the time to the frequency domains [1, 2]. The numerical application of the wavelet transform is born with the work of Morlet and Grossman [3], while the idea of wavelet seems to originate in the work of Gabor and Neumann in the late 1940s. Mallat [1, 4], Meyer [5], Chui [6], Daubechies [7] and others, contributed the pioneering work reported in the early monographs.

The discrete version of the wavelet transform called "Discrete Wavelet Transform (DWT)" mixes both the notions of time and frequency scaling using a compact kernel (bit wave = Wavelet). Mallat [1] generalized the discrete wavelet transform that introduce more flexibility on the time scale of data analysis and lead to the concept of Discrete Wavelet Packet Transform (DWPT).

Currently, the architecture of DWPT and Inverse DWPT (IDWPT) are unsuitable in most fields like image processing, video processing, noise reducing, encryption, coding, modulation and others. This explains the large number of researches and

restrains the usage of DWPT in large applications domains, which leads us to develop a new hardware architecture to improve the performance of hardware implementation of DWPT/IDWPT. The major challenges for modern systems can be summarized as: (i) high throughput rates (increasing the volumes of data processed, especially in multimedia application) and (ii) low cost hardware (limited amount of hardware resource) while still providing high performance.

1.2. Related works

Abundant researches can be found in the literature about the hardware implementations of wavelet packet transform. The first work related to the hardware implementation of DWT was recorded in 1994 [8] Denk and Parhi described an orthonormal DWT architecture, which used a lattice Quadrature Mirror Filters (QMF) structure and digit-serial processing techniques. Similarly, Hatem et al. [9] proposed a parallel/sequential QMF architecture in VLSI for the DWT, to reduce the overall numbers of multipliers. After that, the research evolution was started.

Since DWPT retains all the advantages of the discrete wavelet transform, it has becomes an essential standard tool for signal, image and video processing. For this reason, the DWPT topic has attracted a lot of attention in research area. Notably, the most implementations of DWPT are based on the concept of FIR (Finite Impulse Response filter) filter banks [10]. In the literature, many works are addressing the implementation of DWPT/IDWPT in hardware devices such as Processors, GPUs (Graphic Processing Units) or FPGAs, to take full advantage of the powerful and flexibility offered by the DWPT. In this context, the year 1999 viewed the first work to implement the DWPT on a processor [11] but in the last recent years, FPGA became a popular target technology for many works, which become doubtlessly the target implementation technology for DWPT and IDWPT. However, the implementation of DWPT/IDWPT (based of FIR filter banks) has many caveats, which make it hard to fulfill a high throughput rate and low area consumption simultaneously.

To fulfill these challenges, many works implement the DWPT or IDWPT transformations on FPGA like the work of Wu and Hu [12], the authors implemented DWPT/IDWPT using Embedded Instruction Codes (EIC) and strict number of multipliers and adders for the symmetric filters (two multipliers and four adders). In [13], IDWPT architecture was presented based on classical recursive pyramid algorithm (RPA) by using lifting transformation and polyphase decomposition. In [14], a fast and configurable DWPT was presented based on FIFO inputs, dual-port memory, multipliers and adders.

An inspirational architecture of DWPT was presented in [15] by implement the tree algorithm, their implementation on FPGA was discussed in [16]. In [17], the same authors based on two modes developed pipelined architectures of the direct and inverse DWPT transform: serial-word mode and parallel-word mode. In [18], Farahani and Eshgho presented word-serial pipeline architecture of DWPT/IDWPT transforms by using parallel filter structure. To speed up the transform, the same authors updated their architectures by using high-pass and low-pass filters. In addition, they implemented the wavelet packet transform on FPGA by using three types of multipliers [20].

To increase the performance, in [21] a VLSI architecture of DWPT was presented which based on frame-partition architecture. In [22], authors implemented a flexible architecture of DWPT and IDWPT based on register interface and a multiplexing structure.

An algebraic integer architecture of Daubechies DWPT was presented in [23], this architecture was developed to compute one and two dimensions of Daubechies wavelet order 6 (Db6) with adder-less.

1.3. Contributions

This rich literature around DWPT and IDWPT reveals the importance of wavelet transforms in the current and future applications, which motivates us to deeply excavate in DWPT/IDWPT research area. Furthermore, most of recent approaches present some advantages in terms of flexibility, scalability, reliability, or low costing but not all together within the same architecture.

However, our aim is to develop new architecture generic and configurable (in synthesis and post-synthesis) of DWPT and IDWPT, which can provide an ultra-high speed data processing with minimum hardware usage. We already proposed our first pipeline and configurable architecture of DWPT in 2015 [24] and of IDWPT in 2016 [25] based on Mallat binary tree scheme [4] without any presence of parallelization tools. We transformed successfully an exponential algorithm of DWPT/IDWPT, based on classic Mallat binary tree, to a linear algorithm with the conservation of sample speed processing but with decreasing the usage of hardware by a smart managing and interleaving of data along the depth of transformations.

To increase the throughput, we propose in this literature massive pipeline-parallel architectures of DWPT (P-DWPT architecture) and of IDWPT (P-IDWPT architecture). Moreover, a massive pipeline-parallel DWPT-IDWPT architecture with massive sharing of resources: the core of the architecture is based on sharing hardware resources between P-DWPT and P-IDWPT transforms, and also based on a smart sharing of computational resources (multipliers and adders) between the approximation/details functions related to filter banks.

Those architectures are generic and fully configurable relative to the tree depth, the filters order, the quantization, and selection way of processing method (P-DWPT or P-IDWPT).

To evaluate the performance of our architectures, we study the effect of four parameters: "degree of parallelization, transformation depth of DWPT or/and IDWPT, order of low-pass and high-pass filters, and order of quantization" on the "operating frequency or throughput rate" (computational performance) and "resource consumption" (used hardware or hardware cost). Furthermore, we compared our synthesis results with the existing in literature.

1.4. Paper organization

The rest of this literature is organized as follows: Section II, presents a brief overview of wavelet packet transform concepts and the main challenge of parallelization in our context. Section III is dedicated to present our proposed pipeline-parallel architectures and related design paradigms. Then, we summarize the results

obtained for these architectures on FPGA implementation. At the end, we finalize with the comparison, conclusion and the perspective works.

2. Background

From mathematical point of view, wavelet packet transform decomposes the signal into a two sub-signals: detailed signal and approximated signal. To achieve the transforms there are several methods, the famous one is the multi-resolution or multi-cadence analysis algorithm proposed by Mallat [1], which called "Mallat binary tree" or "pyramid algorithm" of wavelet packet transform. This multi-resolution analysis is based on cascade of low-pass and high-pass digital FIR filters. By definition, using Wavelet Transform (WT) theory is that each signal $x(t)$ can be presented by projecting it into a series of scaled and translated functions $\Psi_{s,\tau}(t)$. The original single function is called "Mother Wavelet" and all scaled and translated functions are obtained by translation and expansion of the mother function, as presented in equation (1):

$$\Psi_{s,\tau}(t) = \frac{1}{\sqrt{s}} \Psi\left(\frac{t-\tau}{s}\right) \qquad (1)$$

where $s \in R^*, \tau \in R$.

The value of s affects strictly the bandwidth of Mother Wavelet. When s is less than 1 ($s < 1$), wavelet variance decreases and the basis function Ψ contracts (used to analyze the high frequency signal). While for $s \geq 1$, wavelet variance increases and the basis function Ψ stretched (used to analyze the low frequency signal).

The Continuous Wavelet Transform (CWT) of $x(t)$ is defined as:

$$(s,\tau) = f, \Psi_{s,\tau} > \frac{1}{\sqrt{s}} \int_{-\infty}^{+\infty} x(t)\Psi^*\left(\frac{t-\tau}{s}\right) dt \qquad (2)$$

where * denotes a complex conjugation.

From the wavelet theory, the CWT is too redundant that make it impractical. To solve this problem, in 1992 Daubechies [26] create discrete wavelets transform (DWT) from the CWT by discretizing of scalable and translatable variables as shown in equation (3):

$$\Psi_{j,k}(t) = \frac{1}{\sqrt{s_0^j}} \Psi\left(\frac{t-ks_0^j\tau_0}{s_0^j}\right) \qquad (3)$$

Where the number of wavelets coefficients is finite. To discretize the scalable and translatable variables, we use the dyadic sampling concept that mean we usually choose $s = 2s_0, s_0 = constant$ and $\tau_0 = 1$ (give us a dyadic sampling in time). The Discretize of the translation and dilation contraction parameters of the wavelet in (2) leads to DWT presented in (4):

$$(j,k) = 2^{-j/2} \int_{-\infty}^{+\infty} x(t)\Psi^*(2^{-j}t - k) dt \qquad (4)$$

where $j, k \in Z$.

Another important function called "scaling function" φ is similar to the wavelet function $\Psi(t)$. According to the multiresolution wavelet theory, the φ is decomposed into finer and finer detail (in multiresolution stairs). As seen in equation (5), the $\varphi(t)$ has also two integer subscripts or parameters j and k.

$$\varphi_{j,k}(t) = 2^{j/2} \varphi(2^j t - k) \qquad (5)$$

where j specifies the magnitude $2^{j/2}$ as well as the scale $2^j t$ of the function, and k specifies the position (integer location, translation or shift) of the function.

In fact, DWT in signal analysis theory can be implemented with a different ways that is presented in equation (4) by using the concept of non-uniform filter bank. The filtering operation of input signal $x(t)$ is done iteratively and generates two separates sub-signals with two different spectrum, the upper half of the spectrum contains the high frequency component of input signal which is analyzed by the wavelet function and the lower part contains low frequency component of the same input signal which is continued to the next stage. The corresponding obtained coefficients that present the first filtered signal are called detailed coefficients and the second filtered signal are called smooth coefficients. Classically, both low-pass digital filter h and high-pass digital filter g are obtained from the scaling function and their corresponding mother wavelets.

We suppose h and g like a FIR filters non-recursive with L length, the transfer functions of h and g can be represented as follows:

$$H(Z) = H_0 + H_1 z^{-1} + H_2 z^{-2} + \cdots + H_{L-1} z^{-(L-1)} \qquad (6)$$
$$G(Z) = G_0 + G_1 z^{-1} + G_2 z^{-2} + \cdots + G_{L-1} z^{-(L-1)} \qquad (7)$$

Where z^{-1} denotes a delays of $1 \times sampling\ period$.

The orthogonal multiresolution decomposition of DWT can be carried out efficiently by using Mallat tree algorithm or pyramid algorithm [1]. This structure constitutes a bank of filters in QMF analysis, where the scaling functions and wavelet functions are realized using scale relations (6) and (7) and can be given as:

$$\phi(t) = \sum_n h(n)\phi(2t - n) \qquad (8)$$
$$\Psi(t) = \sum_n g(n)\Psi(2t - n) \qquad (9)$$

Where $\phi(t) = [\phi_1(t)\phi_2(t)\dots\phi_r(t)]^T$ and

$\Psi(t) = [\Psi_1(t)\Psi_2(t)\dots\Psi_r(t)]^T$ forms the set of scaling functions and their corresponding wavelets. The suffix r denotes the number of wavelets and is dubbed as multiplicity.

According to the wavelet theory based of filter bank concept, any arbitrary signal $x(t)$ can be expanded into a sum of scaling and wavelet functions. The discrete wavelet transform of target signal $x(t) \in L^2(\mathbb{R})$ is given by:

$$(t) = \sum_{(i,l)\in l} \sum \sum_{k\in Z} d_l^i(k)\Psi_{l,k}^i(t) \qquad (10)$$

2.1. DWPT under focus

The decomposition DWPT at each resolution level, based on Mallat binary tree, can be presented as tree shape. To demonstrate a general transform of DWPT based of Filter Bank concept, we use Figure 1.

The input signal ($D_0^0[k]$) in coming data level will be decomposed into a high frequency signal (smooth or

approximation coefficient $D_1^0[k]$) and a low frequency signal (detailed coefficient $D_1^1[k]$) in level 1 by using low-pass / high-pass filters respectively. Then data path will be down sampling by a factor of two (lead to half size of original signal) and so on for the other levels (number of level equal the depth in Mallat tree algorithm). As a global view, the corresponding wavelet coefficients in different level are derived as follows:

$$D_l^{2i}(k) = \sum_n h(n) D_{l-1}^i(2k-n) \tag{11}$$
$$D_l^{2i+1}(k) = \sum_n g(n) D_{l-1}^i(2k-n) \tag{12}$$

where $i = 0, 1, \ldots, (2^{(l-1)} - 1)$ and $h(n)$ and $g(n)$ are low pass and high pass filters, respectively.
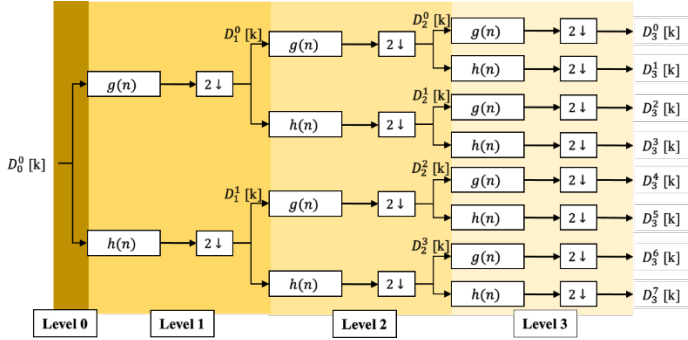
Figure 1. Three-level DWPT decomposition tree

Considering the FIR (Finite Impulse Response) non-recursive implementation scheme of length L for the h and g filters, the corresponding transfer functions can be represented as equations (6) and (7). The number of filters coefficients depends on the mother wavelet. The choice of this later depends upon the required application and its properties. Examples are Daubechies family where the filter length is $2N$ (Order N is strict positive integer), and Coiflets family where the filter length is $6N$ (Order N is strict positive integer), etc.

From the wavelet theory, the reconstruction of original signal that decomposed by direct DWPT is achieved by using the inverse of wavelet packet transform (IDWPT). Like the direct way, the reconstruction operation is also performed by using an iterative method. This mean, for each pair coefficients at level $l + 1$ of the tree we can calculate the wavelet packets coefficients at the previous level l as shown in equation 13:

$$D_l^i(k) = \sum_n \bar{h}(n) D_{l+1}^{2i}(2k-n)$$
$$+ \sum_n \bar{g}(n) D_{l+1}^{2i}(2k-n) \tag{13}$$

In sample representation of equation (13), it can be represented in Figure 2.

For example to reconstruct the D_{j+1}^0 coefficient at level $j + 1$, we used the approximation D_j^0 and detailed D_j^1 coefficients at the previous level j. The reconstruction operation is provided by adding zeros and convolving the results with the reconstruction filters.
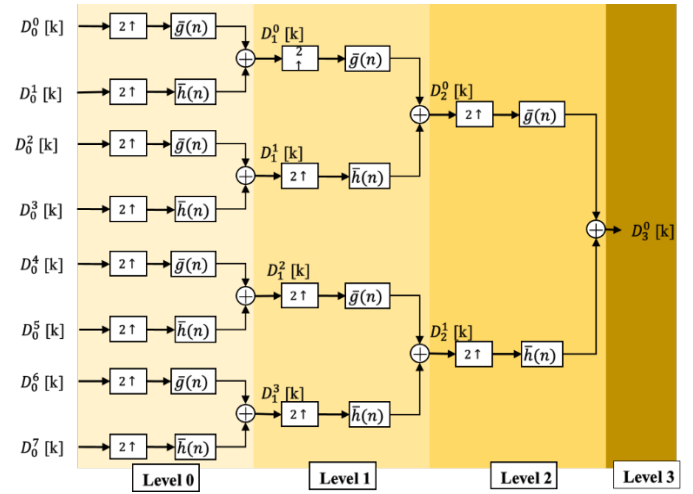
Figure 2. Three-level IDWPT reconstruction tree

To perfectly reconstruct the original signal (before the DWPT transformation), it is sufficient to use the concept of QMF filters that satisfy the following relations:

$$g(i) = (-1)^i h(L - 1 + i)$$

$$\bar{h}(i) = h(L - 1 - i)$$

$$\bar{g}(i) = (-1)^{i+1} h(i) = g(L - 1 + i) \tag{14}$$

2.3. Parallelization challenge of DWPT and IDWPT

In order to achieve our aim to develop a high performance hardware architecture of DWPT and IDWPT, we think for the first time by a classic way to parallelize the Mallat binary tree as shown in figures 3 and 4.
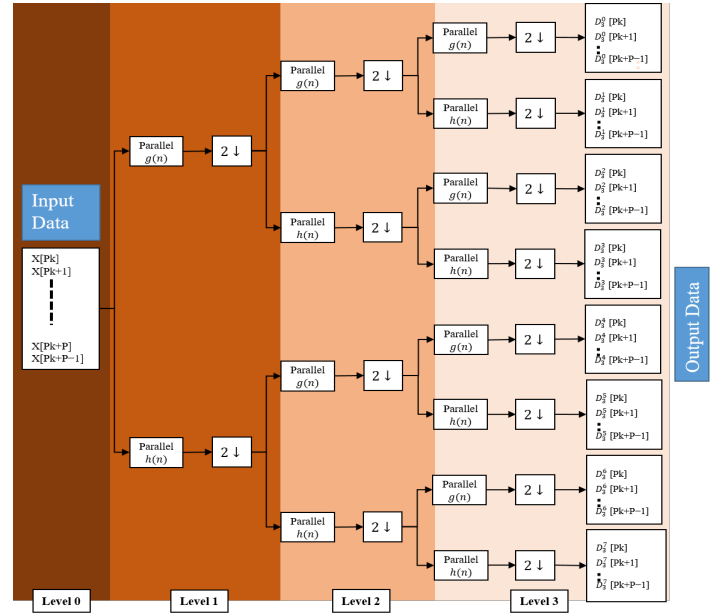
Figure 3. Classic P-parallel, Three-level DWPT
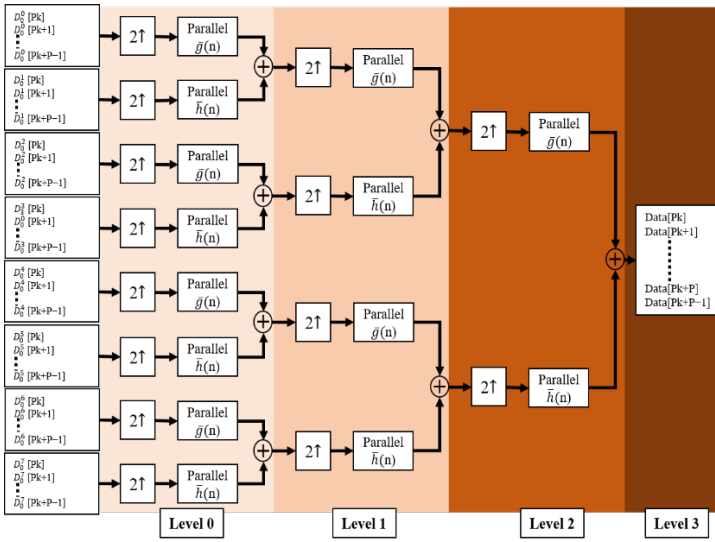
Figure 4. Classic P-parallel, Three-level IDWPT

A traditional possible parallelization solution is presented in figures 3 and 4, where we show a three-level decomposition, reconstruction tree respectively, with theoretical P-parallel degree and a P-parallel filters bank consists of wavelet functions to be able to treat P sampling in each slot time. This classic P-parallel architecture is unrealizable because the number of reconstruction filters and/or decomposition filters increases exponentially as a function of depth order. In precision, the number of filters needed to implement this architecture is around $P * (2^{depth+1} - 1)$. For example, with degree of parallelism $P = 16$ and $depth = 5$, we need 1008 low/high pass filters for DWPT and another 1008 low/high pass filters for IDWPT. The implementation of 2016 filters is a potential problem, which make it unable to implement.

3. Our Parallel-Pipeline architectures with sharing resources

- At a given stage of the Mallat binary tree, the data rate processed by any filter is twice as fast as that of any filter on the adjacent stage on the input side in Figure 1.
- While, the data rate processed by any filter is twice as slow as that of any filter on the adjacent output-side stage in Figure 2 (factor 2 upper sampling from stage to stage).
- In addition, the amount of data processed in a k-level is $2^{depth} - k$ times the amount to process at level 1. Thus, the total amount of data to be processed in level k is the same as in the first or last level ($2^{-k} \times 2^k = 1$).

Based on this big regularity that provide high throughput rate with lower hardware resources, we build an evolution parallel-pipeline architecture of DWPT and IDWPT based on Mallat binary tree as shown in section III. To achieve our objective that ensure high throughput with low hardware consumption in our parallel architecture of DWPT and IDWPT, we have to modify the high/low pass FIR filters.

Under the strategy to reduce area consumption, instead of using two filters we merge in the same architecture the functionality of low-pass and high-pass filters in a single block filter. Furthermore, we modify the single block to process P sampling in two-clock cycle. In figure 5, we present our P-Parallel modified transposed FIR filter.

This new architecture is linearized to be similar to coding theory where the Serial transposed FIR filter is like a Single-Input Single-Output (SISO) system and the Parallel transposed FIR filter is like a Multiple-Input Multiple-Output (MIMO) system. Therefore, this heterogeneous architecture provides the processing of P inputs signals and consequently P outputs signals, in each clock cycle.
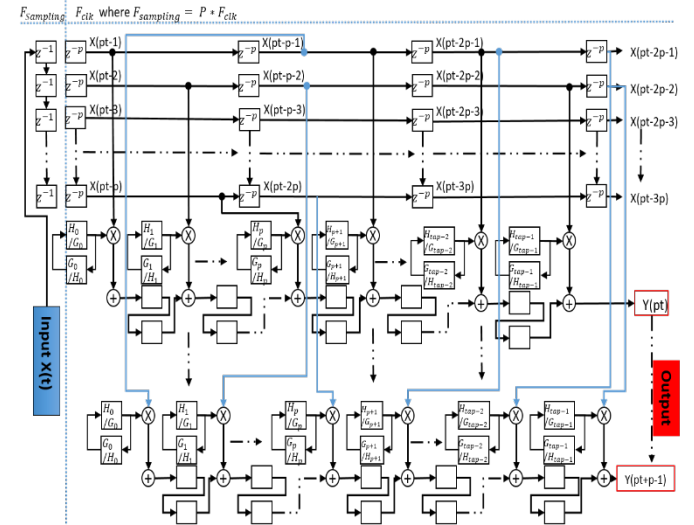
Figure 5. Architecture of a single modified FIR filter

The main difference with the filters, proposed in [24, 25] is related to the handling of the low-pass and high-pass filter coefficients. Furthermore, this filter requires to be feed by correctly scheduled data by a smart shift of data between different stages to serve P sampling in each clock cycle.

Respecting the z^{-p} buffer order, it is the most sensitive and the core of the serial FIR filter to parallel operation. This role of manage and interleaving data is devoted to the key block in our model. We called this block "buffers block" situated between the filters in different level. The structure of a single buffer is dependent on the direct or inverse DWPT.

In our following proposed architecture, we respect these constraints:

- The degree of parallelism and the characteristics of the target card (technology) must be taken into consideration in order not to deposit available resources.
- The degree of parallelism must respect the dyadic rule, that is, $P = 2^x, \forall x \in \mathbb{N}^+$. Where P affects directly the data managing between different levels of the transformation tree and simplifies the up/down sampling operation.
- To synthesis our proposed hardware architecture, we used

Altera Quartus premium lite edition software that is targeted on Altera FPGA belonging to the Cyclone V family with a speed grade of –7.

3.1. First proposed architecture: Parallel-Pipeline architecture of DWPT with sharing resources

This part is dedicated to present a P-Parallel DWPT architecture, which provides high throughput, by using the modified transposed P-parallel digital FIR filters (presented in figure 6).
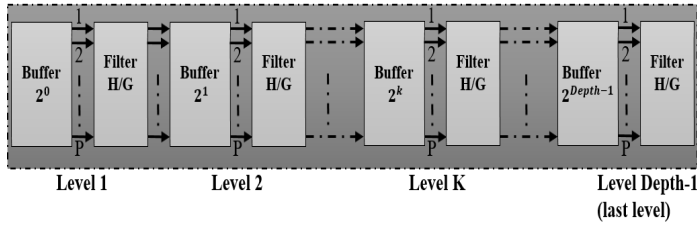
Figure 6. Datapath diagram of the proposed P-parallel DWPT architecture

As we mention below, this modified filter or H/G block, presented in Figure 5, ensures the processing of the same amount of data on any stage k in the original Mallat binary tree.

Concerning our plan to reduce the hardware usage, the functionality of high-pass filter and low-pass filter is provided by the H/G blocks. Consequently instead of using two separate similar filters (high-pass filter and low-pass filter), we propose using a single filter in an alternatively process. Hence, this alternatively process (on consecutive clock cycle) works by taking a sample for the $H(Z)$ and then for $G(Z)$ and so on. The critical point in this model is to manage correctly the data between filters and different levels, this role is devoted to a specific buffers situated between the filters in different stages. The structure of blocks buffer is shown in figure 7.

A. Buffer Block structure of DWPT

The k parameter in figure 7 is related to the number of the stage in which the buffer is implemented. Then in each level k in figure 6, the buffer blocks is built up on two types of shift register (2^{k-1} positions shift for each one): a "fast shift buffer" which takes data from the previous stage, and a "slow shift buffer" which feeds its data from its own stage filter. Overall, the buffer size depends on the parallel degree P and the level k on which it is implemented.

all related stage. On each 2^k cycles, only half samples can be transferred from the fast buffer part to the slow buffer part. Furthermore, each sample flows out from the slow buffer to the next stage must be presented twice (granted by the low frequency rate of the slow buffer) in order to be processed by both $H(z)$ and $G(z)$ filters. To manage different control signal and the data interleaved in different stage, we developed a control unit called "$Control\_block$".
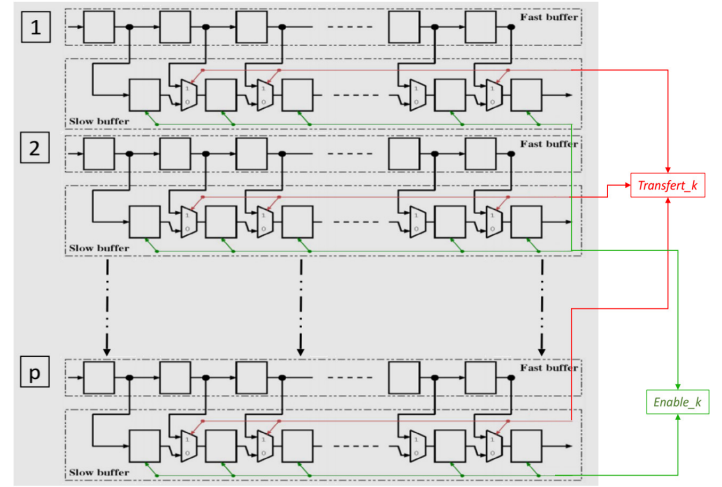
Figure 7. Structure of the DWPT buffer block in stage k

The "$Control\_block$" is dedicated to generate and manage the various controls signals "$enable_k$" and "$transfer_k$" related to stage K.
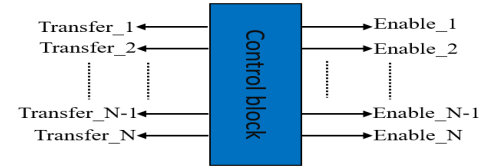
Figure 8. Control unit block

B. Synthesis results

As we mentioned, this architecture is fully configurable in synthesis by three parameters: the wavelet scale (the tree depth), the order of the filters and the filter coefficient quantization (generic parameters in the VHDL-RTL model). Moreover, it is partially configurable after synthesis of filters coefficient values corresponding to different wavelet family (with the same filter order). This means we can change the type of wavelet without re-synthesis our architecture (the coefficients are loaded dynamically after synthesis) where it was not the case for all the previous work.

In this new scheme, the fast buffer (and hence in the slow buffer too) are stored, by interleaving ordered disposition, filtered data of

In tables 1, 2 and 3, the configuration parameters are presented as 3-tuples: Depth of DWPT tree, Order of filter, and Quantization

with the synthesis results of implementation for 4, 8 and 16 parallel DWPT architecture. Although the highest rate of our architecture is not requiring any memory or DSP blocks.

Table 1. Implementation results of 4-parallel DWPT architecture

Design parameters (Depth, Order *Quantization*)	Data processing rate	Resources usage (l_e, l_r)
(2, 2, 5)	415.36	(471,296)
(3, 2, 5)	400.86	(756,510)
(4, 2, 5)	389.56	(1204,899)
(2, 4, 5)	403.52	(879,456)
(3, 4, 5)	342.46	(1299,719)
(4, 4, 5)	375.24	(1941,1171)
(2, 16, 5)	357.99	(3299,1416)
(3, 16, 5)	369.83	(4794,1924)
(4, 16, 5)	3425.69	(6397,2614)
(2, 2, 16)	208.14	(2571,905)
(3, 2,16)	198.62	(4216,1599)
(4, 2,16)	209.89	(5850,2853)
(2, 4,16)	201.18	(5038,1324)
(3, 4,16)	194.60	(7521,2260)
(4, 4,16)	184.120	(10374,3636)
(2, 16,16)	180.41	(4902,4402)
(3, 16,16)	220.31	(6805,5729)
(4, 16,16)	213.63	(9107,7752)

Table 2. Implementation results of 8-parallel DWPT architecture

Design parameters (Depth, Order *Quantization*)	Data processing rate	Resources usage (l_e, l_r)
(2, 2, 5)	583.03	(1109,504)
(3, 2, 5)	557.23	(1699,935)
(4, 2, 5)	562.25	(2754,1531)
(2, 4, 5)	586.20	(2120,897)
(3, 4, 5)	506.77	(3050,1197)
(4, 4, 5)	526.91	(4603,2023)
(2, 16, 5)	499.48	(7689,2447)
(3, 16, 5)	504.56	(12176,3166)
(4, 16, 5)	478.26	(14956,4571)
(2, 2, 16)	282.35	(6079,1696)
(3, 2,16)	294.41	(9279,2735)
(4, 2,16)	300.60	(13489,5011)
(2, 4,16)	289.86	(12032,2582)
(3, 4,16)	273.15	(17549,3965)
(4, 4,16)	238.45	(24311,6363)
(2, 16,16)	302.62	(11263,7856)
(3, 16,16)	325.85	(14750,11451)
(4, 16,16)	309.31	(21314,13091)

3.2. Second proposed architecture: Parallel-Pipeline architecture of IDWPT with sharing resources

A new P-parallel architecture of IDWPT will proposed in the following part, which can provide ultra-high speed data processing and low cost resources consummation.

Table 3. Implementation results of 16-parallel DWPT architecture

Design parameters (Depth, Order *Quantization*)	Data processing rate	Resources usage (l_e, l_r)
(2, 2, 5)	718.13	(3668, 652)
(3, 2, 5)	710.42	(6019, 960)
(4, 2, 5)	709.04	(8655, 1243)
(2, 4, 5)	536.55	(5991, 1689)
(3, 4, 5)	517.53	(8380, 2363)
(4, 4, 5)	516.33	(11181, 2601)
(2, 16, 5)	464.04	(30012, 4881)
(3, 16, 5)	454.15	(37172, 6575)
(4, 16, 5)	454.92	(38374, 7680)
(2, 2, 16)	455.49	(11116, 3395)
(3, 2,16)	447.30	(17679, 4330)
(4, 2,16)	446.25	(25389, 4830)
(2, 4,16)	336.75	(31336, 5137)
(3, 4,16)	331.99	(39361, 7764)
(4, 4,16)	341.59	(42687, 10342)
(2, 16,16)	302.46	(26408, 15572)
(3, 16,16)	303.44	(33859, 20959)
(4, 16,16)	298.79	(36348, 24456)

The reconstruction process of the P-parallel IDWPT architecture is simply the reversed form of P-parallel DWPT. Considering the diagram in Figure 4, we can observe that any filter of level k is able to process $P * 2^{(depth-k)}$ times the amount of data to be processed on level 1. Thus, the total amount of data by all filters to be processed in level k is the same as in level 1 $(P * 2^{(depth-k)} * 2^{-(depth-k)} = P)$. It is the same remark that is observed at the DWPT in the previous part (III.1 a). Moreover, we can notice that at any filter, in a given level, half the data is processed than its neighbors on the output side, and twice its neighbor on the input side. This implies that the construction of structure tree have repeated functionally blocks and the complexity of the filters is the same for all stages.

To eliminate the exponential evolution of filters number as function of depth order, we linearize and serialize the filtered data in different stages. Consequently, instead of using $P * 2^k/2$ low pass filters and $P * 2^k/2$ low pass filters to filter data in stage k, we just implement only one modified filter in each stage. As shown in figure 10, the number of used modified transposed FIR filter bank increases linearly a function of depth order.
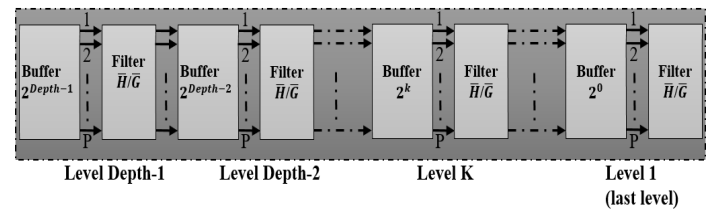
Figure 10. Data path diagram of the proposed P-parallel IDWPT architecture

The modified single filter ($\frac{\bar{H}}{\bar{G}}$ Filter Block) can process the same amount of data on stage k in the original tree (Figure 4). In figure 10, $\frac{\bar{H}}{\bar{G}} block$ is the high-pass filter and low-pass filter, which are

related to $h(n)$ and $g(n)$, in equation (16).

The structure for the modified FIR filter is the same as that presented in figure 5. The only difference is the coefficients of FIR filters concerning $\bar{H}(Z)$ and $\bar{G}(Z)$. This filter provides the same functionality of that presented by Mallat in their binary tree, we also reduce their occupied area by merging low-pass and high-pass reconstruction filters function in the structure in a single block. The processing of filtering is similar to that used in decomposition part (sub section III-1) where we take a sample for $\bar{H}(Z)$ filter and for $\bar{G}(Z)$ filter by an alternatively process (on consecutive cycles) and so on.

Consequently, this modified block can process one sample in two clock cycles. To ensure the best interleaving and managing of data in different stages, we developed a key block in our entire model that is the "Buffer Block", which is situated between the filters in different levels.

A. Buffer Block structure of IDWPT

The most important thing in our P-parallel IDWPT (Figure 10) is, as previously, the link between neighbor levels, which does not need any reorganize of data set. We use the same concept of buffer blocks as illustrated in Figure 6, so the size of the buffer shown in figure 11 depends on the stage in which it is implemented. The buffer block is then built up on two sub-blocks in each k positions (k is an indicator of the stage where the block buffer is implemented).

Then in each level k in figure 10, the buffer block is built up on two types of shift register (2^{k-1} positions shift for each one): a "fast shift buffer" takes data from the previous stage, and a "slow shift buffer" which feeds its data from its own stage filter. Overall, the buffer size depends on the parallel degree P and the stage k in which it is implemented.

The "fast shift buffer" is faster than "slow shift buffer" where it achieves P-shift on each clock cycle while the "slow shift buffer" achieves P-shift on two-clock cycles. To handle this latency, we used an enable signal called "$enable_k$" which manages the slower shift rate in level k. To control the transfer data from the fast buffer to slow buffer, we insert a counter signal 2^k to activate the $transfer_k$ signal on every 2^k cycles. This transfer operation is performed P times at each stage.

The synthesis results of implementation of our P-parallel IDWPT architecture based on modified P-parallel FIR filter structure will be the subject of this part.

This architecture of P-parallel IDWPT is fully configurable in synthesis by three parameters: the wavelet scale (the tree depth), the order of the filters and the filter coefficient quantization (generic parameters in the VHDL-RTL model). Moreover, it is partially configurable after synthesis of filters coefficients values corresponding to different wavelet families (which have the same filter order). The resources consumed by our proposed architecture resources are given as 2-tuples (l_e, l_r): where l_e stands for logic elements and l_r for logic registers. To generate the differences counter, the different control signals "$enable_k$" and "$transfer_k$" related to stage k, we proposed also like in figure 8 a block control unit.
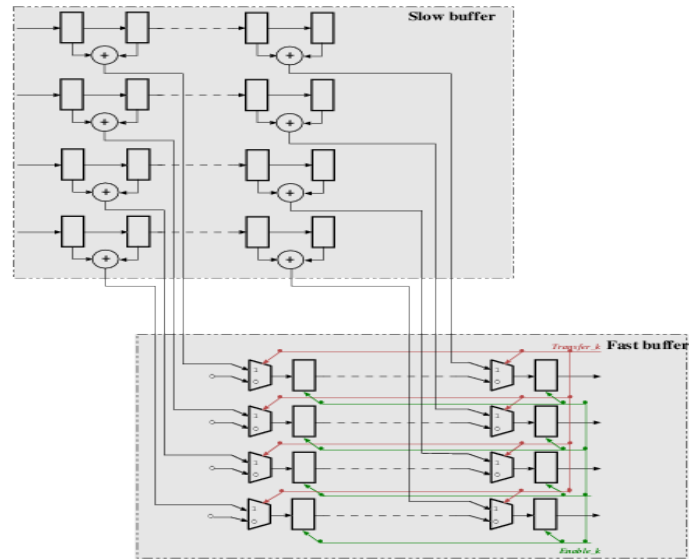
Figure 11. Structure of the IDWPT buffer block in stage k with degree of parallelization P=4

Table 4. Implementation results of 4-parallel IDWPT architecture

Design parameters (Depth, Order Quantization)	Clock frequency	Resources usage (l_e, l_r)
(2, 2, 5)	205	(109, 186)
(3, 2, 5)	201.82	(166,312)
(4, 2, 5)	196.16	(244,505)
(2, 4, 5)	152.88	(265,286)
(3, 4, 5)	152.58	(379,442)
(4, 4, 5)	153.37	(483,665)
(2, 16, 5)	144.03	(1447,886)
(3, 16, 5)	137.44	(1983,1222)
(4, 16, 5)	136.24	(2457,1625)
(2, 2, 16)	132.36	(578,582)
(3, 2,16)	135.34	(833,972)
(4, 2,16)	133.69	(1102,1572)
(2, 4,16)	104.57	(1594,902)
(3, 4,16)	102.77	(2174,1388)
(4, 4,16)	100.61	(2772,2084)
(2, 16,16)	94.14	(7719,2822)
(3, 16,16)	92.08	(10557,3884)
(4, 16,16)	90.49	(13469,5156)

B.   Synthesis results

In tables 4, 5 and 6, the configuration parameters are presented as 3-tuples: Depth of DWPT tree, Order of filter, and Quantization with the synthesis results of implementation for 4, 8 and 16 parallel DWPT architecture. Although the highest rate of our architecture does not require memory or DSP blocks.

3.1. Third proposed architecture: Parallel-Pipeline architecture of IDWPT-DWPT with sharing resources and shared FIR modified filter

This part is dedicated to present another version of parallel DWPT/IDWPT architecture. In some applications and especially in digital communication, we find that there is a need to implement

the DWPT and IDWPT transforms on the same FPGA split. Certainly, we can implement our architecture (P-parallel DWPT and P-parallel IDWPT) on the same board to ensure the transformation functionality using the same constraints of work with ultra-high throughput and low resources consumption (consequently low power consumption). We notice a big regularity in the architecture of P-parallel DWPT (Figure 6) and P-parallel IDWPT (Figure 10), particularly in the modified FIR filters on the different levels where the structure of modified FIR filter is the same and the difference is particularly in the loaded coefficients after synthesis. We think to develop, at last but not at least, new architecture, which retains all features from previous architectures (ultra-high throughput, generic, configurable with depth and with order of filters and with quantization). In addition, the same architecture can be implemented with P-parallel DWPT or P-parallel DWPT by the selection of the active way as shown in figure 12.

In this scheme, we represent a new pipeline-parallel architecture of DWPT/IDWPT with shared P-parallel FIR filter. Furthermore, instead of the implementation of 2*Depth P-parallel FIR filter in this architecture, we used the half of hardware architecture with the conservation of the ultra-throughput rate.

The buffers block are similarly presented in previous part in P-parallel DWPT and P-parallel IDWPT. Respectively, Mux_Data and DeMux_Data blocks are implemented to manage data between the buffers and the shared P-parallel FIR filter in P-parallel DWPT or P-parallel IDWPT. The Mux_Filter blocks are multiplexer, used to precise the transform direction and to manage the loading FIR filters coefficients after synthesis.

A. Synthesis results

To evaluate the performance of our P parallel DWPT/IDWPT architecture, we just implement the Blocks of P-parallel FIR filters, which gives us a percentage of the resources quantity consumption that have been dispensed by using a shared P-parallel FIR filters.

The compilation results of area consumption for different values of the configuration parameters, that mean tree depth, filter order, coefficient quantization (in number of bits), and parallel degree are presented in table 7 with different parallel degree.

Table 5. Implementation results of 8-parallel IDWPT architecture

Design parameters (Depth, Order Quantization)	Clock frequency	Resources usage (l_e, l_r)
(2, 2, 5)	207.04	(202, 347)
(3, 2, 5)	195.77	(301,574)
(4, 2, 5)	198.73	(409,863)
(2, 4, 5)	147.15	(527,527)
(3, 4, 5)	148.39	(705,804)
(4, 4, 5)	147.65	(813,1123)
(2, 16, 5)	136.44	(2955,1607)
(3, 16, 5)	133.05	(3802,2184)
(4, 16, 5)	131.56	(4288,2683)
(2, 2, 16)	128.75	(1140,1095)
(3, 2,16)	123.08	(1529,1806)
(4, 2,16)	128.04	(1828,2711)
(2, 4,16)	99.98	(3098,1671)
(3, 4,16)	98.87	(4094,2542)
(4, 4,16)	98.95	(4733,3543)
(2, 16,16)	90.16	(15416,5127)
(3, 16,16)	86.02	(20279,6958)
(4, 16,16)	86.1	(23194,8535)

Table 6. Implementation results of 16-parallel IDWPT architecture

Design parameters (Depth, Order *Quantization*)	Clock frequency	Resources usage (l_e, l_r)
(2, 2, 5)	197.47	(389,668)
(3, 2, 5)	195.35	(573,1096)
(4, 2, 5)	194.97	(742,1576)
(2, 4, 5)	147.54	(1091,1008)
(3, 4, 5)	142.31	(1410,1526)
(4, 4, 5)	141.98	(1552,2036)
(2, 16, 5)	127.6	(5894,3048)
(3, 16, 5)	124.88	(7300,4106)
(4, 16, 5)	125.09	(7536,4796)
(2, 2, 16)	125.25	(2183,2120)
(3, 2,16)	123.02	(2704,3472)
(4, 2,16)	122.71	(3016,4986)
(2, 4,16)	92.6	(6154,3208)
(3, 4,16)	91.29	(7730,4848)
(4, 4,16)	93.93	(8383,6458)
(2, 16,16)	83.17	(30619,9736)
(3, 16,16)	83.44	(39257,13104)
(4, 16,16)	82.16	(42143,15290)

Table 7. Implementation results of 4, 8 and 16-parallel degree of Parallel FIR architecture

Design parameters (Depth, Order *Quantization*)	FIR Resources usage (l_e, l_r)		
	P=4	P=8	P=16
(2, 2, 5)	(41,27)	(128,64)	(298,106)
(3, 2, 5)	(41,28)	(138,67)	(315,110)
(4, 2, 5)	(49,30)	(138,70)	(315,113)
(2, 4, 5)	(53,37)	(148,83)	(371,127)
(3, 4, 5)	(53,38)	(167,86)	(371,131)
(4, 4, 5)	(61,40)	(167,89)	(371,135)
(2, 16, 5)	(112,94)	(281,189)	(701,319)
(3, 16, 5)	(112,96)	(298,195)	(701,322)
(4, 16, 5)	(119,98)	(307,195)	(701,327)
(2, 2, 16)	(198,83)	(302,107)	(368,161)
(3, 2,16)	(198,88)	(328,112)	(368,166)
(4, 2,16)	(218,93)	(353,117)	(368,172)
(2, 4,16)	(257,114)	(378,138)	(431,337)
(3, 4,16)	(257,119)	(428,143)	(431,198)
(4, 4,16)	(297,124)	(428,148)	(431,203)
(2, 16,16)	(557,278)	(760,314)	(1164,408)
(3, 16,16)	(557,283)	(806,319)	(1164,408)
(4, 16,16)	(593,288)	(806,324)	(1164,413)

From table 7, we notice that the resources consumption are the same with the same order filter (which is expected). Moreover, with half usage of P-parallel FIR filter in P-parallel DWPT/IDWPT, we gain resources percentage from 3 to 5 % for a depth = 4 and filter order = 16, which is significant especially with the large depth and high filter order (like discrete Meyer wavelet).

This is a very important result because classically it is too difficult to implement in the same board simultaneously both parallel architectures.

4. Comparison

To evaluate the performance of each architecture, we make a comparison chart (Table 8) with several DWPT and IDWPT architectures designs achieved in literature.

A quantitative and comprehensive comparative are presented in Table 8, where we can observe that:

- Our architectures are implemented without using any memory or DSP block where most of other architecture required that.
- Our models DWPT/IDWPT architectures are generic and fully configurable. They are suitable for all wavelet families (where the wavelet families are simplified by the filters orders) where it is not the case for all previous works.
- In our architectures; the quantization, the depth order and the filters order are theoretically unlimited (depends on the limit of manufacturing technology)

From these observations, table 8 presents the superiority of our architectures that are clearly presented on the high bite-rate driven by high frequency (around 200MHz).

5. Conclusion

In this work, we have proposed three powerful P-parallel/pipeline configurable architecture of Direct Discrete Wavelet Packet Transform (DWPT) and Inverse Discrete Wavelet Packet Transform (IDWPT) based on Mallat binary tree using bank filter concept, which provide high throughput with minimal hardware resources. The considered problem is to accelerate data processing in clock cycle and decrease the total hardware resources used.

To solve these problems, we develop a P-Parallel modified FIR filter based on transposed FIR filters that can share hardware resources between low-pass and high-pass filters (by merging their functionality). The effective data path maintains a short critical path allowing high operation frequency to be reached, where synthesis results indicate that our architecture provides a very high speed data processing with minimum resources. For example: for a parallel degree P=16, depth order = 2, filter order = 2 and order quantization = 5 we reach a very high bit rate equal to 3159.52 Mega samples. We also developed a new effective architecture for both DWPT and IDWPT implemented in the same programmable board, conserving the high throughput performance with more decreased hardware resources consumption.

Those proposed architectures (IDWPT and DWPT) are fully configurable at synthesis as function of P-parallel degree, depth (number of tree stages), filter order and filter coefficient quantization (generic parameters in the VHDL-RTL model). Furthermore, they are re-configurable after synthesis by loading the filter coefficients, which depend on different wavelet family during operations (after synthesis) that providing high flexibility of DWPT/IDWPT transforms.

Perspectives

In this work, we present a different parallel version of hardware implementation in FPGA of DWPT/IDWPT. This work is still in progress where we are constructing another generation of DWPT/IDWPT implementation, which can provide more performance and can be used in new application domains.
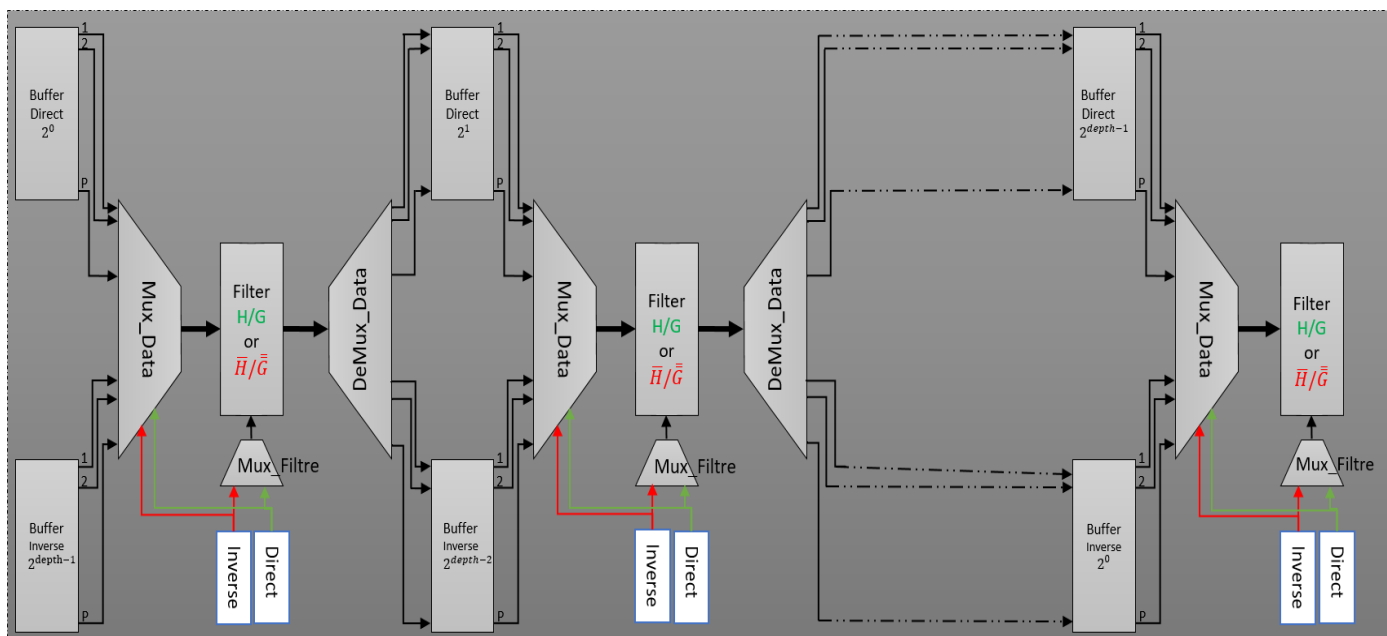
Figure 12. Proposed data-path diagram of both P-parallel DWPT and IDWPT architecture.

Table 8. Comparison of proposed architectures with existing DWPT and IDWPT architectures.

	Tze-Yun et al.[28]	Marino et al. [29]	Mohanty et al. [30]	Madishetty et al. [23]	Wang et al. [27]	Wu et al. [12]	Meihua et al. [21]	Proposed architecture N°1	Proposed architecture N°2	Proposed architecture N°3
Wavelet	Daub-4	Quadri-filter	Daub-4	Daub-6	Lifting-Db4	Quadri-filter	Quadri-filter	arbitrary	arbitrary	arbitrary
DWPT/IDWPT	Both	DWPT	DWPT	DWPT	Both	Both	DWPT	DWPT	IDWPT	Both
Logic cell	N/A	N/A	426	1040	N/A	30192 (logic gate)	1835 (logic unit)	3668 (logic element)	389 (logic element)	
Technology	Xilinx XC2V4000	N/A	CMOS 90nm	Xilinx Virtex 6	CMOS 180 nm	0.35 µm	Altera EP20K200E	Altera Cyclone IV	Altera Cyclone V	Altera Cyclone V
Max. Freq.(Mhz)/Bitrate	N/A	N/A	20	306.15	20	100	29	Bitrate:718.13	Feq.: 197.47	~200
Memory	Yes	N/A	yes	yes	Yes	N/A	yes	No	No	No
Quantization (bits)	N/A	N/A	N/A	N/A		32	N/A	Test with 5 to 16 (up to unlimited theatrically)	Test with 5 to 16 (up to unlimited theatrically)	Test with 5 to 16 (up to unlimited theatrically)
DSP	N/A	N/A	N/A			N/A	N/A	No	No	No
Depth	3	N/A	N/A	4	3	3 (up to 6)	3	2, 3 and 4 (up to unlimited theatrically)	2, 3 and 4 (up to unlimited theatrically)	2, 3 and 4 (up to unlimited theatrically)
Parallel degree								4,8 and 16 (up to unlimited theatrically)	4,8 and 16 (up to unlimited theatrically)	4,8 and 16 (up to unlimited theatrically)

References

[1] S. Mallat, A wavelet tour of signal processing, Academic Press, 1999.

[2] R. Dilmaghani and M. Ghavami, "Comparison between wavelet-based and Fourier-based multicarrier UWB systems," IET Communications, Vol. 2, Issue 2, Feb. 2008. http://dx.doi.org/10.1049/iet-com:20070181

[3] A. Grossmann and J. Morlet, "Decomposition of Hardy functions into square integrable wavelets of constant shape," SIAM journal on mathematical analysis, 15(4), 723-736, 1984. https://doi.org/10.1137/0515056

[4] S. Mallat, "Multiresolution approximations and wavelet orthonormal bases of L2(R)," Transactions of the American Mathematical Society, vol. 315, no. 1, pp. 69–87, 1989. https://doi.org/10.1090/S0002-9947-1989-1008470-5

[5] Y. Meyer, "Wavelets and operators," vol. 37 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1992.

[6] C.K. Chui, "An Introduction to Wavelets," Philadelphia, SIAM, vol. 38, 1992.

[7] I. Daubechies, Ten lectures on wavelets, Philadelphia, SIAM, vol. 61, pp. 198–202, 1992.

[8] T.C. Denk and K.K. Parhi, "Architectures for lattice structure based orthonormal discrete wavelet transforms," IEEE International Conference on Application Specific Array Processors, pp. 259-270, Aug. 1994. https://doi.org/10.1109/ASAP.1994.331798

[9] H. H. A. Hatem, M. El-Matbouly, N. Hamdy, K. A. Shehata, "VLSI architecture of QMF for DWT integrated system," Circuits and Systems, MWSCAS 2001. Proceedings of the 44th IEEE 2001 Midwest Symposium on (Volume: 2), pp. 560–563, 2001. https://doi.org/10.1109/MWSCAS.2001.986253

[10] G. Strang and T. Nguyen, Wavelets and Filter Banks, Wellesley-Cambridge Press, 1997.

[11] W. Xiaodong, L. Yongming and C. Hongyi, "Programmable wavelet packet transform processor," In Electronics Letters, vol. 35, no. 6, pp. 449-450, 18 Mar 1999. https://doi.org/10.1049/el:19990330

[12] B-F. Wu and Y-Q. Hu, "An efficient VLSI implementation of the discrete wavelet transform using embedded instruction codes for symmetric filters," IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 9, pp. 936–943, September 2003. https://doi.org/10.1109/TCSVT.2003.816509

[13] G. Paya, M. M. Peiro, F. J. Ballester, V. Herrero and J. Cerda, "A new inverse architecture discrete wavelet packet transform architecture," IEEE, Signal Processing and Its Applications, 7803-7946, 443 - 446 vol.2, 2003. https://doi.org/10.1109/ISSPA.2003.1224909

[14] A. Jamin and P. Mähönen, "FPGA implementation of the wavelet packet transform for high speed communications," International Conference on Field Programmable Logic and Applications. Springer Berlin Heidelberg, 2002. https://doi.org/10.1007/3-540-46117-5_23

[15] M. Trenas, J. López, M. Sanchez, F. Argüello and E. L. Zapata, "Architecture for Wavelet Pack et Transform with Best Tree Searching," In Application-Specific Systems, Architectures, and Processors, 2000. Proceedings. IEEE International Conference on (pp. 289-298). IEEE, 2000. https://doi.org/10.1109/ASAP.2000.862399

[16] M. Trenas, J. López and E. Zapata, "FPGA Implementation of Wavelet Packet transform with Reconfigurable Tree Structure," Euro micro Conference, 2000. Proceedings of the 26th Volume 1, 5-7, pp. 244 - 251 vol.1, September 2000. https://doi.org/10.1109/EURMIC.2000.874639

[17] M. Trenas, J. López, E. Zapata and F. Argüello, "A Configurable Architecture for the Wavelet Packet Transform," Journal of VLSI signal processing systems for signal, image and video technology, Volume 32, Issue 3, pp 255-273, November 2002. https://doi.org/10.1023/A:1020221003822

[18] M. Farahani and M. Eshghi, "Architecture of a wavelet packet transform using parallel filters," TENCON 2006 - IEEE Region 10 Conference, p. 7-10, 2006. https://doi.org/10.1109/TENCON.2006.343965

[19] M. Farahani and M. Eshghi, "Implementing a new architecture of wavelet packet transform on FPGA," Proceedings of the 8th WSEAS International Conference on Acoustics & Music: Theory & Applications, Vancouver, Canada. 2007.

[20] M.A. Farahani, S. Mirzaei and H.A. Farahani, "Implementation of a reconfigurable architecture of discrete wavelet packet transform with three types of multipliers on FPGA," 24th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), 2011. https://doi.org/10.1109/CCECE.2011.6030704

[21] X. Mei-hua, C. Zhang-jin, R. Feng, and C. Yu-lan, "Architecture research and VLSI implementation for discrete wavelet packet transform," High Density Microsystem Design and Packaging and Component Failure Analysis, 2006. HDP'06. Conference on. IEEE, 2006. https://doi.org/10.1109/HDP.2006.1707554

[22] D. Garcia, M. Mansour and M. Ali, "A flexible hardware architecture for wavelet packet transform with arbitrary tree structure," IEEE Transactions on circuits and systems – II: Express briefs, vol. 60, no. 10, p. 657-661, October 2013.   https://doi.org/10.1109/TCSII.2013.2277964

[23] S. K. Madishetty, A. Madanayake, R. J. Cintra and V. S. Dimitrov, "Precise VLSI Architecture for AI Based 1-D/ 2-D Daub-6 Wavelet Filter Banks with Low Adder-Count," IEEE Transactions on circuits and systems-I: regular paper, Vol. 61, No. 7, 1984 - 1993, July2014. https://doi.org/10.1109/TCSI.2014.2298283

[24] M. Chehaitly, M. Tabaa, F. Monteiro and A. Dandache, "A fast and configurable architecture for Discrete Wavelet Packet Transform," In Design of Circuits and Integrated Systems (DCIS), 2015 Conference on (pp. 1-6), 2015. https://doi.org/10.1109/DCIS.2015.7388599

[25] M. Chehaitly, M. Tabaa, F. Monteiro and A. Dandache, "A VHDL-RTL implementation for a fast and configurable design of inverse discrete wavelet packet transform," Design of Circuits and Integrated Systems (DCIS), 2016 Conference on, p. 1 – 5, 2016.

[26] M. Antonini, M. Barlaud, P. Mathieu and I. Daubechies, "Image coding using wavelet transform," IEEE Transactions on image processing, 1(2), 205-220, 1992.

[27] C. Wang, J. Zhou,L. Liao, J. Lan, J. Luo, X. Liu and M. Je, "Near-threshold energy-and area-efficient reconfigurable DWPT/DWT processor for healthcare - monitoring Applications," IEEE Transactions on Circuits and Systems II: Express Briefs, 62(1), 70-74, 2015. https://doi.org/10.1109/TCSII.2014.2362791

[28] T.-Y. Sung, H.-C. Hsin, Y.-S. Shieh and C.-W. Yu, "Low-power multiplierless 2-D DWT and IDWT architectures using 4-tap Daubechies filters," In Proc. Seventh Int. Conf. PDCAT, pp. 185–190, 2006. https://doi.org/10.1109/PDCAT.2006.78

[29] F. Marino, "Two fast architectures for the direct 2-D discrete wavelet transform," IEEE Trans. Signal Process, vol. 49, no. 6, pp. 1248–1259, 2001. https://doi.org/10.1109/78.923307

[30] B. K. Mohanty and P. K. Meher, "Memory-efficient high-speed convolution-based generic structure for multilevel 2-D DWT," IEEE Trans. Circuits Syst. Video Technol., vol. 23, pp. 353–363, 2013. https://doi.org/10.1109/TCSVT.2012.2203745

ASTES

Developing A Conceptual Framework of Product-Service System Management Toward Firms' Sustainability for Indonesian Industrial Estate Firms

Christina Wirawan[*,1,2], Gatot Yudoko[1], Yuliani Dwi Lestari[1]

[1]School of Business Department, Institut Teknologi Bandung, Bandung 40123, Indonesia

[2]Industrial Engineering Department, Universitas Kristen Maranatha, Bandung 40164, Indonesia

A R T I C L E   I N F O

A B S T R A C T

Industrial estate firms have important roles in national development through the industrial development. In the course of their life, industrial estate firms ought to follow the firm life cycle pattern. The common pattern shows that the period, in which the firm experiences profit increase, will be directly followed by consistent profit and profit decrease. Should the strategy realization are not precisely implemented during each stage, especially the decreases period, it will be difficult for the firm to survive.

A strategy that can be implemented by industrial estate firms in order to be sustained is to manage Product-Service System (PSS). It is a well-known concept in creating opportunities for a new business model in which firms can offer bundles of products and services to their customers instead of merely tangible products. PSS can be distinguished into three types of categorization consist of different compositions of products and services and different business activities. Each type of PSS and business activities can be applied at the stages of the firm life cycle in accordance with its condition to maintain the company's sustainability at every stage of its life cycle. In this article, we construct a conceptual framework to help Indonesian industrial estate firm to be alert on the need to focuses on providing services instead of only selling products. This article also shows that different PSS management would be needed in particular stages of firm life cycle. Applying suitable PSS and business activities based on PSS in each stage of the firm life cycle can lead to gaining a competitive advantage and sustainable industrial estate firms. This article's novelty is to combine the PSS concept concerning each stage of firm life cycle at industrial estate firms.

1. Introduction

This article is an extension of our previous work published at 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM) [1]. Our previous work discussed about types of PSS and business activities implementation in Indonesian industrial estate firms in order to help them sustain [1]. Industrial estates firstly initiated in Europe, such as England; Netherland; Germany; and the United States of America in the late 1800s [2]. They are built as a center and localization for industrial companies [2]. Following after, many countries also built industrial estates for several reasons such as environmental and economic growth [3]. Industrial estate firms have obligation to provide land; building; infrastructures; utility

services [4]; and amenities [5] for industrial companies in the estate and community. Industrial estates, primarily built to support industrial growth [2] by spurring industrial growth [6] and efficient resources usage [5] while also aimed to protect the environment [3]. Based on their purposes, industrial estate firms becoming important for a nation and need to be sustained. However, along their lifespan, industrial estate firm will then follow the firm life cycle pattern [7].

Following the common firm life cycle, industrial estate firms will experience profit ups and downs [8]. At the beginning of the firm life cycle, industrial estate firms have land and building to be sold, thus they will be able to make enough profit. However, along the time period, if the business runs well, these land and buildings will eventually be sold out. Thus, in the later stages, industrial estate firms need to find other sources of profit to keep them

[*]Christina Wirawan, Ganesha 10, Bandung, Indonesia, +62 8122119907
Email: christina.wirawan@sbm-itb.ac.id

sustained. Many industrial estate firms have tried to sustain by expanding the industrial estate of building another one in a different location. But this strategy seems to be unsuitable for long-term business since the land will eventually run out and the industrial estate market will also be saturated. Moreover, the local government will also impose several regulations that will limit the building of industrial estate firms. Then, to help industrial estate firms to be sustained, the concept of PSS is seemed to be suitable.

PSS concept attempt to shift tangible products consumption into services with the same performance, function, and benefit [9]. By applying the concept of PSS, the production and consumption of tangible products can be reduced so as to reduce the environmental burden, providing opportunities for new business models [9,10], and improving the social conditions of the community through co-creation [11]. According to experts, PSS is divided into three types and for each type of there are several types of business activities. In its implementation, the company can choose the type of PSS and business activities. Each type and business activities have their own characteristic that will build different strategies. These strategies would be suitable with the condition of the industrial estate firms at each stage of the life cycle. The appropriate business activity at the beginning of the life cycle stage may be different from the appropriate business activity at the later life cycle stages. The concept of PSS that combines the tangible products with services is considered to be suitable for the firms to achieve sustainability in each stage of its firm life cycle. It can be achieved by combining the limited tangible products with services.

In Indonesia, the first industrial estate was built in Cilacap in 1968 and the number was continuously increasing each year. The rapidly increasing number of industrial estate firms is partly due to the fact that since 1989 the private sector has been allowed to build industrial estate firms in Indonesia. Until now there are about 232 industrial estates in Indonesia [12]. About 68.24% industrial estates were located in Java, the most populous and developed island in Indonesia, where Jakarta as Indonesian capital is also located [12]. Therefore the Indonesian government encourages further industrial estates development outside of Java [13]. This regulation is applied to increase the distribution of industrial, economic and welfare development, especially outside of Java [12].

In Indonesia, the importance of industrial estate roles is strengthened by the Constitution of Republic of Indonesia number 3 the year 2014 (*Undang-Undang Republik Indonesia*) on the industry, which states the development of the industry as one of the pillars of national economic development [14]. Indonesian industrial estates were built to support national industries. According to Indonesian government's regulations, industrial estates in Indonesia were built for several purposes, such as to accelerate the spread and equality of industrial development; to develop environmentally friendly industrial development; to encourage investment and industrial competitiveness; and to give certainty to the site [15]. With these regulations, it is obvious that industrial estate firms play important roles to support economic development [14] through raising companies efficiency; attract investors [16]; create job opportunities; foster innovation [3]; and support welfare enhancement in surrounding area. Industrial estates also help the government to control environmental impact and industrial development [15]. Considering the importance of

the roles of industrial estate companies, it is necessary to think applicable efforts to support their sustainability.

The Indonesian government has already been aware of the important role of industrial estate firms by encouraging industrial estate firms' performance and the effort for sustainability. One of the efforts done by Indonesian government is by setting government regulation in 2009 and 2015 that remark sustainability in term of economic [14-15,17] and competitiveness; social [17]; community [6]; and environmental [15,17] which are deemed suitable with sustainable development concept [18]. As an effort to improve the performance and sustainability of industrial estates, Indonesian government through The Indonesian Industrial Ministry holding Indonesian Industrial Estate Award(s), which has been held at 2013 and 2015 [17].

Indonesian industrial estate firm will also face a life cycle pattern problem, which in a certain period they will experience a decrease in profit due to exhaustion or reduction in product sales. But given its important role, they need to be sustainable. As discussed previously, the PSS concept would be necessary to help industrial estate firms to be sustained.

There have been many studies and discussions about PSS related to the life cycle carried out by researchers. Most researchers write about PSS to prolong product and PSS life cycle and further to recycle or rebuilt them.PSS concept discussed managing PSS product life cycle stages: design, manufacturing, usage, remanufacturing or disposal [19-20] toward sustainability. While for of PSS life cycle, PSS management discussed for the beginning of life, middle of life and end of life [20-21]. Beginning of life stage consist of ideation, requirement, design, middle of life consist of realization, delivery, support, and end of life consist of evolution [20, 22].

However, the article that discussed PSS management to address the firm life cycle has not been found. Although commonly the pattern of product, PSS, and firm life cycle may be similar, the strategy to manage the PSS would be different due to the unique characteristic of each industrial estate firms. The article discusses the implementation of PSS at industrial estate firms has also could not be found. Therefore, this article will discuss PSS implementation at industrial estate firms.

As pointed earlier, this article is an extent of our previous work. In the previous work, we found that Indonesian industrial estate began to lead to the provision of services, in addition to merely selling products [1]. They already practice some of the business activities as an effort to gain new profit toward sustainability [1]. This article will extend the finding by considering the firm life cycle. Relation of the type of PSS and business activities practiced with the stages of the firm life cycle would be studied. This study is conducted due to the fact that for industrial estate firm, there are different characteristics in each stage of the firm life cycle that will logically differ on what are the suitable type of PSS that need to be implemented.

The objective of this article is to build a conceptual framework regarding sustainable issues of industrial estate firms in each stage of firm life cycle using the PSS concept. This article will study each PSS types and business activities implementation at 5 industrial estates firms to understand PSS implementation in

Indonesian industrial estate firms while testing the theoretical based framework. This article will contribute to help industrial estate firm by raising awareness and giving suggestion to the industrial estate firms so that they will be more focus on providing services as the alternative income generator. This article also highlights the differences of PSS concept that is needed among stages of firm life cycle to help industrial estate firms sustain in each stage. For the government, this article will give insight for industrial estate planning and policy-making to encourage sustainability of the industrial estate firms to suitably play important roles. For academic field, this article will contribute to combine PSS concept with life cycle concept in an effort to help industrial estate firms to sustain.

In the following section, the literature review will be discussed, especially about the concepts that will be used as the basis for the formation of the framework and the previous studies followed by the research objective and methodology. Furthermore, the conceptual framework of PSS management will be presented at each stage of product life cycle followed by a discussion on PSS management framework and practice in Indonesian industrial area. In the end, this article will be closed by the conclusion, including implications and further research directions.

2. Literature Review

As previously mentioned, in most countries, industrial estates managed by industrial estate firms, have been built to help companies to operate more efficiently due to they are providing facilities, amenities, and waste treatment infrastructure to support companies' in their operation [4,5]. Industrial estates are built with the objectives to encourage employment rate and economic growth [6,23], to attract new firms [5] and foreign direct investments [6], to foster technological learning and innovation [3], to spur an industrial dynamic [6], to efficient resource used, and to lower environmental burden [23]. With this supports, companies may benefit from the reduced cost of getting water, energy, material, etc., as well as disposal cost [24]. Another benefit that will be gained by the company is an opportunity to build relationships and make collaborations with other companies and stakeholders to build value co-creation [25].
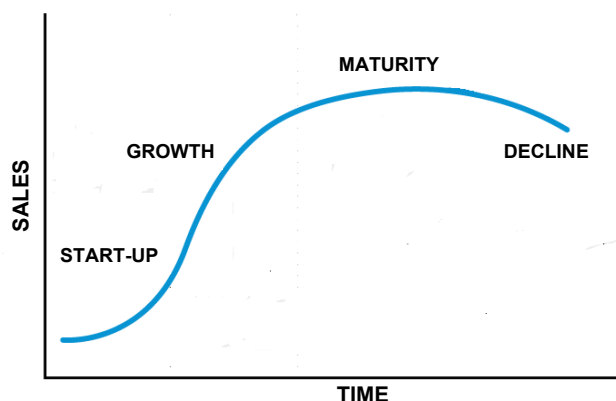
Figure 1. Common life cycle model[8], [26]

In the course of their growth and development, firms will experience the stages that form the general pattern, namely the firm life cycle [7]. Several authors identified divisions of firm life cycle stages. However, common stages are divided into start-up, growth,

maturity and decline stages [26] with the S-shaped pattern [27]. The common pattern of the life cycle is as shown in Figure 1 [8,27]. At each stage of its life cycle, the company will experience different conditions of different pressures, threats, and opportunities [28]. Therefore, the company will need a different strategy and performance [29], structure [30], effectiveness criteria [31], profit-maximizing [27] to keep sustaining in each stage of the life cycle.

In Indonesia, industrial estate firms have an important role in national economic development. This is indicated by the Republic of Indonesian government regulation requiring industrial companies to operate within industrial zones administered by licensed industrial estate companies, except in limited circumstances [15]. While in Indonesian republican law, the industry is declared as one of the pillars of national economic development [14]. Thus, it is clear that the industrial estate firms have an important role to play in national economic development. Indonesian Ministry of Industrial and Trade at 1987 already stated that industrial estate firms have obligations to provide road network in industrial estate, drainage, installation of clean water supply along with distribution channels to each industrial plot, wastewater treatment plant which can accommodate all wastewater from factories in industrial estates, electricity supply installations and distribution networks, telecommunications networks, environmental impact control facilities, lighting, industrial estates corporate offices, fire-fighters, housing for industrial workers, social and public facilities in accordance with the relevant provisions of the relevant agencies [12]. The Indonesian government already paid attention to the importance of industrial estate firms, as well as efforts to support the Indonesian industrial estate firms' sustainability by issuing some regulations, such as that remark sustainability in term of economic [14–16] and competitiveness [17]; social [17]; community [6]; and environmental [15,17] which are deemed suitable with sustainable development concept [18], as well as formulating a roadmap for industrial estates and organizing industrial estate awards for high performer industrial estates [17].

The first industrial estate in Indonesia was built in Cilacap in 1968, and since then some new industrial estates are built each year and the number increases from year to year. Until now, prospects for industrial estates is still considered good since demand is rising [12]. Geographically, the industrial area in Indonesia is still not evenly distributed. Most of them or as much as 68.24% have been built on Java, while in Riau Island, Riau and Bintan is 17,17% [12]. In Sumatra, Kalimantan, Sulawesi, Ambon and Papua there are 7,73%, 3.43%, 2.58%, 0.43% and 0.43% respectively [12]. Java is attractive to industrial estate firms due to the density of the population thus able to provide advantages of localization and urbanization economy[32]. Moreover, the complete facilities and infrastructure in Java and the presence of the nation's capital making it more interesting for further gain on profit. With this condition, the Industrial Master Plan or RIPIN 2015-2035 prioritizes the development of industrial areas outside Java to encourage welfare distribution throughout Indonesia.

Indonesian industrial estate firm main business is selling products in the form of land and buildings that make most of the profits. In addition, industrial estate firms also provide a range of services to support and to assist companies in the industrial estates

such as renting warehouses, land, buildings, industrial vehicles and other industrial equipment; providing water supply, waste treatment, and disposal; shipping and delivery facilities, etc. [1]. Product in the form of land and building usually give enough profit for industrial estate firms, but it is still limited. Once the land and buildings are sold out, the profits will only come from the provision of services, which is, of course, lower than the sales of the products. Some industrial estate firms maintain their sustainability by expanding the land or opening industrial estates elsewhere. However, this strategy cannot be done continuously since in the near future, the land will run out. This condition is also barred by the limitation by the rules of fringe benefit that some allocated area is not supposed to be used for industries [33]. Therefore, a new business model that emphasizes the provision of services will be appropriate to be applied to industrial park companies in order to remain profitable enough to be sustained.

Related to the firms' life cycle concept, in the early stages of its life cycle, industrial estate firms will gain enough profit from product sales, but in the later stages where the product is sold out, the industrial estate company needs to find a new business model to earn enough profit to be sustained. Regarding the importance of the role of industrial estate companies in Indonesia in contributing to the development of the national economy, an effort to maintain their sustainability is highly needed. The concept of PSS that combines product sales and service provision will be suitable for application in industrial estate firms, especially in the later stage of the firm life cycle.

PSS is a system of products; services; network; and infrastructure designed to meet customer needs and cause lower environmental impact than traditional business models while seek to increase firm competitiveness [34]. PSS promote dematerialization [9,35]. It is the attempt to decrease tangible products consumption, and wherever possible change it into the bundle of products and services [9]. PSS also attempt to tangible products' useful life and increase the utilization of tangible products [10] that both will reduce tangible product usage. With PSS concept, consumer or user will receive relatively equal function, performance, and benefit as they buy tangible products in a different way. PSS concept attempt to decreased environmental burden and meeting the needs of the world's population that increased dramatically caused by the sharp increase in population and consumption of large quantities of products in recent decades [9].
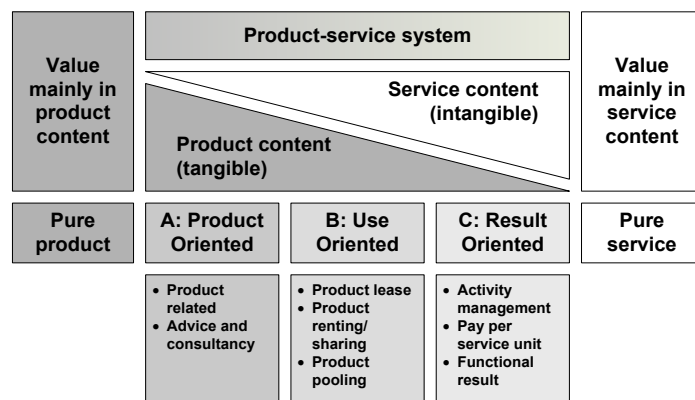
To attain their sustainability, PSS would be appropriate to implemented by industrial estate firms since it brings out an opportunity to find new added value and competitiveness. PSS can also increase customers' satisfaction by offering customized product and services, building a good customer relationship, and creating faster innovation opportunity [35]. All of these conditions can lead industrial estate firm to gain competitive advantage and then sustainability. A reference [36] stated that land and building selling in industrial estates need to be shifted into services in the future because of the availability of the limited products.

According to experts, the PSS is categorized into three different types, namely product-oriented, use-oriented, and result-oriented [10,36] as shown in Figure 2. Each of the type related to particular business activities and particular composition of products and services.

The first type is Product-oriented PSS. Product-oriented PSS is the type of PSS that most closely resembling a traditional business, where products ownership transferred to the customer as the main business, but with the addition of services to extend products' life and retrieve them when the product is unusable for recycling[10]. Reference [35] stated 2 business activities of product-oriented PSS, namely product-related services and also advice and consultation to prolong the product's useful life as an effort to decrease the product consumption [35]. Example of product-oriented PSS implementations in industries is Eastern Energy, which provides energy management; consumption and process monitoring; utility awareness; and training to help their customer maintain the products that will prolong their life cycle [10]. The other example is Polyplank AB that creates cheap, recyclable and moisture-resistant composite material to be used for paper mill's core plug from plastic waste [37]. Polyplank AB also takes back the core plug to be reused four times before need to be replaced [37]. With this system, products consumption can be reduced.

The second type is use-oriented PSS. The Use-oriented PSS is a PSS type in which the tangible product is used, but the ownership of the product is not transferred to the customer for the purpose of improving the utilization and extending the useful life of the product [10]. Business activities in this type are product lease, product renting/ sharing, and product pooling [35]. Examples of this type are vehicle sharing group, where vehicle provided by the provider and used by several customers by sharing scheme [10]. In Seoul, a public bicycle system operated to reduce energy consumption, raise community health status as well as reduce bicycle used and budget for buying bicycle by individuals [38]. The other example is Conway's water purifier which built a rental system with the regular visiting maintenance service for water purifier [39]. With this rental system, the customer does not need to buy the product, and other customers can use the product when he/she does not intend to use it. Regular visiting of maintenance will help the product to be in good condition and can be used for a longer time.

The last type is result-oriented PSS. Result-oriented PSS is the type of PSS with the greatest service composition. In this type, customers buy services in the form of capabilities and results [10]. Business activities in this type are activity management, pay-per-service unit, and functional result [35]. Example of this type o PSS implementations is Xerox pay-per-copy system, where the

Figure 2. Types of Product-Service System (PSS) [35]

customer will lease the copy machine and they will pay per copy they made counted by the machine [10]. Xerox will also provide maintenance service and will take the machine back after the leased in finished. With more or less similar system, Electrolux also create a system, which customers pay initial fee for placing the washing machine in their place, then payment would be made per wash [10]. Electrolux also is responsible for the maintenance, reparation, and another financial service of the washing machine [10]. With this system, the washing machine useful life will be prolonged, since the cost for maintenance and repair done by the experts, while customers benefit from efficient cost.

Among these three types of PSS, there would be differences in product and service proportion, as well as business activities. Product-oriented PSS has the largest proportion of the product, followed by the use-oriented PSS, thus conversely, result-oriented PSS has the largest service proportion. Business activities of each type of PSS would be appropriate to be applied at each stage of the firm life cycle of industrial estate firms to help industrial estate firms to be sustained in each stage of the life cycle.

This article will discuss the conceptual framework of PSS management at each stage of the firm life cycle. The discussion will take the context of Indonesian industrial estate firms that have important roles in Indonesian economic development as stated in the Indonesian Constitution Number 3 the year 2014. This article will provide a novelty on the application of PSS types at the stages of the firm life cycle within the context of industrial estate firms in Indonesia.

3. Research Objectives and Methodology

From the beginning, the Government of Indonesia built industrial estate firms with the main objective of controlling industrial development, controlling the environmental burden, overcoming the limitations of industrial support, and accommodating housing needs around industrial sites [36]. Later on, the purpose of an industrial estate building is growing primarily to develop the industry [36] in an attempt to support national economic growth [14]. Different from most countries, starting from 1989 until now, industrial estates in Indonesia are allowed to be built by the private sector [36]. With this permission, industrial estates in Indonesia are developing even faster. The fast development is due to capital addition and new innovation from private firms. In future development, industrial estates in Indonesia are directed and expected to adopt the concept of a modern industrial estate for the development of a sustainable national industrial competitiveness [36]. In addition, they are also expected to contribute to aspects of sustainability, namely economic, social, and environmental development [36]. The future development of Indonesian industrial estates is also directed at providing services rather than merely selling products [36].

This article will use the PSS concept to help Indonesian industrial estate firms that play important roles in national economic development to gain enough profit and can be sustained in each stage of the life cycle. According to the previous research, actually Indonesian industrial estate firms are starting to do business activities that match with PSS concept [28]. This finding indicates that Indonesian industrial estate firms are suitable for PSS implementation. Therefore, this article will provide a conceptual framework to manage PSS in each stage of the

industrial estate firms' life cycle to sustain the industrial estate firms. This article will also provide direction for relevant further research. The conceptual framework also attempts to help the Indonesian government to set a plan, policy, and regulation that will be suitable for the industrial estate firms' development.
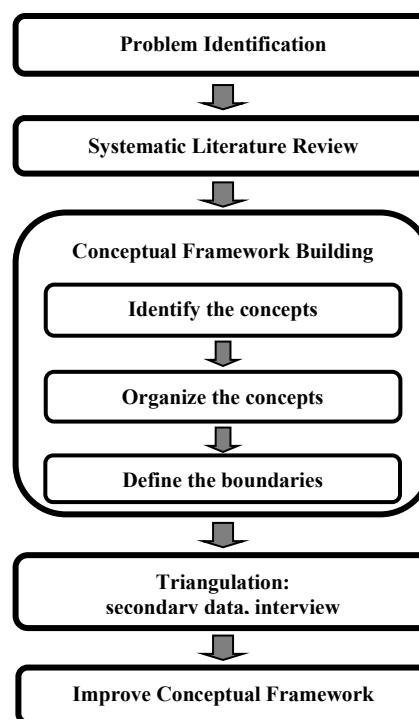
Figure 3. Conceptual Framework Building

To create the conceptual framework, deductive and inductive approach used as shown in Figure 3. A deductive approach will be used to build the framework based on literature reviews while inductive approach will be used to test and improve the framework.

The research begins by identifying problems by examining the condition of industrial estates in Indonesia. Next, authors conduct a systematic literature review on the industrial estate, firm life cycle and PSS to find suitable concepts to build the conceptual framework. The conceptual framework then was built deductively. To build the conceptual framework, there are several methodologies known. This article uses a methodology that suggests three steps of conceptual framework building: (1) identify the concepts, (2) organize the concepts, and (3) define the boundaries [40]. After that, some cases of industrial estate firms will be used to test the conceptual framework. The data will be obtained from secondary data from websites and results. The framework testing will use an inductive approach. The conceptual framework would be improved if needed based on the test result.

To test the conceptual framework, 5 industrial estate firms would be obtained. Four of them are 2015 industrial estate awards winners, namely First Industrial Estate, Second Industrial Estate, Third Industrial Estate, and Fourth Industrial Estate. Industrial estate awards event was conducted in 2015 by Indonesian Industrial Ministry especially to monitor industrial estate development. These industrial estate firms were chosen because they are well-established, good performers and already proven their performance and sustainability for about 30 years. One of the

industrial estate firms taken is the latecomer or new industrial estate firms, namely Fifth Industrial Estate. This industrial estate firm is taken to compare the condition with the four well-established industrial estate firms. Interviews were conducted with experts whose are works in industrial estate firms and consultant of industrial estate firms. The names of the industrial estate firms are disguised in this article, to maintain the confidentiality of the firms.
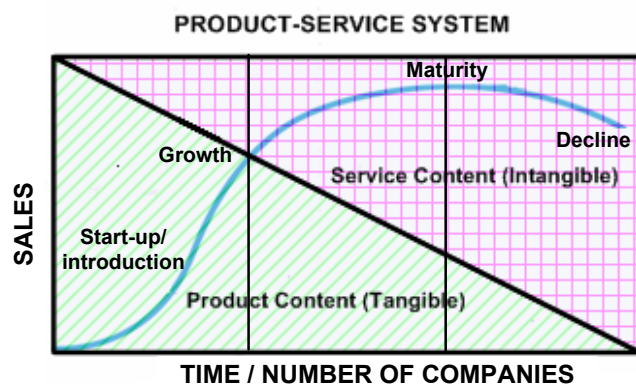
Figure 4. Conceptual Framework of PSS Management

4. Conceptual Framework

As discussed earlier, the problem faced by Indonesian industrial estate firms is how to sustain at every stage of the firm life cycle. Based on this problem, systematic literature reviews were conducted to build the conceptual framework. The first step in building a conceptual framework according to the chosen methodology is to identify the concepts. From the literature review, observation, and interview, it is found that industrial estate companies experience the firm life cycle pattern [41]. At the initial stage of the life cycle, industrial estate firms gain enough profit

from selling the product in the form of land and building. But after that, the profit will be decreased because the land and building would be sold out.

Some of Indonesia's long-standing industrial areas maintain their sustainability by expanding and/or creating new business models based on services. The literature review is also carried out to find a suitable concept to help industrial estate companies to be sustainable at each stage of the firm life cycle. The literature review found that the article that links between PSS with the firm life cycle is still rare. In addition, articles on the application of PSS to industrial estate companies have not been found. Therefore, this article will try to apply the concept of PSS related to the life cycle of the company in the Indonesian industrial estate firms. The concept of PSS is found to be appropriate to address this problem, as it offers a new business model that combines products with services. Then it can be concluded that this article will use the concepts of industrial estate firms, firm life cycle, and PSS.

The second step is to organize the concepts. As discussed previously, Indonesian industrial estate firms will generally follow the firm life cycle [41] although the details of the form and the pattern period may be different from the common pattern. Then to be sustained at each stage of the life cycle, the firms need to apply particularly suitable strategies [29]. Considering the characteristics of industrial estate firms that have limited tangible resources, the PSS concept that seeks to shift the use of products to services will be appropriate. The concept of PSS consists of 3 types, each of which consists of different business activities to shift the consumption of products into services [35]. Each of these types of PSS can be implemented as a strategy at each stage of the product life cycle in an attempt to be sustained. For example, for the introduction and growth stages, product-oriented PSS would be suitable, because mostly in these stages industrial estate firms sell products. But this is not the case for other stages. The relationship of the concepts built the conceptual framework for industrial estate firm as described in Figure 4.

This figure shows that in course of their life, industrial estate firm, as also most firms, will face stages of firm life cycle, namely start-up/introduction, growth, maturity, and decline [8]. Due to the limitations of products to be sold, then in the early stages of the firm life cycle, industrial estate firms can sell more products than services. However, in the later stages, when the product is sold out, in order to sustain they need to provide more services. This condition is matching with the characteristics of the PSS types' distribution, namely product-oriented PSS, use-oriented PSS, and result-oriented PSS. Industrial estate firms can appropriately apply product-oriented PSS in the early stages of the product life cycle, where they still have products to sell or product sales are still increasing. After that, use-oriented PSS will be appropriately applied, when the product starts to run out or product sales begin to decline. At last, result-oriented PSS will be appropriate to be applied when the product is fewer or even exhausted and product sales are declining significantly, so new income from the new business model is needed. With the new business, the industrial estate will be survived even in decline stage.

For each type of PSS, there are business activities that can be implemented by the firms. Using the product-oriented type, most firms sell product and added services [35]. Most Indonesian

industrial estate firms' core business is selling products in form of land and buildings. Some industrial estate firms also sell supporting facilities and equipment, but usually, this is not the core business. Service provided by the firms that accompanying product, mainly with purpose to prolong its useful life, in form of:

(1) Product-related services. The business activity where firms sell products to the customer but also add services to prolong the product's useful life, efficient use and take back after usage [35]. Industrial estate firms provide maintenance, repair and supplies services for companies' land, buildings, and facilities, such as the industrial vehicle; machinery; waste; water treatment, etc.

(2) Advice and consultancy. The business activity where firms give advice and consultancy regarding companies operations, management or setup to help prolong the product's useful life and efficient use [35]. In these business activities, the industrial estate firms can provide assistance in form of providing training and consultancy for several activities, such as inventory control, company setting, organizational management, waste treatment, maintenance, etc.

Using use-oriented type, firms sell the use or availability of product [10] instead of sell product [35]. Indonesian industrial estate firms' do business activities of this type in form of :

(1) Product lease. The business activity where firms lease products instead of selling them [35]. Products ownership would still be at producer instead of customer [35]. The producer also has responsibilities of maintenance and taking back products at the end of its useful life [35]. With this business activity, customers get unlimited access to the products leased [35]. Industrial estate firms can lease land; building; industrial vehicles; warehouse; and other equipment to customer/user. Customer/user can use and access the products unlimitedly during the period of leasing. Hence, customer/user is defined as industrial companies operated in the industrial estates firms. Companies, as the customer/user, pay regular fee periodically, while maintenance and repairment ensured by industrial estate firms. This action will ensure cost-efficiency in the whole industrial estate firms due to economic scale and utilization.

(2) Product renting or sharing. The business activity where customer/user rent products and the products can be used by more than one customer [35]. Here, customer/user does not have unlimited access [35]. Industrial estate firms can provide facilities, such as a warehouse, office building, and equipment, such as the vehicle, and machinery to be rent to companies. Customers can share in using these products with other. Industrial estate firms can also rent it to more than one customer with the particular arrangement. This scheme will increase the utilization and productivity of products, decrease total products produce, as well as benefit the customer at an efficiency, practical and lower cost.

(3) Product pooling. The business activity where products can be used together and be operated for more than one customers simultaneously [35]. Industrial estate firms can provide some industrial vehicles, equipment, and facilities that can be used by customers simultaneously. For instance, to deliver production results, more than one customers can use the same vehicle. Another example is water treatment, which more than

one customers can use the same equipment. This scheme would be applicable for industrial vehicle and equipment. However, for land, buildings, and specific machinery this scheme will only be applicable for companies that only need the products occasionally or operated in a low capacity.

Using the result-oriented type, the firm sells result or capability instead of products, that can be customized to what customers/users needs [10]. Customers/users will then pay only for the result used [10]. Industrial estate firm can do business activities in form of:

(1) Activity management/outsourcing. The business activity where firms provide outsourcing for customers [35]. Industrial estate firms can provide various services that will support customers to operate more efficiently. Some services that can be provided are cleaning services; garden maintenance; security services, etc. Customers may use those services with the particular amount of payment included in the contract. With this business activities, customers can focus on their own core business, and do not need to waste much energy and effort to manage the supporting activities and the workforce who carry it out.

(2) Pay per service unit. The business activity where firms provide services for customers that are charged per services used by the customer [35]. Industrial estate firms can provide document center; food service; laundry service; load or unload service; and transportation service. Customers can use the services by paying the result according to the number of output needed and used. With this business activity, customers will benefit from efficient cost spent on facilities that actually not continuously or intensely used.

(3) Functional result. The business activity where firms sell functional results instead of products [35]. Industrial estate firms can provide results as services for customers, such as a clean working space; comfortable office; healthy worker; skillful worker, etc. Customers will only pay for the results and get the benefit, without the need to manage or have the skill to do the process by themselves.

The third step of conceptual framework building is to define boundaries. This conceptual framework will be build for Indonesian industrial estate firms because the data were taken from Indonesian industrial estate firms. The basic concept would be compatible for industrial estate firms in other countries but the data must be obtained to adjust whether there are differences in the regulations, conditions, roles, and characteristics of industrial estate firms.

The next step is framework testing to confirm the deductive conceptual framework. Framework testing will be conducted using official websites, papers, and interviews with experts as summarized in Figure 4.

5. Result and Discussion

Indonesian government issuing regulation that industrial estate firms have obligation to provide at least: (1) standard water treatment installation; (2) sewage water treatment installation; (3) drainage; (4) street-lighting; (5) and road for the companies operated in the estate[15]. Then we can conclude that these are the minimum requirement of industrial estate firms. In Indonesia,

according to several sources, it is known that most industrial estate firms depend so much on products selling since it gives large enough profit compares to providing services.

To investigate the implementation of types of PSS and their business activities in Indonesian industrial estate firms, data were obtained from the official websites, the papers, and interviews with experts that closely related to industrial estate firms. As mentioned before, in this article we study 5 industrial estate firms to test the framework. The first four industrial estate firms are 2015 industrial estates award winners whose operated about 30 years, and the fifth is a new industrial estate firm, operated below 5 years. The first three industrial estate firms are located in Java island and the last two are outside Java island.

From our study, we found that all industrial estate firms studied already fulfilled, even exceed, all standards determined by the government. They already provide water supply, water treatment and waste treatment services, security, lighting, road and network, and other infrastructure. The industrial estate firms also found already doing business activities in types of PSS, as summarized in table 1.

From the data, we found that all industrial estate firms studied are focus on selling land and/or buildings as their main business in the early firm life cycle period. Beside selling land and/or buildings, all industrial estate firms also provide maintenance and security services to help customers in operating their businesses. Additionally, the fourth industrial estate provides consultancy and advice services to assist companies (as customers) that will invest in the industrial estate, and assist license application. The fifth Industrial Estate also provides 3 hours investment licensing system to ease companies joining the estate. This fact has proven that Indonesian industrial estate firms in the early stage of life cycle already implement business activities that comply with product-oriented PSS.

Although all industrial estate firms implement product-related business activities, the fourth industrial estate, and the fifth industrial estate also provide implementation advice and consultancy. Among the industrial estate firms, it is found that the kind of services and products provided are almost the same. All of them get the most profit that comes from selling products and provide services in form of maintenance, repair, and security. The fourth industrial estate and the fifth industrial estate have uniqueness in form of license application assistance that maybe become the competitive advantage.

As mentioned in table 1, all industrial estate firms provide warehouses for lease, while the first industrial estate and the second industrial estate also lease factories to customers and the fifth industrial estate rent tank farms. This fact shows that all industrial estate firms already implement product lease business activity with almost the same kind of services. The third industrial estate found rent vehicle, while the fifth industrial estate plan to built meeting incentives, convention, and exhibition (MICE) facility. In many countries, MICE recently believed can attract visitors and bring economic benefit, raising good image, and have a positive impact on the development of the local region as well as the country.

Due to implementing and planning rent business activities, the third and the fifth industrial estate already implement the product rent/sharing business activity. All industrial estate firms provide logistic service with pool system. Additionally, the second and the fifth industrial estate provide a dry port facility that can ease and fast the distribution process. This condition shows that the first to the fourth industrial estate already implement use-oriented PSS through the business activities. Whereas the fifth industrial estate as the considerably new industrial estate has also implement use-oriented PSS, although some activities are still in the planning phase. In use-oriented PSS, it can be seen that there slightly different services provided by industrial estate firms. The differences mainly because of difference in resource and capability owned by the industrial estate firms.

The third industrial estate provides custom office in the estate, while the fourth industrial estate provides manpower management and recruitment service and provide immigration clearance to help companies in the estate. The second industrial estate provides security, office boy and cleaning service outsources, managed by the industrial estate firm. Paying periodical fee, companies in the estate can use these outsourcing services. These services show that the industrial estates already do management business activity, helping customers to do their non-core business with outsourcing system. The first to the fourth industrial estate already do pay per service unit business activities, while the fifth industrial estate still finds the one pay per service business activity. All industrial estates provide hotel. The first to the fourth industrial estate provide health service. The unique pay per services business activity provided by the second industrial estate by renting helicopter; the third industrial estate by providing vehicle service; and the fourth industrial estate by providing dormitory and executive village. Furthermore, the first to the fourth industrial estate already do functional result business activity. The first to the fourth industrial estate provide commercial facilities or shop houses and sport and entertainment facilities, while the first and the second industrial estate build a city around the industrial estate, the second and the third industrial estate build educational facilities, and the second industrial estate built senior living facility. The fifth industrial estate plan to build community academy, but the training already conducted for Sei Mangkei's human resource. From these business activity implementations, it can be concluded that the first to the fourth industrial estate firms already implement result-oriented PSS, while the fifth industrial estate firm still starting to plan to implement result-oriented PSS, and already start to provide training. Based on the analysis, in result-oriented PSS, the difference of services provided even more. These difference can become the competitive advantage for each industrial estate firm.

All types of PSS already found in the first to the fourth industrial estate practice that means services already taken as an important point in their business. However, they were still emphasizing the limited land and building sales, especially during the early stages of their life cycle, since it is providing a great profit. That is why most industrial estate firms attempt to expand the land or find another location once the products sold out. This is due to the fact that selling products still becoming the greatest income generator for industrial estate firms. Nevertheless expanding the industrial estate not always a simple or possible option, because of the high investment and spatial regulation. On the other hand, they still have

Table 1. PSS Implementation in Industrial Estate Firms

Types of PSS	Business activities	First Industrial Estate	Second Industrial Estate	Third Industrial Estate	Fourth Industrial Estate	Fifth Industrial Estate
Product-oriented PSS	**Product related**	• Sell land and building • Provide maintenance • Provide security	• Sell land and building • Provide maintenance • Provide security	• Sell land and building • Provide maintenance • Provide security	• Sell land and building • Provide maintenance • Provide security • Advice and assist companies on incorporation • Advice and assist license application	• Sell land and building • Provide maintenance • Provide security • Provide assistance with 3 hours of investment licensing service
	Advice and consultancy					
Use-oriented PSS	**Product lease**	• Lease warehouses • Lease factories	• Lease warehouses • Lease factories	• Lease warehouses	• Lease warehouses	• Lease warehouses • Lease tank farms • Rent meeting incentives, convention, and exhibition facility (in the plan)
	Product renting/ sharing			• Rent vehicle		
	Product pooling	• Provide Logistic service	• Provide logistic service • Provide dry port facility	• Provide logistic service	• Provide logistic service	• Provide logistic service • Provide dry port facility
Result-oriented PSS	**Activity management**		• Provide security, cleaning service, office boy outsource	• Provide custom office	• Provide manpower management and recruitment service • Provide immigration clearance	
	Pay per service unit	• Provide hotel • Provide health service	• Provide hotel • Provide health service • Provide helicopter rent	• Provide hotel • Provide health service • Provide vehicle service	• Provide hotel • Provide health service Provide dormitory • Provide executive village	• Provide hotel
	Functional result	• Provide commercial facilities • Provide sport and entertainment facilities • Built city	• Provide commercial facilities • Provide sport and entertainment facilities • University • Built city • Provide senior living facility	• Provide shophouses • Provide sport and entertainment facilities • Vocational high school	• Provide commercial facilities • Provide sport and entertainment facilities	• Provide training as an impetus of community academy

Source: summarized from official websites [42]–[45], a conference material [46]and interviews with experts

obligations to their customers as in the regulation that they still must manage the industrial estate and provide service to the companies inside. Therefore, they manage to provide various services to be sustained. The more they mature they will find more effort to innovate to provide more various services.

The first to the fourth industrial estate was built in the 1990s, range from 1989 until 1993. It means that they have operated more than 30 years and already in the middle or later stages of the life cycle. This is one of the reasons why their business activities tend to innovate in providing services. By providing more services,

industrial estate firms have more opportunity to do innovation. With innovation, firms have more opportunity to gain new profit to be sustained. In contrast to the four industrial estates that has long been operating, the fifth industrial estate has only been operating for about 3 years and still at the early stage of firm life cycle. So the fifth industrial estate still provides a little kind of service or still planning the services. This conforms with the framework that the more mature the industrial estate the more they need to provide service. It can be said that use-oriented PSS and result-oriented PSS would be needed when an industrial estate firm going to runs out their land and buildings to keep on sustain.

Among the industrial estate firms, while implementing product-oriented PSS, they mainly sell and provide similar products and services with similar business activities. Nevertheless, by using use-oriented PSS and moreover, result-oriented PSS, which is the composition of providing services more than products, among industrial estate firms there are differences in business activities and services provided. For example, there are industrial estate firms that rent factories, build a city, provide sport and entertainment facilities while the other create services such as healthcare, immigration clearance, etc. Then, it can be concluded that by providing more services, industrial estate firms can do more various business activities.

From the conceptual framework and discussion, there are propositions that can be built regarding PSS implementation in industrial estate firms toward sustainability, considering the firm life cycle.

Proposition 1: Industrial estate firm would need to provide more services in the later stage of the firm life cycle to be sustained because they have limited tangible products.

Proposition 2: PSS concept supposed to be suitable to be implemented for industrial estate firms, due to the emphasizes on services providing rather than tangible products.

Proposition 3: Among types of PSS, product-oriented, use-oriented product-oriented composition of PSS is getting bigger, leading to more possibilities for innovation, as services are given more opportunity for it.

Proposition 4: More service composition leads to the industrial estate firms that have more flexibility to create business activities. That is why in the application of results-oriented PSS, there are many variations in the types of services provided between industrial estate firms. This variation can then become a uniqueness that will form a competitive advantage for industrial estate firms.

Proposition 5: In relation to Proposition 2, industrial estate firms with unique business activities and competitive advantage based on services will have the greater likelihood of profit and sustainability.

Proposition 6: All three types of PSS are appropriate to be applied to each product life cycle in order to maintain the sustainability of the industrial estate firm at each stage of the product life cycle. Product-oriented appropriate PSS is applied in the early stages of the company's life cycle when the firm is new and sells many products in the form of land and buildings. The appropriate use-oriented PSS is applied at the mid-life stage of a company when the product begins so little that the service has begun to be emphasized. Result-oriented PSS is appropriate to apply to the end of the firm life cycle when the product is getting smaller or exhausted, so the service needs to be the main business activity.

6. Conclusion

As already been explained, this article is extending the previous work in 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM) [1]. In this article, we added the concept of firm life cycle and data from interviews with experts. The novelty of this article is by added firm life cycle concept in PSS management. This concept is added because of the knowledge regarding industrial estate firms will follow a firm life cycle pattern [41]. As discussed earlier, that would create differences with the condition [28] and strategy needed [29]. Therefore, different PSS management would be promptly needed.

From the above discussion, it can be concluded that the concept of PSS can be applied to help Indonesian industrial estate firms to be sustainable across their life cycle. This is strengthened by data and interview results that many Indonesian industrial estate firms have started to think and provide services combined with products sales. Among firms, there are differences of services provided and business activities done. These findings are in line with the reference that state industrial estate firms need to shift to provide service for further development to be sustained [36].

Facing the firm life cycle, industrial estate firms can use the three types of PSS with the business activities. Industrial estate firms can do business activities that are suitable for their conditions regarding with firm life cycle. The type and business activities can be changed when the condition of the firm change along the firm life cycle. The suitable PSS management would lead to firms' sustainability in each life cycle. The conceptual framework suggests that the types of PSS implementations can assist industrial estate firms at each stage of the firm life cycle through creative innovation of the services provided to enable the industrial area to play its role well.

From the data, it is also found that the more the service composition is provided, the industrial estate firm will have more flexibility to innovate creatively. The more innovation would lead to more opportunity creation and the ability to choose the right business and service type. With this benefit, the industrial estate firms would get a competitive advantage, not to mention sustainability.

The implication of the findings is that industrial estate firms need to manage their PSS as an attempt to be sustained. They better plan the PSS in the beginning for they can get more stable and continuous profit. The planning for PSS from the beginning will also ease industrial estate firms to allocate the service facilities in the industrial estate and maintain good integration in the estate.

This article will help industrial estate firms to see opportunities for business activities that meet the conditions that can be applied at each stage of the firm life cycle based on the PSS concept. This concept also provides suggestions for industrial estate firms in determining strategies at each stage of the firm life cycle to be sustainable. For the government, this article will support the

government's efforts to develop appropriate policies in the development of the industry through the role of the industrial estate, which in turn improves the national economy and welfare. In the academic field, this article presents the incorporation of industrial estate concepts, corporate life cycles and PSS that are currently unexamined. This conceptual framework can generally be used for industrial estate firms because the basic theory used is general. However, since the empirical data in this article are only taken from industrial estate firms in Indonesia, there may be differences in government regulations, and conditions of competition, business, economy, industry conditions, etc for other countries that may require adjustment.

The implementations of PSS in this article are identified from the websites of industrial estates award winners and the interviews with industrial estate firms experts. For a better study, data from the websites need to be triangulated to get better descriptions. In addition, for further research, it is necessary to study more industrial estates to strengthen the results obtained. The study with non-award winners industrial estate firms also needs to be conducted to study to further test the concept. Moreover, the new and middle age industrial estate firms also need to be studied, to confirm the concept. Then our next research is to test the conceptual framework that has been built with interviews with several industrial estate firms with different conditions.

Further research about the comparison between private and public industrial areas as well as the comparison between industrial areas located in Java and those located outside Java should be employed. Research can also be extended by combining other concepts of operational strategies. Another research can also be conducted to study the firm life cycle pattern. Further study would also be needed to study the industrial estate firms with various time operated to better see the pattern of implementation of each type of PSS in industrial firms.

Conflict of Interest

The authors declare no conflict of interest

Acknowledgment

References

[1] C. Wirawan, G. Yudoko, and Y. D. Lestari, 'Product-service system for Indonesian industrial estate firms : A conceptual framework', in *Proceedings of 2017 IEEE IEEM*, Singapore, 2017, 1812–1816.https://doi.org/10.1109/IEEM.2017.8290204

[2] P. Scott, 'Industrial Estates and British Industrial Development, 1897-1939', *Bus. Hist.*, 43(2), 73–98, 2001.http://dx.doi.org/10.1080/713999223

[3] A. R. R. Ramos and F. P. Fonseca, 'A methodology to identify a network of industrial parks in the Ave Valley, Portugal', *Eur. Plan. Stud.*, 20(10), 1844–1862, 2016.http://dx.doi.org/10.1080/09654313.2016.1202201

[4] World Bank Group, *Getting to Green - A Sourcebook of Pollution Management Policy Tools for Growth and Competitiveness.* www.worldbank.org, 2012.

[5] F. Fonseca, R. A. R. Ramos, and A. N. R. da Silva, 'An agent-based model to assess the attractiveness of industrial estates', *Jasss*, 18(4), 1–11, 2015.DOI: 10.18564/jasss.2893

[6] United Nations Industrial Development Organization, 'Economic Zones in the ASEAN: Industrial parks, special economic zones, eco-industrial parks, innovation districts as strategies for industrial competitiveness', August. 2015. DOI:10.1016/j.indmarman.2016.03.001

[7] B. R. Scott, *Stages of corporate development.* Boston: Harvard Business School Press, 1971.

[8] H. Mintzberg, 'Power and Organization Life Cycle', *Acad. Manag. Rev.*, 9(2), 207–224, 1984.

[9] O. K. Mont, 'Clarifying the concept of product-service system', *J. Clean. Prod.*, 10(3), 237–245, 2002. https://doi.org/10.1016/S0959-6526(01)00039-7

[10] T. S. Baines, H. Lightfoot, E. Steve, A. Neely, R. Greenough, J. Peppard, R. Roy, E. Shehab, A. Braganza, A. Tiwari, J. Alcock, J. Angus, M. Bastl, A. Cousens, P. Irving, M. Johnson, J. Kingston, H. Lockett, V. Martinez, P. Michele, D. Tranfield, I. Walton, and H. Wilson, 'State-of-the-art in product service-systems', in *Proceedings IMechE, 221 Part B: J. Engineering Manufacture*, 221, 1543-1552, 2007.DOI: 10.1243/09544054JEM858

[11] A. P. B. Barquet, M. G. de Oliveira, C. R. Amigo, V. P. Cunha, and H. Rozenfeld, 'Employing the business model concept to support the adoption of product-service systems (PSS)', *Ind. Mark. Manag.*, 42, 693–704, 2013.http://dx.doi.org/10.1016/j.indmarman.2013.05.003

[12] CDMI Consulting Group, 'Studi potensi bisnis dan pelaku utama kawasan industri di Indonesia 2016-2020', Jakarta, 2015.

[13] Republic of Indonesia Industrial Ministry, *Rencana Induk Pembangunan Industri Nasional 2015-2035 (National Industrial Master Plan 2015-2035).* 2015.

[14] Republic of Indonesia., 'Undang-Undang Republik Indonesia Nomor 3 Tahun 2014 Tentang Perindustrian (Republic of Indonesia Law no. 3 the year of 2014)'.

[15] Republic of Indonesia. Industrial Ministry., 'Peraturan Pemerintah Republik Indonesia Nomor 142 Tahun 2015 Tentang Kawasan Industri (Republic of Indonesia Government Regulation No. 142 the year of 2015 about industrial estate)'.

[16] G. Yudoko, 'Strategy map for industrial estate firms: model building in the Indonesia Regulatory Context', *Adv. Sci. Lett.*, 22(12), 4316–4320, 2016.https://doi.org/10.1166/asl.2016.8136

[17] G. Yudoko, 'The Indonesia Industrial Estate Firm Award's: Theory Building and Testing', *Soc. Sci.*, 11(6), 7341–2345, 2016.

[18] J. Elkington, 'Enter the Triple Bottom Line', in *The Triple Bottom Line: Does It All Add up?* A. Henriques and J. Richardson, Eds. London: Earthscan, 2004, 1–16.

[19] E. Dewberry, M. Cook, A. Angus, A. Gottberg, and P. Longhurst, 'Critical Reflections on Designing Product Service Systems', *Des. Journal*, 6(4), 408–430, 2013.DOI: 10.1016/j.compind.2012.02.006

[20] S. Wiesner, M. Freitag, I. Westphal, K.-D. Thoben, and X. Boucher, 'Interactions between service and product lifecycle management', in *Procedia CIRP*, 2015, 221, 36–41.DOI: 10.1016/j.compind.2012.02.006

[21] S. Cavalieri and G. Pezzotta, 'Product-service systems engineering: State of the art and research challenges', *Comput. Ind.*, 63, 278–288, 2012.DOI: 10.1016/j.compind.2012.02.006

[22] V. Zanetti, S. Cavalieri, and G. Pezzotta, 'Additive Manufacturing and PSS: a Solution Life-Cycle Perspective', *IFAC-PapersOnLine*, 49(12), 1573–1578, 2016.DOI: 10.1016/j.ifacol.2016.07.804

[23] S. Singhal and A. Kapur, 'Industrial estate planning and management in India - An integrated approach towards industrial ecology', *J. Environ. Manage.*, 66, 19–29, 2016.DOI: 10.1006/jema.2002.0571

[24] V. Veleva, S. Todorova, P. Lowitt, N. Angus, and D. Neely, 'Understanding and addressing business needs and sustainability challenges : lessons from Devens eco-industrial park', *J. Clean. Prod.*, 87, 375–384, 2015.http://dx.doi.org/10.1016/j.jclepro.2014.09.014

[25] C. K. Prahalad and V. Ramaswamy, 'Co-creating unique value with customers', *Strateg. Leadersh.*, 32(3), 4–9, 2004.

[26] I. M. Jawahar and G. L. McLaughlin, 'Toward a descriptive stakeholder theory: An organizational life cycle approach', *Acad. Manag. Rev.*, 26(3), 397–414, 2001.https://doi.org/10.5465/amr.2001.4845803

[27] D. C. Mueller, 'A Life Cycle Theory of the Firm', 20(3), 199–219, 2013.DOI: 10.2307/2098055

[28] C. R. Anderson and C. P. Zeithaml, 'Stage of product life cycle, business strategy, and business performance', *Acad. Manag. J.*, 27, 5–24, 1984.https://doi.org/10.5465/255954

[29] S. H. Hanks, C. J. Watson, E. Jansen, and G. N. Chandler, 'Tightening the life-cycle construct : A taxonomic study of growth stage configurations in organizations', *Entrep. Theory Pract.*, 18(2), 5–30, 1993.https://doi.org/10.1177/104225879401800201

[30] A. Chandler, 'Strategy and structure'. Cambridge. MA, MIT Press, 1962.

[31] R. E. Quinn and K. Cameron, 'Organizational life cycles and shifting criteria of effectiveness: some preliminary evidence', *Manage. Sci.*,29(1), 33–51, 1993.https://doi.org/10.1287/mnsc.29.1.33

[32] M. Kuncoro and S. Wahyuni, 'FDI impacts on industrial agglomeration: the case of Java, Indonesia', *J. Asia Bus. Stud.*, 3, 65–77, 2009.https://dx.doi.org/10.1108/03068290910921226

[33] T. Firman, 'Rural to urban land conversion in Indonesia during boom and bust periods', *Land use policy*, 17(1), 13–20, 2000.DOI: 10.1016/S0264-8377(99)00037-X

[34] M. J. Goedkoop, C. J. van Halen, H. R. te Riele, and P. J. Rommens, 'Product Service systems, Ecological and Economic Basics', PricewaterhouseCoopers N.V. / Pi!MC, Storm C.S., Pre consultants, 1999.

[35] A. Tukker, 'Eight types of product-service system: Eight ways to sustainability? Experiences from suspronet', *Bus. Strateg. Environ.*, 13, 246–260, 2004.DOI: 10.1002/bse.414

[36] D. Mulyadi, *Manajemen Perwilayahan Industri*, First. Jakarta, Indonesia: Leuser Cita Pustaka, 2012.

[37] M. Lindahl, E. Sundin, and T. Sakao, 'Environmental and economic benefits of Integrated Product Service Offerings quantified with real business cases', *J. Clean. Prod.*, 64, 288-296, 2014. DOI: 10.1016/j.jclepro.2013.07.047

[38] S. M. Lee, D. L. Olson, and S. Trimi, 'Co-innovation : Convergenomics, collaboration, and co-creation for organizational values', *Manag. Decis.*, 50(5), 817–831, 2012. DOI: 10.1108/00251741211227528

[39] Y. Geum and Y. Park, 'Designing the sustainable product-service integration: A product-service blueprint approach', *J. Clean. Prod.*, 19, 1601–1614, 2011.DOI: 10.1016/j.jclepro.2011.05.017

[40] R. A. Swanson and T. J. Chermack, *Theory Building in Applied Disciplines*. San Francisco: Berret-Koehler Publishers, Inc., 2013.

[41] F. Belussi and S. R. Sedita, 'Life Cycle vs. Multiple Path Dependency in Industrial Districts', *Eur. Plan. Stud.*, 17(4), 505–528, 2009.DOI: 10.1080/09654310802682065

[42] 'Karawang International Industrial City'. [Online]. Available: www.kiic.com. [Accessed: 09-Jun-2017].

[43] 'Jababeka & Co. Industrial'. [Online]. Available: www.jababeka.com. [Accessed: 09-Jun-2017].

[44] 'MM2100 Industrial Town'. [Online]. Available: www.mm2100.ac.id. [Accessed: 09-Jun-2017].

[45] 'Batamindo Investment Cakrawala'. [Online]. Available: www.batamindoindustrial.com. [Accessed: 09-Jun-2017].

[46] Nurhidayat, 'Kawasan Ekonomi Khusus (KEK) Sei Mangkei Menuju Eco-Industrial Park', *National Seminar on Palm Oil Insight Panel Speech*. Yogyakarta, 2018.

SDN-based Network Control Method for Distributed Storage Systems

Luis Guillen[*,1], Satoru Izumi[1], Toru Abe[1,2], Hiroaki Muraoka[3,4], Takuo Suganuma[1,2]

[1]*Graduate School of Information Sciences, Tohoku University, 980-8577, Japan*

[2]*Cyberscience Center, Tohoku University, 980-8577, Japan*

[3]*Tohoku University, 980-8577, Japan*

[4]*Tohoku Institute of Technology, 982-8588, Japan*

A R T I C L E   I N F O

A B S T R A C T

With the increasing need for effective storage management due to ever-growing content-generation over the Internet, Distributed Storage Systems (DSS) has arisen as a valuable tool. Although DSS has considerably improved in the past years, it still leverages legacy techniques in its networking. To cope with the demanding requirements, Software Defined Networking (SDN) has revolutionized the way we manage networks and can significantly help in improving DSS network management. In this paper, we propose an SDN-based network control method that is capable of handling DSS network management and improving its performance. This paper presents the design, implementation, and evaluation using an emulated environment of a typical Data Center Network (DCN) deployment. The experiment results show that by applying the proposed method, DSS can increase the performance and service resilience compared to existing solutions.

1. Introduction

This paper is an extension of a previous work originally presented at the 2017 International Conference on Network and Service Management (CNSM2017) [1]. User-generated content is growing exponentially. From 30 Zettabytes (ZB) of content generated in 2017, it is foreseen to reach 160 ZB by 2025 [2]. Although end-users create most of this content, with the widespread use of the Internet of Things (IoT) and the advances in cloud technologies, content generation will also increase at the core of the network. It is also worth noting that in 2017 approximately 40% of the content was stored in enterprise storages [2], but due to the paradigm shift from expensive and large data-centers to cloud-based virtualized infrastructures, it is also projected to increase to 60% by 2025. The massive scaling and flexibility required in those infrastructures will demand more efficient ways to handle the additional traffic.

At the outlook of such demanding requirements, Distributed Storage Systems (DSS) became more popular, since they provide highly reliable services by networking nodes to provide enhanced storage [3]. Over the years, DSS has progressively achieved better performance by improving propagation and recovery methods, from simple replication to more advanced techniques [4,5]. However, little attention has been paid to improvements at the network level, as they still rely on legacy techniques.

To cope with the increasing need for efficient storage managed by DSS, in this paper, we propose a network control method that is capable of handling the generated traffic by specific DSS tasks. The proposed method is based on Software Defined Networking (SDN) [6], which separates the control plane from the data plane and will allow a more flexible programmable network. The contribution of this paper is to show the potential of applying this paradigm to improve DSS performance. Furthermore, we describe the inherent problems of DSS when using legacy techniques in Section 2, and the minimum requirements that DSSs demand from the network perspective, namely aggregated bandwidth and practical use of resource. Based on those requirements, we designed the proposed method as described in Section 3, whose main strength is its simplicity. More concretely, we define a solution to handle the bandwidth aggregation called *on-demand inverse multiplexing*, and we detail the solution for the practical use of resources called *multipath hybrid load balancing*.

*Corresponding Author: Luis Guillen, 2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan, +810222175080, Email: lguillen@ci.cc.tohoku.ac.jp

To test the feasibility of the proposed method, in Section 4 we evaluated the implementation of typical DSS scenarios, the results showed that our method outperforms traditional ones, and is capable of delivering the required features by DSS.
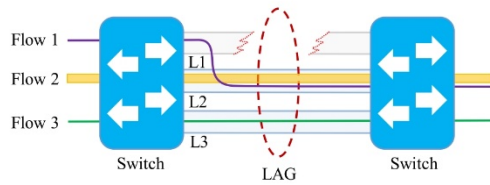
Figure 1: Programmability problem with LAG

2. Related Work

2.1. Problem Description with Legacy Networking Techniques

To the best of our knowledge, solutions that combine features of both technologies (DSS and SDN) have not been fully explored yet, however, in this section we present related work on individual features.

Initially, we should mention that there are two minimum requirements DSS will need from the networking perspective, namely: aggregated bandwidth and practical use of resources. In this sub-section, we describe the specific problems with legacy techniques when applied to DSS regarding these two aspects.

In the case of bandwidth aggregation, since DDS nodes might be allocated in different storage servers, a bottleneck is created at the server gateway when various clients try to access it at the same time, due to the limited bandwidth of a single link. Kaneko et al. [7] tried to overcome this problem by using Link Aggregation (LAG) [8], which allows bandwidth aggregation by grouping a limited number of physical links as a single logical bundle. LAG offers communication resilience by redirecting the incoming traffic to another active link in case of failure. However, there is no control on the selection process, e.g., in the simple LAG deployment depicted in Figure 1, three links (L1, L2, and L3) are grouped in a single bundle, if L1 fails then the protocol redirects the traffic to another link, but the traffic may well be sent through the most congested link (L2) instead of the L3 which has less traffic. Moreover, for a link to be part of a bundle, all of them need to have the same configuration, and it is limited to a hard-coded number of links, which significantly limits the flexibility of the system.

In the second case, namely the practical use of resources, DSSs are typically deployed on Data Center Networks (DCNs) using common topologies, such as a three-layer non-blocking fully populated network (FPN) or three-layer fat-tree network (FTN) [9]. In these environments, network traffic management is usually leveraged to techniques such as Equal Cost Multipath (ECMP) [10]. However, ECMP is not efficient regarding resource usage, and since DSS clients access several storages concurrently, congestion mostly occurs at some segments of the network. Additionally, since it is limited to legacy protocols such as Spanning Tree Protocol (STP), it will prune redundant links, which can be used for creating alternative paths. For example, in the topology shown in Figure 2, if two flows go from H1 and H2 to H3, some routes might be preferred (e.g., the red dashed path),

despite the availability of other paths. Moreover, as shown in Figure 3, a bottleneck is generated at the segment nearest to the end-device, for instance, if H5 have to reply flows from requests send by H1-H4, the link connected to the nearest switch will be highly congested (i.e., red dashed path). Apart from the number of
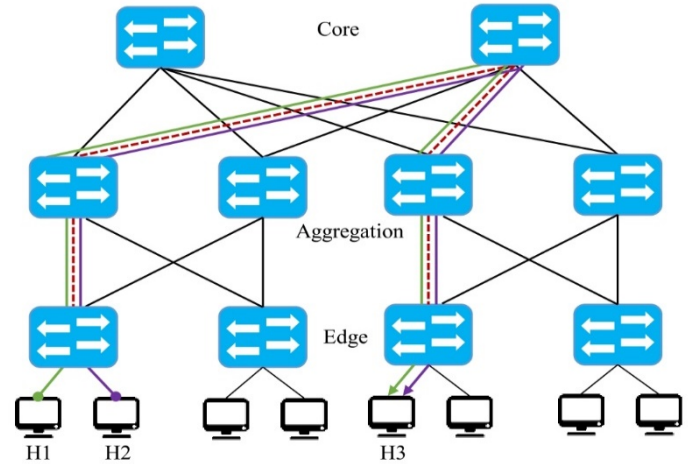
Figure 2: Traffic congestion issue in DCNs
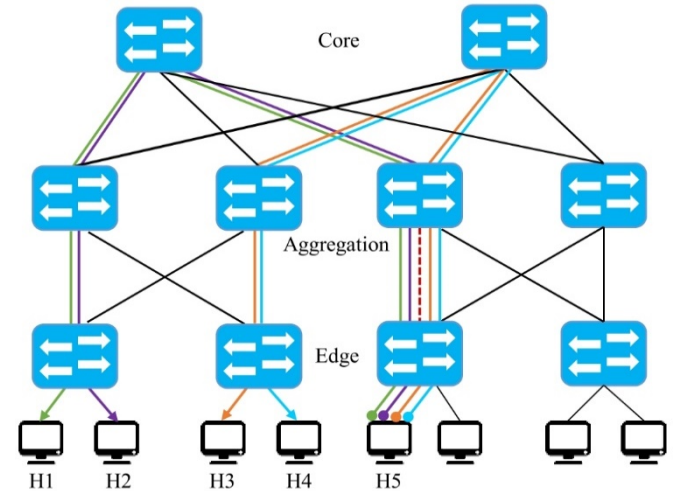
Figure 3: The last-mile bottleneck issue in DCNs

flows going through a single link, that particular segment of the topology the connection speeds are not as fast as they are in the core, we call this problem "last-mile bottleneck."

To sum up, legacy techniques cannot provide DSS with neither aggregated bandwidth or efficient use of resources due to the last-mile bottleneck issue, and limitations with end-to-end routing programmability with techniques such as ECMP.

2.2. Related work on SDN-based Multipath Load-balancing

The primary task towards a resource-efficient method is to balance the load among the available resources, and using multiple paths is a practical way to relieve the network traffic. Thus, in this section, we present load balancing solutions that focus on multi-path solutions using SDN.

Initially, it is worth noting that load balancing is a topic that has been extensively explored for decades, but SDN-based solutions are relatively more recent. SDN-based load balancing can be

categorized depending on their architecture in Centralized and Distributed [11], we focus on the centralized architectures wherein the management is performed in the data plane by a central controller that has an overview of the entire network. In this context, a pioneer work was *Hedera*, a dynamic flow scheduling for DCN capable of balancing the load among the available routes. The main contribution of the authors is the introduction of two placement heuristics to allocates flows, and the estimation of flow demands that uses OpenFlow (OF) for routing control. However, the authors did not consider last-mile bottlenecks nor the load on the server side.

OLiMPS [13] is a real implementation of Multipath TCP (MPTCP) [14] in an intercontinental OF-network that achieved high throughput. The drawback of this work is that, as other solutions [15, 16], they rely on MPTCP, which is an experimental protocol capable of handling resource pooling using multiple paths. However, the inherent problem is that it needs modifications in the end-point kernel. To avoid this restriction, Banfi et al. [17] presented MPSDN, a multipath packet forwarding solution for aggregated bandwidth, and load balancing that achieves similar results to MPTCP with the added value of not requiring end-point modifications. Their main contribution is the idea of including a threshold called Maximum Delay Imbalance (MDI), used to place flows in paths. However, even if they do not require end-point modification, they need to modify Open Virtual Switch (OVS).

DiffFlow [18] used a selection mechanism to handle flows in a DCN, so that short flows will be handled by ECMP and long ones using a process called Random Packet Spraying (RPS). However, by partially relying on ECMP they inherit the same issues concerning the use of resources.

Li and Pan [19] proposed a dynamic routing algorithm with load balancing for Fat-tree topology in OpenFlow based DCN. Their flow distribution strategy provided multiple alternative paths from a pair of end nodes and placed the flow to the one with the highest bandwidth. However, the hop-by-hop recursive calculation overlooked the available bandwidth of the entire network. A similar approach was presented by Izumi et al. [20], who proposed a dynamic multipath routing to enhance network performance, they introduce an index based on the risk and the use parallel data transmission and distribute the traffic using multiple paths. Dinh et al. [21] also presented a dynamic multipath routing, capable of selecting k-paths to distribute the traffic based on the load of the links; however, they only handled the initial assignment, and the selection of the number of paths is unclear.

Finally, Tang et al. [9] present an OF-based scheduling scheme that dynamically balances the network load in data centers. Their approach aimed to maximize the throughput by designing a heuristic based on the available resources in the network. A significant contribution of their work is that they present a theoretical model for load-balancing and specific metrics for network utilization and load imbalance. The drawback, however, is that they heavily rely on the use of a pre-calculated table (ToR Switch-to-ToR Switch Path Table S2SPT) for path selection, which dramatically limits the dynamicity in case of real implementation, as the computation time will increase if the number of links is relatively high. Moreover, they are still subject of the last-mile bottleneck problem due to the limitation in the DCN topology.

2.3. Target Issues

From the related work presented in the preceding sub-sections, and considering DSS requirements, we summarize our target issues as follows.

(P1) Last-mile bottleneck: Links have lower connection speeds at layers closer to the end-device in a DCN topology causes a network bottleneck. Therefore, DSS performance will be limited when performing parallel tasks.

(P2) Limited use of multipath routing: Despite the available redundant links in the topology, paths from end-to-end nodes are usually mapped as single-paths via some preferred routes, this provokes unnecessary congestion in specific segments of the network.

(P3) Load Imbalance: Due to the limited use of the available resources DSS neither the network nor the servers are used in a balanced manner. The existing solutions focus on either one of those aspects, which leads to overlooking the importance of both variables for an effective control method.

3. SDN-based Network Control Method for Distributed Storage Systems

3.1 Motivating example

We describe the proposed SDN-based network control method to solve the problems described above. However, before describing the proposed approach, let us consider a simple scenario of a DSS process.

Suppose we have a DSS that uses network codes [3] to recover failed nodes. In this scenario, the information of a node (D) is fragmented into p=4 pieces (p_1, p_2, p_3, p_4) stored in different nodes from the set of nodes (N), such that any piece can be reconstructed from any k=3 pieces in N, but it needs the four pieces to recover D. If a node fails, the DSS needs to conduct two processes, namely, regenerate the piece and resume the primary data transmission.

The recovery process is described in Figure 4. Initially, *D* collects information from p_1, p_2, p_3, and p_4, but imagine p_4 fails; then the system needs to identify and locate the other pieces (p_1, p_2, and p_3), establish the corresponding recovery links (marked with dashed lines) and transfer the recovery data from the surviving pieces to a new node (p_4'). Once the transmission of the regenerated piece is over, the newly created piece (p_4') will need to continue the primary data transmission, to do so, the system needs to identify the location of the new piece, establish the new link (blue line), and resume.
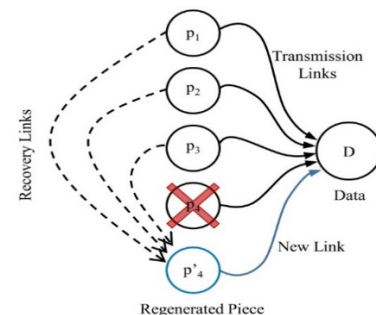
Figure 4: Sample process in DSS

As observed, even in this small example, various processes occur from the networking point of view. Needless to say that the efficiency of the DSS will depend on how fast a node can be recovered and how efficiently uses the resources.
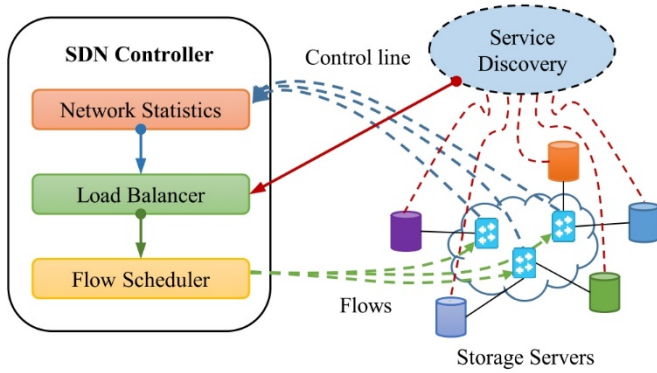
Figure 5: Overview of the Proposed Method

3.2 Overview of the Proposal

To solve the target issues, we describe the proposed SDN-based Network Control Method for DSS. The overall scheme is depicted in Figure 5. As observed, the architecture consists of a centralized SDN controller that is comprised by three internal modules: *Network Statistics*, *Load Balancer*, and *Flow Scheduler*; additionally, an external module (*Service Discovery*) interacts with the controller and the storage servers connected in the underlying network. The role of each of these modules is described as follows:

- *Service Discovery* - this module is in charge of tracking the status of all the storage servers. This information includes the physical load of each of the servers (e.g., memory, CPU, number of processes being served), and the specific DSS configuration (i.e., the number of pieces and their location). The interaction with the SDN Controller is direct, and the status is sent when requested.

- *Network Statistics* - periodically collects network statistics of the entire network, i.e., network topology changes, transmitted and received packets per port. This polling process will allow having an overview of the whole network infrastructure before calculating the appropriate paths. Although the polling happens periodically, it can also be triggered directly by request. We use this information to calculate parameters such as the available bandwidth (1)

$$b_{i,j} = \frac{(Rx+Tx)_{current} - (Rx+Tx)_{previous}}{\Delta t} \qquad (1)$$

- *Load Balancer* - this is the central module in charge of calculating the paths based on the information collected by the *Service Discovery* and *Network Statistics* modules. The primary goal of this module is to distribute the traffic among the available paths based on not only the state of the network but also the load of the storages containing the required pieces. The overall process is subdivided into three phases, namely Discovery, Load Balancing, and
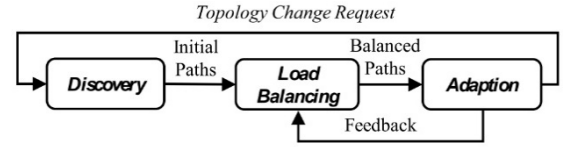
Figure 6: Load Balancing Steps

Adaption (as depicted in Figure 6). In brief, once the request is sent from a source, in the Discovery phase, this module calculates the available paths from source to destination with a process we call *on-demand inverse multiplexing* (described in Section 3.3). Then, based on the current status of both network and storage servers, in the Load-balancing phase, a process we call *hybrid Multipath Load-balancing* assigns the paths that best suit the request. Finally, in the Adaption phase, once the transmission has started the status is updated periodically in case better routes become available or if any change occurred in the topology. The main load-balancing procedure is described in Algorithm 1, where a control loop checks if there has been any change in the topology within a fixed period. As observed, the function *adjustWeights()* in line 8, requests an update from the *Network Statistics* module, and in line 9 the function *balanceTraffic()* is in charge of performing the load-balancing.

- *Flow Scheduler* - once the most suitable paths have been selected, this module writes the flow rules into the specific network devices that take part in the transmission process.

Algorithm 1: Topology Monitor

1:   function monitorTopology ();
2:   *startTimer(t);*
5:   do
6:       if *hasTimeElapsed()* then
7:           if *hasTopologyChanged()* then
8:               *adjustWeights(),*
9:               *balanceTraffic(client, Servers, NW),*
10:          else *restartTimer()* then
11:          end
12:      else *addTimeSpan();*
13:      end
14: While *true;*
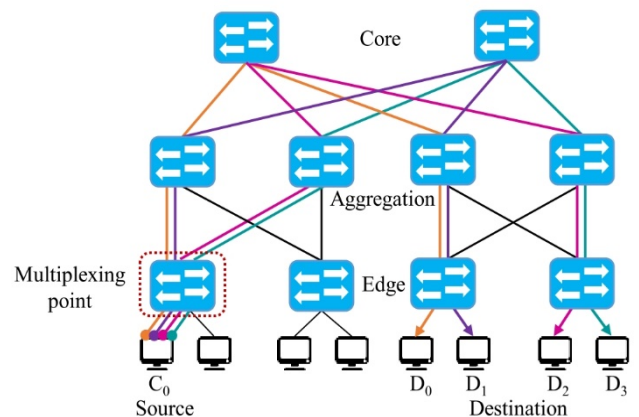
Figure 7: On-demand Inverse Multiplexing

3.3 On-demand Inverse Multiplexing

As mentioned in the previous subsection, the initial step towards an effective network control is the path discovery and traffic distribution. In this section, we describe a process we called *On-demand Inverse Multiplexing*. Initially, it is worth noting that *inverse multiplexing* was explored in the past [22] and mainly used for bandwidth aggregation. It refers to the simple process of sending traffic over multiple paths. In essence, our particular contribution is the idea of distributing the traffic among a specific number of paths (k), which depend on the requirements in a particular DSS configuration. However, in a typical DCN deployment only single links are assigned from device-to-device; moreover, connections at the last layer in the topology are usually low-speed connections, which create a phenomenon we call a *last-mile bottleneck*. To solve this problem, we propose to augment the topology by using parallel connections from the layer closest to the end-devices (edge layer) to the next hop (aggregation layer). This topology augmentation is central for our overall proposal and would involve the overprovision of links according to the particular configuration of the DSS, i.e., setup *n* links if the system split data in a maximum of *n* pieces. Although it is not a common suggestion, organizations still prefer to use various cheap links instead of expensive limited services, such as MPLS or single fat-line [23] due to the cost constraints. Note that, by performing this augmentation, the power consumption of the involved network devices will increase proportionally to the number of parallel connections, therefore, it is necessary to think of a strategy that is capable of reaching a tradeoff between the performance and the power consumptions, however, such a strategy is out of the scope of this paper.

To illustrate the implementation of the on-demand inverse multiplexing process, let us consider the scenario depicted in Figure 7, in this simple deployment, a Fat-tree fully-connected topology comprises three layers (Edge, Aggregation, and Core). Furthermore, assume the particular DSS uses four pieces out of 5 to regenerate as the maximum level of division. In this case, the number of links that the network needs to provision from the edge to the aggregation layer is four. Of course, the augmentation need not be among each device, but only in those particular segments wherein the requirements are higher and can be configured statically or calculated on-demand based on the configuration in the DSS gathered by the *Service Discovery* module.

Once we have the infrastructure prepared, the remaining process works as follows. Consider again the scenario depicted in Figure 7, where we emulate the data recovery which will be allocated in a storage server (C_0), the pieces needed for recovery are located in different servers (D_0, D_1, D_2, D_3). Therefore, in this simple example, the number of parallel links needed are k=4 paths. When the request arrives at the multiplexing point (the closest Top of Rack TOR switch), the controller calculates the appropriate number of paths (*k*) based on the information at the *Service Discovery* module and the current state of the network captured by the *Network Statistics* module.

To solve the initial path discovery, we calculate the Maximum Disjoint Paths (MDP) by a process described in Algorithm 2. This process, which is the initial step of this work and was introduced in [1], calculates the MDPs by an adapted version of the Suurballe's [24] algorithm which calculates the path candidates. Note

that we use a multigraph $G = (V, E)$, where V is the set of network devices (switches), and E is the set of links. Since it is a multigraph, more than a single link can connect two nodes in the set V. We assume that the bandwidth (b_i) for each edge is symmetric, which means that both the uplink and the downlink are the same. Moreover, each edge has a cost $\mu_{u,v}$ as shown in (2), where B is the set of all individual bandwidths b_i. This value is updated periodically by the *Network Statistics* module.

$$\mu_{u,v} = \frac{b_i}{max(B)} \qquad (2)$$

Algorithm 2: Path selection algorithm to find the kmax disjoint paths from source to destination

---

1: <u>function selectKPaths</u> (s, t, k, G);
     Input : *s, t, k, G(the Vthe ,E)*
     Output: Set of *k* path candidates *P*
2: $P \leftarrow \emptyset;$
3: *currentPath* $\leftarrow \emptyset;$
4: *nPath* $\leftarrow 1;$
5: do
6:      if *nPath* > 1 then
7:          *adjustWeights (P[nPath-2]),*
8:      end
9:      *currentPath* $\leftarrow$ *getDijkstraShortestPath(s,t)*
10:     if *currentPath* $\neq \emptyset$ then
11:         *nPath++;*
12:         *P.add(currentPath);*
13:     end
14: While *currentPath* $\neq \emptyset$ and *nPath* $\leq k$
15: return *P*

3.4 Hybrid Multipath Load-balancing

This part of the paper was partially presented in a previous work [25]. Usually, load-balancing techniques focus on either server or network load. In the proposed method, we consider both variables to balance the traffic, and that is why we call it *hybrid*. Moreover, since the topology was augmented with parallel links, it is necessary to take into account multiple paths to balance the traffic. The main procedure is described in Algorithm 3. Initially, from a pool of servers *S*, the *lookupServers* function (line 2) searches for the server candidates which can provide the required service (e.g., in the case of a regeneration process in DSS, it locates the nodes containing the pieces necessary to regenerate the current piece), which are then ordered based on the load of the node. In line 8, we search *k* alternative paths from the source *c* to each of the server candidates *s* that can provide the service. In line 9, the function *assignPaths* calculates the minimum combination of server and path cost. It is worth noting that, based on initial experimentation, the difference in overall cost from all the paths should be less than 25% otherwise the transmission will be delayed in the paths that have costs with the higher difference. This process will guaranty that the distribution is homogeneous in both, the network and servers. Finally, once the destination servers have been identified and the paths selected, the function *writeFlows* in line 11 will send the balanced paths to the Flow Scheduler module, which sends the instructions to the network devices involved in the path.

Algorithm 3: Hybrid Load Balancing Algorithm to calculate the best paths
1: <u>function balanceTraffic (c, S, G);</u>
Input : *Source, set of Servers, G(V,E)*
2:   *serverCandidates ← lookupServers(c,S);*
5:   if *isNotEmpty(serverCandidates)* then
6:       foreach s in *serverCandidates* do
7:         k← *s.lenght*
8:         *Candidates ← selectKPaths(c,s,k,G)*
9:         *BalancedPaths←assignPaths(c,Candidates),*
10:      end
11:     *writeFlows (BalancedPaths);*
14: end

To illustrate the selection process, consider the contrived topology depicted in Figure 8. In this particular example, the features of the system are as follows:

- C_0 is a newly created node that needs to restore the data, to do so, it requires four pieces (p_1, p_2, p_3, and p_4)

- From the edge layer, there are two connections to the aggregation layer

- All links in the topology have weight=1

- None of the servers have any activity yet, and therefore the load is 0%.

- The storage servers that contain the replicas, among others, C_0 are S_1, S_2, S_3, S_4

- Each request consumes 25% of the server resources, and 100% of the link capacity, and therefore only one of the request can be served per path

When the recovery process starts, the *lookupServers* function will identify that S_1-S_4 have the pieces and their location, and all the candidates will be generated in pairs {(C_0, S_1), (C_0, S_2), (C_0, S_3), (C_0, S_4)} and since there are four pieces (p_1, p_2, p_3, and p_4) the function *selectKPaths* will search for four paths between each pair. Then the selection process in function *assignPaths* will start assigning the appropriate paths. For example, p_1 in C_0 to S_1, via the path: $a \rightarrow b \rightarrow e \rightarrow h \rightarrow g$, adjust the weights and the load percentages and continue the process recursively until all the pieces have a service provider; the final result will look as shown in Figure 9, in which all the servers have a perfect balance, and the network traffic is evenly distributed among the available paths. Of course, this is not the unique process, neither are the request homogeneous, therefore, in case another process starts, i.e., from sources *X*, *Y*, or *Z* the values will vary in the service providers and the network load.

Note that this is just the initial assignment, but in case there is a change in the topology (e.g., a link disconnection, or a network device fails), the whole process needs to adapt and restart the process. In the proposed method, the adaption can happen proactively and reactively. In the first case, as shown in Algorithm 1 in Section 3.2, a timer determines how frequent the system needs to control the changes. Although fail-tolerance is not the focus of this paper, the value needs to have a trade-off between performance and
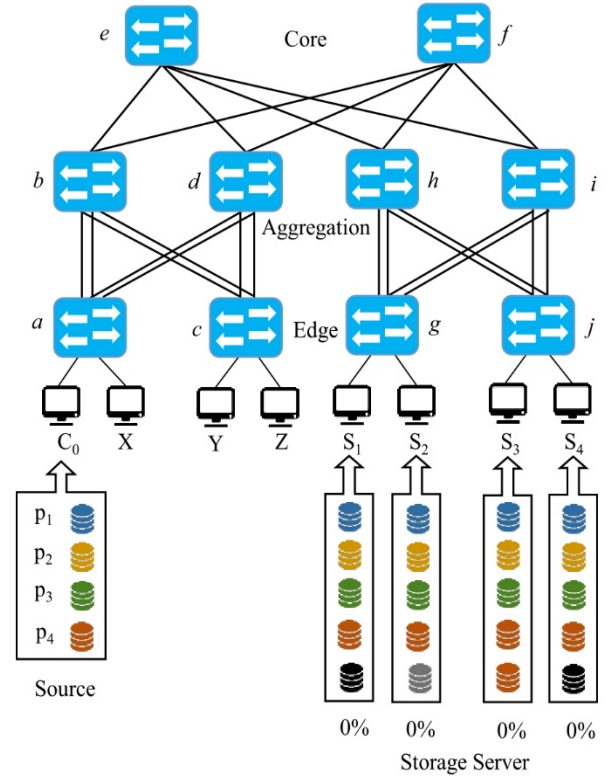
Figure 8: Sample hybrid load balancing (Initial state)
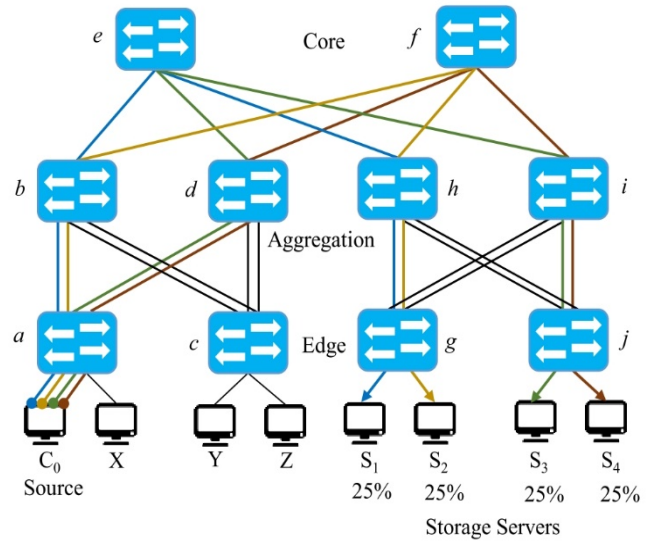
Figure 9: Sample hybrid load balancing (Final state)

service resilience, since having a small refresh time in the order of milliseconds would offer better service resilience at the cost of performance due to the constant polling. In case of reactive adaption, which is an event-based procedure triggered by any change in the topology during the transmission, the system needs to recalculate all the paths in the affected segments.

3.5 Implementation

Based on the overall scheme presented in the previous sub-sections, we implemented the proposed approach using a commonly

used SDN controller OpenDaylight[1] Belirium-SR3 (ODL), OF 1.3 as the communication protocol, and OVS as the back-end deployment.
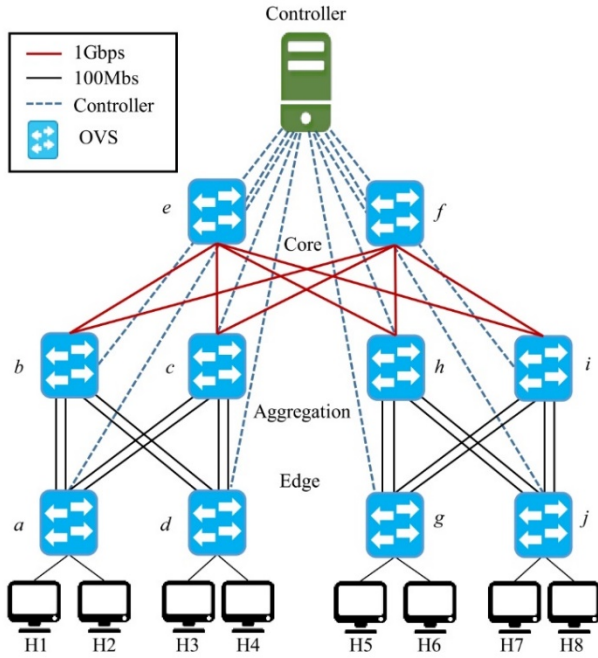
Figure 10: Testbed Experiments

Each of the modules was implemented as part of the controller as follows. The *Service Discovery* module provides a global record of servers' status so that every time a new request arrives at the controller it verifies and updates the status of all the servers involved, and when the transmission is over the status is once again updated. This module could keep track of various parameters, but for simplicity in the current implementation, we only use the following attributes.

ServerID, MaxLoad, Ports[ID, CurrentRequest, Location]

MaxLoad is the maximum number of requests that the server can handle, which we then use to calculate the server load based on the individual requests, and finally the parameter *Ports* is an array of ports that represent a different service, note that for each of the ports we store the current state and where they can be located.

The *Network Statistics* module, collects the network variables every 10 seconds in the current implementation, although the polling time can be configured directly in the controller if the time is too short; correspondingly, the number of control messages increases. The initial topology discovery is conducted by *L2Switch*, which is a feature available in ODL that handles, among others, the ARP handling, host tracking, and so forth. However, once the initial connectivity is ensured, all path decisions will be made by the *Load Balancer* module.

Finally, the *Load Balancer* is integrated a separate feature and is triggered every time a new request arrives at the controller. Once

the paths are selected, the *Flow writer* module will send the *flow_mod* messages to the devices and set up the proper flow rules with an expiration time equal to the refreshing time, so that in case the flows are not in use in a cycle they will be deleted from the table.

4. Evaluation

4.1. Overview

To evaluate the proposed approach, we used an emulated environment created using Mininet[2] v.2.2.2, ODL Beryllium-SR3 as the controller, *iperf*[3] to create the network traffic, and Wireshark to analyze the results. Initially, we explored the correctness and behavior in controlled conditions in which the only traffic is the stream being tested. Then, we performed measurements on environments where other transmissions are happening concurrently and compare the performance with existing solutions.

4.2. Transmission without Background Traffic

In this section, we describe a set of experiments that demonstrate the correctness and effectiveness in conditions where there is no background noise, which means that no other transmissions are happening at the same time. Figure 10 depicts the evaluation

(a) Average throughput one-to-one test

(b) Average throughput one-to-many test
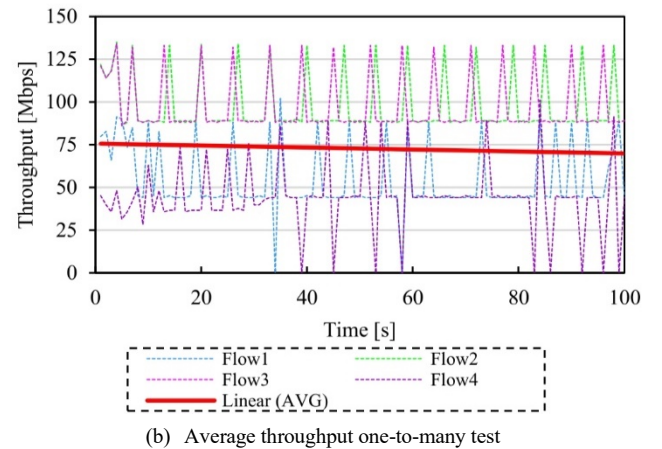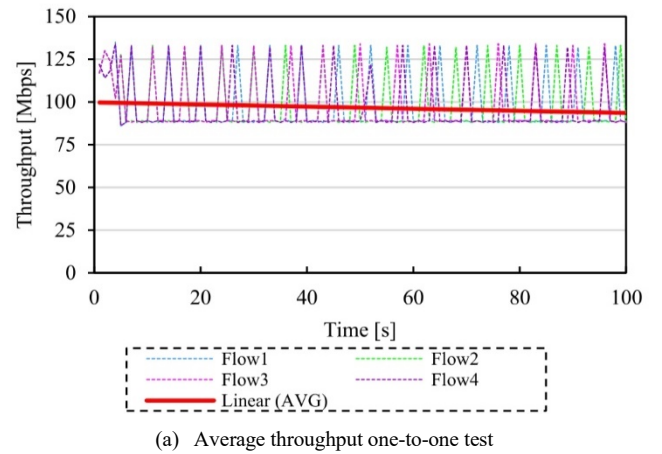
Figure 11: Results average throughput

---

[1] https://www.opendaylight.org/

[2] http://mininet.org/

[3] https://iperf.fr/
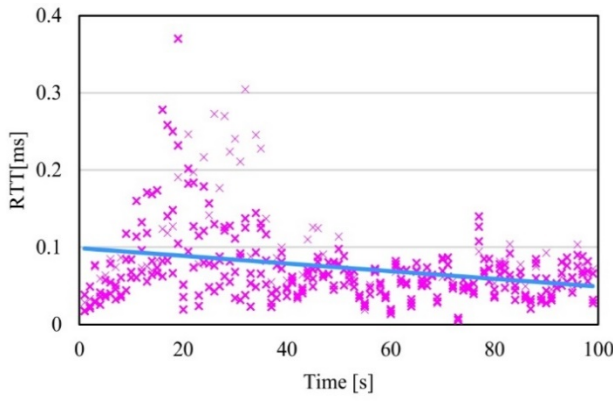
(a) Average RTT one-to-one test
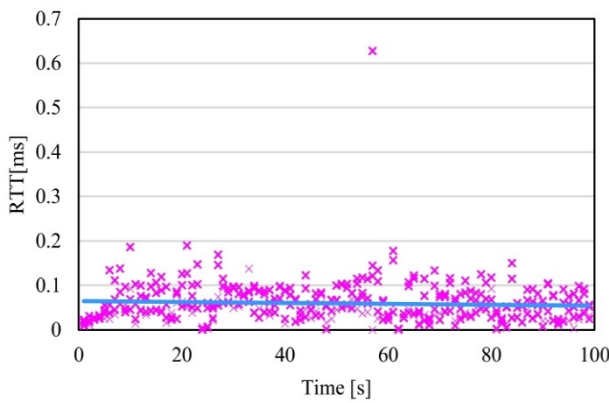
(b) Average RTT one-to-many test

Figure 12: Results average RTT

setup of the testbed used in the evaluation. As observed, a DCN is deployed in a fully connected Fat-Tree-based topology with eight servers, each containing four pieces represented as TCP ports. Each of the links connecting the *edge* to the *aggregation* layer was configured with 100Mbps, whereas the links connecting the *aggregation* to the *core* layer were set up at 1Gbps. Note that the topology was augmented by four parallel links which connect the devices from the edge layer as the number of pieces is also four, each of them will be named as the assigned letter and a subscript index based on the connection port, e.g., the link a_1 means OVS *a* port 1. Also, the SDN application and the emulated network were hosted in the same Virtual Machine using Ubuntu 16.04 LTS, with two 2.60GHz CPUs and 4 GB of memory.

- One-to-One Transmission Test

The goal of this first experiment is to test the correctness of the proposal when the traffic distribution is continuous from a single source to a single destination. In DSS, this case applies when peer-to-peer storages share the same information as simple replication. To conduct this experiment, we used *iperf* to send a continuous TCP stream to four different TCP ports from H1 to H8 for 100s. Note that just for testing purposes no delay was set up for any of the links in mininet.

Initially, the path discovery assigned the following paths $\{[a_1 \rightarrow b_1 \rightarrow e \rightarrow h_3 \rightarrow j_3], [a_2 \rightarrow b_2 \rightarrow f \rightarrow i_3 \rightarrow j_3], [a_3 \rightarrow c_1 \rightarrow e \rightarrow h_4 \rightarrow$

$j_4], [a_4 \rightarrow c_2 \rightarrow f \rightarrow i_4 \rightarrow j_4]\}$, which is the ideal distribution this particular case. Figure 11a shows the throughput achieved by each of the flows, note the linear trend (in red) is within the best theoretical threshold for each link. Then, we also measured the average RTT of the entire stream, the results are shown in Figure 12a which is relatively high at the beginning of the stream as there is a single destination, but it is still within the boundaries of the standard parameters.

Table 1 shows a summary of the obtained results. From all the flows a total of 4.5Gbytes were transmitted in the 100s period, and the average of throughput was 96.6 Mbps with an average overall RTT of 0.07ms as expected since most of the lines were expedited.

- One-to-many Transmission Test

In this experiment, the continuous stream was sent from a single source to multiple destinations. In DSS, this case applies when a recovering a failed node or when reconstructing the data. To conduct this experiment, we sent an *iperf* request from H1 to a different TCP port in H5-H8 for 100s. A summary of the obtained results is described in Table 2. The discovered paths were $\{[a_1 \rightarrow b_1 \rightarrow e \rightarrow h_1 \rightarrow g_1]$ for H1, $[a_3 \rightarrow c_1 \rightarrow f \rightarrow i_1 \rightarrow g_3]$ for H2, $[a_4 \rightarrow c_2 \rightarrow e \rightarrow h_3 \rightarrow j_1]$ for H7, and $[a_1 \rightarrow b_1 \rightarrow f \rightarrow i_4 \rightarrow j_4]$ for H8$\}$, note that there is an overlap in the first segment of the paths which affected the overall throughput, as shown in Figure 11b, note the linear trend (in red) achieved 25% less than in the previous case. After tracing the costs of the paths, we found out that there were two paths with the same cost and the algorithm assigned the one first discovered, which shared common segments. Nonetheless, the loss in throughput and the amount of data transferred (see Table 2) was compensated with a more homogeneous overall RTT, as seen in Figure 12b.

Table 1: Results Experiment 1 (one-to-one)

Flow#	Data [Gb]	AVG TP [Mbps]	AVG RTT[ms]
Flow 1	1.13	96.6	0.09
Flow 2	1.12	96.7	0.05
Flow 3	1.13	96.6	0.07
Flow 4	1.13	96.7	0.07

Table 2: Results Experiment 2 (one-to-many)

Flow#	Data [Gb]	AVG TP [Mbps]	AVG RTT[ms]
Flow 1	0.65	53.0	0.04
Flow 2	1.13	96.5	0.08
Flow 3	1.12	96.5	0.08
Flow 4	0.52	42.8	0.04

- Path Adaption Test

In this experiment, we test the resilience of the service by sending a continuous *iperf* request to a single server for 60*s* from H1 to H8, and after a random period, one of the links was shut down, Table 3 shows the results obtained. As observed, in the 60s a total of 624Mbytes were transmitted, with an average throughput of 85.6Mbps. Moreover, Figure 13 depicts the transmission trend, which was relatively stable. The initial path was calculated as follows $a_1 \rightarrow b \rightarrow e \rightarrow h_4 \rightarrow j_2$, then the link $h_4 \rightarrow j_2$ failed and recovered later on to a new path $a_4 \rightarrow c \rightarrow f \rightarrow i_4 \rightarrow j_4$, the outage time (shown in the red shaded part in the figure) was approximately 5s. Although the time could be reduced if only part of the path was modified, in practice it is more efficient to recalculate the whole path than look for the part that is affected as the time to add a flow is much lower
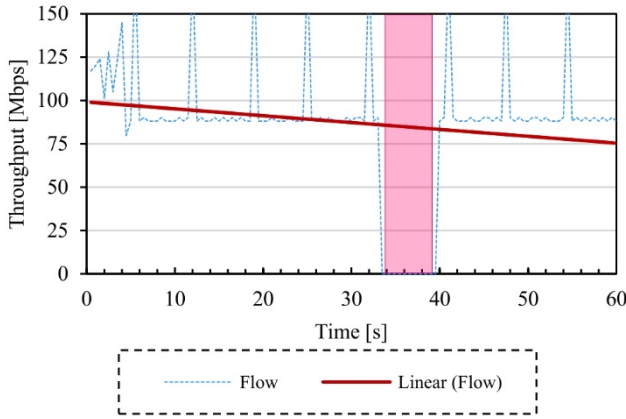
(a) Average throughput single source and destination
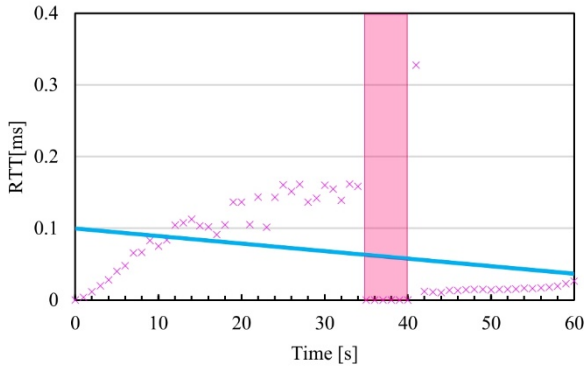
(b) Average RTT single source and destination

Figure 13: Results path adaption test
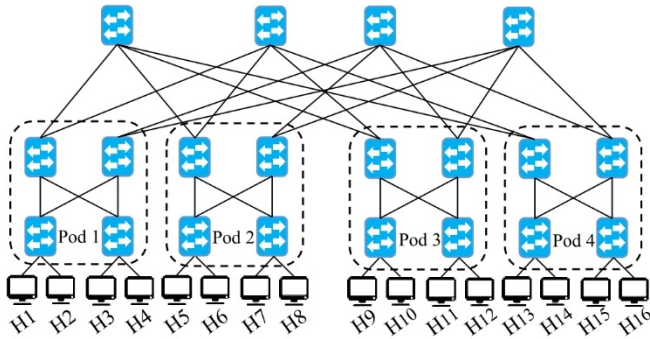
Figure 14: Evaluation testbed 4-ary Fat-tree topology

than when it is updated [26]. Moreover, since the other flows will not have activity anymore, they will be removed in the next refresh cycle.

4.3. Transmission with Background Traffic

In this section, we compare the approach with existing related work. Figure 14 shows the evaluation environment, which consists of a k-ary Fat-tree topology with K=4 pods. Eight servers (Pod 3 and Pod 4) will provide four storages represented as TCP ports. Without losing generality, for testing purposes, all of the links were

configured to 100Mbps with a *max_queue_size* of 1000. Note that the topology was not augmented for this test, as the traditional techniques and the existing solutions could not handle the increased amount of redundancy, as the percentage of loss and drop packets was high. However, in the comparison, we also include the results in the case where the topology was augmented.

Table 3: Results Experiment 3 (one-to-many)

Flow#	Data [Mbytes]	AVG Throughput [Mbps]	AVG RTT [ms]
Flow 1	624	85.7	0.06

To benchmark the approach, we compared the results with the following:

- Single path; the deployment was entirely handled by the L2Switch project in ODL, which uses the STP protocol and Dijkstra's algorithm to calculate the shortest path for full connectivity.
- ECMP; An implementation of ECMP using a modified version of a publicly available code[4], that uses OF *group rules* to switch traffic when there are multiple paths.
- MPSDN [17]; the code is also publicly available[5], both the controller (Ryu[6]) and the mininet were hosted in the same virtual machine using Ubuntu 14.04 LTS (as the modified version of OVS was initially designed for a particular version of the kernel). We selected this work, as they proved that for certain conditions their performance equals the one of MPTCP, and thus, we implicitly benchmark our approach against this experimental protocol as well.
- Our approach; The deployment for our approach without modifying the topology, which means that no extra links were setup between the *edge* and *aggregation* layer in the topology. Moreover, the initial paths remained the same for the entire transmission, with no adaption.
- Extended version of our approach; For the extended version of the proposed approach, we modified the original Three-layer non-blocking FTN by adding an extra parallel link from the *edge* to the *aggregation* layer, and the paths were adapted after every 10s (refresh cycle) based on the information given by the *Network Statistics* module.

In the benchmark test, each host in Pod 3 and Pod 4 were listening to TCP ports {6001, 6002, 6003, and 6004}. Moreover, random UDP background traffic continuously sent from all host in Pod 1 and Pod 2 to all hosts in pods 3 and 4, i.e., H1 to H9, H2 to H10, H3 to H11 and so forth. We used continuous UDP *iperf* requests with bandwidths ranging from 1 to 10Mbps using the default paths and values. Then a request consisting of four parallel 100Mbps TCP *iperf* of were sent from both H1 and H5 (recovery nodes) which will be handled by the servers in Pod 3 and Pod 4 respectively. Therefore, the total amount of data sent was

---

[4] https://github.com/Huangmachi/ECMP/blob/master/fattree4_ecmp.py
[5] https://github.com/dariobanfi/multipath-sdn-controller
[6] https://osrg.github.io/ryu/

800Mbps, and we measured the time in which all the transmissions finished.

- *Completion time*

The first metric we measured was the completion time. The results are depicted in Figure 15; as can be seen, the slowest solution was the single path approach, which completed all the data transferences in 275 seconds with a standard deviation of 16s. By contrast, the fastest one was the proposed approach in both of the cases, with and without topology augmentation, completing all the transference in a total of 30s with a standard deviation of 7.8s. In the case of the extended version of the proposal, the standard deviation was only of 4.8s which is the lowest among the benchmarked solutions. Note that ECMP completion time is also relatively low, 36.5s was the completion time of the last flow with a standard deviation of 6.4s. Finally, note that although MPSDN had a long completion time, 114s for the last flow, the variation of flow arrival was slightly higher than ours, which is due to the condition they use to ensure paths with compatible delays.
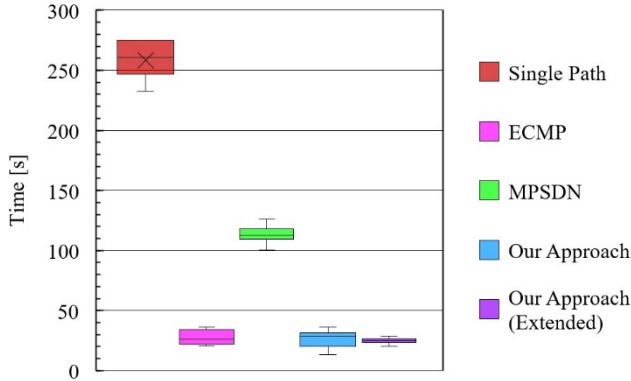
Figure 15: Completion time of 800Mbytes in a 4-ary FTN

- *Throughput*

The second metric we measured was the average throughput per flow. As shown in Figure 16, MPSDN and the simple path approach showed a steady average throughput in all the flows. However, the overall maximum performance was low, which affected the completion time. In the other cases, even though the distribution was not as regular, the overall performance was much higher. Also, it is also worth noting that since the UDP traffic was sent at random bandwidths, some of the paths were more saturated than others, which might have caused the irregular distribution. Nevertheless, as can be seen in Figure 17, in our approach, the aggregated throughput was much higher than the other approaches. In the case of the augmented version, consider that although the number of links was duplicated from the edge to the aggregation layer, the increase was not linear.

- *Network Load*

The last variable we measured was the overall network load. Although the proposed approach can handle the dynamic load balance of servers and network at the same time as shown in [25],
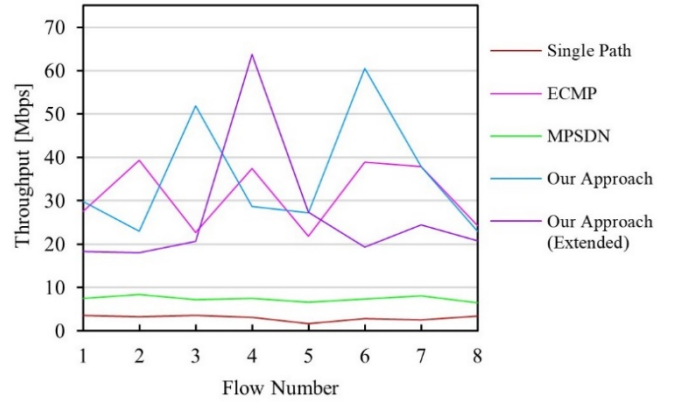
Figure 16: Average throughput per flow
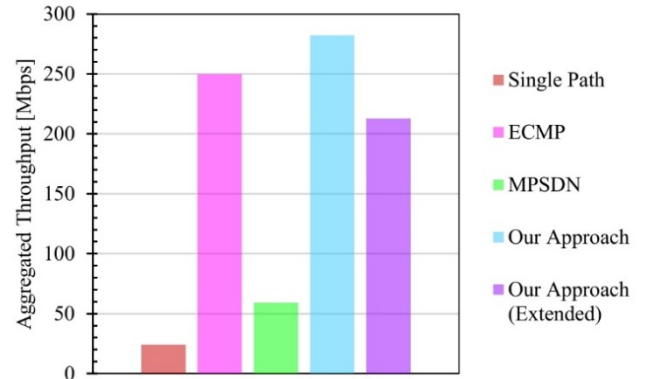
Figure 17: Aggregated throughput at eight 100Mbps flows

since the existing solutions focus mainly on the network load balancing we measured only that parameter. To perform this calculation, we used Tang's [9] formula to calculate the network bandwidth utilization ratio $\overline{\lambda(t)}$, which gives the total bandwidth utilization in a time t and is expressed as in (3)

$$\overline{\lambda(t)} = \frac{\sum_{1 \leq i,j \leq N} \lambda_{i,j}(t)}{N} \qquad (3)$$

Where N is the number of links, and $\lambda_{i,j}$ is the link bandwidth utilization ratio defined as in (4), where $b_{i,j}$ is the used bandwidth and $B_{i,j}$ is the capacity of the link.

$$\lambda_{i,j}(t) = \frac{b_{i,j}}{B_{i,j}} \qquad (4)$$

To make a fair comparison, we only measured the time until the time the first flow finished, which was about 30s from the beginning of the experiment. Figure 18 shows the Normalized the Cumulative Load over those 30s. As observed, ECMP was the solution that generated the most load over that time and across the whole network. However, the proposed approach used slightly more than the load of the single path solution with the added value of having a faster completion time and better overall throughput. Note that in our case, the load was not concentrated in a specific part of the network as in the other solutions, but spread among the

discovered paths. Moreover, the results for the augmented version of the topology shows better traffic distribution, which will allow more concurrent operations to be performed in less time, a condition that is desirable for DSS.
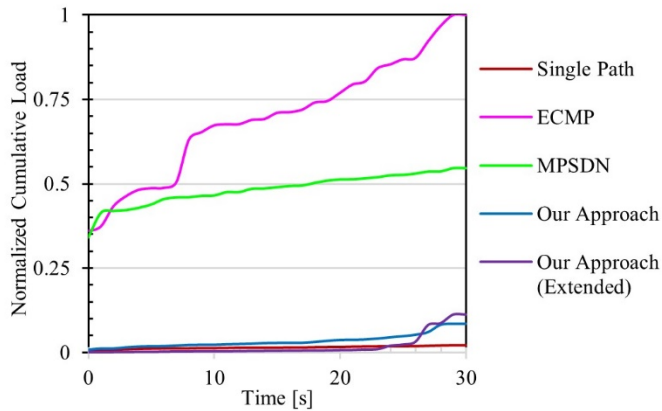
Figure 18: Normalized Cumulative Network Load in 30s

Finally, Table 4 shows a summary of the results per flow. As observed, the proposed approach outperformed the others in most of the cases. Moreover, the overall load of the entire network is compared to the one on a single path, but the traffic is more evenly distributed. Note that the network load in ECMP and MPSDN per flow is relatively high, in the case of MPSDN, it might be due to the high amount of control messages sent by the controller, but in the case of ECMP, the use of multiple paths caused several retransmissions due to packet loss.

Table 4: Average benchmark results per flow

Metric	Single Path	ECMP	MPSDN	Our App.	Our App. (extended)
Completion Time [s]	256.5	27.5	113.6	26	24.4
Throughput [Mbps]	3.02	31.24	7.41	35.28	26.61
Network Utilization [%]	0.05%	2.06%	1.45%	0.11%	0.05%

5. Conclusions

Due to the restrictiveness of using legacy network techniques for DSS, it may not be possible to cope with the ever-increasing user content-generation in the future. Therefore, in this paper, we presented a network control method to improve DSS performance by using SDN. The proposed approach used pragmatic and straightforward solutions to handle DSS's main requirements, namely aggregated bandwidth and effective use of resources. We have evaluated the proposal in various scenarios, and compared it with both network legacy techniques and existing solutions. Experimental results show that our method could achieve faster data transference and maintain balanced network and server loads. Additionally, in contrast to other proposals, our method does not re-

quire protocol or end-point modification but instead a topology enhancement based on the system demands. We showed that the overall performance of routinely DSS tasks can be dramatically increase by augmenting the DCN using parallel links, a proper path discovery, and dynamic adaption based on the network and servers load. As a future work, we still have to test the solution in large-scale and real DSS deployments. Furthermore, with the rise of new multipath transport protocols, such as QUIC [27] it might be interesting to study the impact on the way traffic is handled in non-traditional transport protocols in future networks. Likewise, we have not fully exploited the adaptive aspect of the proposal, which can be used to solve problems concerning resilient and fail-tolerant networks. Finally, it is also important to find a mechanism that achieves a tradeoff between high-performance and energy-efficiency of the whole network to cope with the proposed overprovisioned infrastructure.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] L. Guillen, S. Izumi, T. Abe, T. Suganuma, H. Muraoka, "SDN implementation of multipath discovery to improve network performance in distributed storage systems" in 13th International Conference on Network and Service Management, Japan, 2017. https://doi.org/10.23919/CNSM.2017.8256054

[2] D. Reinsel, J. Gantz, J. Rydning, "Data Age 2025: The evolution of data to life-critical - Don't focus on big data; focus on the data that's big", IDC White Paper, 1–25, 2017.

[3] A. Dimakis, P. Godfrey, Y. Wu, M. Wainwright, K. Ramchandran, "Network coding for distributed storage systems" IEEE Transactions on Information Theory, **56**(9), 4539–4551. 2010. https://doi.org/10.1109/TIT.2010.2054295

[4] C. Suh, K. Ramchandran, "Exact-Repair MDS Code Construction Using Interference Alignment" in IEEE Transactions on Information Theory, **57**(3), 1425–1442, 2011. https://doi.org/10.1109/TIT.2011.2105003

[5] N. B. Shah, K. V. Rashmi, P. V. Kumar, K. Ramchandran, "Distributed storage codes with repair-by-transfer and nonachievability of interior points on the storage-bandwidth Tradeoff" in IEEE Transactions on Information Theory, **58**(3), 1837–1852, 2012. https://doi.org/10.1109/TIT.2011.2173792

[6] Open Networking Foundation (ONF) Software-Defined Networking (SDN) Definition Available at https://www.opennetworking.org/sdn-definition/ Accessed on 2018-07-20

[7] S. Kaneko, T. Nakamura, H. Kamei, and H. Muraoka, "A Guideline for Data Placement in Heterogeneous Distributed Storage Systems" in 5th International Congress on Advanced Applied Informatics, Kumamoto Japan, 942–945, 2016. https://doi.org/10.1109/IIAI-AAI.2016.162

[8] IEEE Std. 802.3ad "IEEE Standard for Local and metropolitan area networks - Link Aggregation," available at http://www.ieee802.org/3/ad/ Accessed on 2018-06-10

[9] F. Tang, L. Yang, C. Tang, J. Li, M. Guo, "A Dynamical and Load-Balanced Flow Scheduling Approach for Big Data Centers in Clouds" IEEE Transactions on Cloud Computing, 2016. https://doi.org/10.1109/TCC.2016.2543722

[10] IETF RFC2992 "Analysis of an Equal-Cost Multi-Path Algorithm" available at https://tools.ietf.org/html/rfc2992, Accessed on 2018-01-30

[11] L. Li, Q. Xu, "Load Balancing Researches in SDN: A Survey," in 7th IEEE Int. Conference on Electronics Information and Emergency Communication, 403–408, Macau China, 2017. https://doi.org/10.1109/ICEIEC.2017.8076592

[12] M. Al-Fares, S. Radhakrishnan, B. Raghavan, N. Huang, A. Vahdat, "Hedera: Dynamic flow scheduling for data center networks," In 7th Conference on Networked Systems Design and Implementation, 19–34, San Jose CA USA, 2010. http://dl.acm.org/citation.cfm?id=1855711.1855730

[13] R. van der Pol, M. Bredel, A. Barczyk, B. Overeinder, N. van Adrichem, O. Kuipers, "Experiences with MPTCP in an Intercontinental OpenFlow Network," In 29th Trans European Research and Education Networking Conference,1–8, Amsterdam The Netherlands, 2013.

[14] IETF RFC 6182 "Architectural Guidelines for Multipath TCP Development," available at https://datatracker.ietf.org/doc/rfc6182/, Accessed on 2017-06-15

[15] J.P. Sheu, L. Liu, R.B. Jagadeesha, Y. Chang, "An Efficient Multipath Routing Algorithm for Multipath TCP in Software-Defined Networks," in European Conference Networks and Communications, 1–6, Athens Greece, 2016. https://doi.org/10.1109/EuCNC.2016.7561065

[16] D. J. Kalpana, K. Kataoka, "SFO: SubFlow Optimizer for MPTCP in SDN," in 26th Int. Telecommunication Networks and Applications Conference, Dunedin New Zealand, 2016. https://doi.org/10.1109/ATNAC.2016.7878804

[17] D. Banfi, O. Mehani, G. Jourjon, L. Schwaighofer, R. Holz, "Endpoint-transparent Multipath Transport with Software-defined Networks," in IEEE 41st Conference on Local Computer Networks, 307-315, Dubai UAE, 2016. https://doi.org/10.1109/LCN.2016.29

[18] F. Carpio, A. Engelmann, A. Jukan, "DiffFlow: Differentiating Short and Long Flows for Load Balancing in Data Center Networks," in IEEE Global Communications Conference, 1–6, Washington DC USA, 2016. https://doi.org/10.1109/GLOCOM.2016.7841733

[19] Y. Li, D. Pan, "OpenFlow based Load Balancing for Fat-Tree Networks with Multipath Support," In 12th IEEE International Conference on Communications, Budapest Hungary, 2013.

[20] S. Izumi, M. Hata , H. Takahira, M. Soylu, A. Edo,T. Abe and T. Suganuma, "A Proposal of SDN Based Disaster-Aware Smart Routing for Highly-available Information Storage Systems and Its Evaluation," International Journal of Software Science and Computational Intelligence, **9**(1), 68–82, 2017. https://doi.org/10.4018/IJSSCI.2017010105

[21] K.T. Dinh, S. Kuklinski, W. Kujawa, M. Ulaski, "MSDN-TE: Multipath Based Traffic Engineering for SDN," in 8th Asian Conference on Intelligent Information and Database Systems, 630–639, Da Nang, Vietnam 2016. https://doi.org/10.1007/978-3-662-49390-8_61

[22] P.H. Fredette, "The past, present, and future of inverse multiplexing," in IEEE Com. Magazine, **32**(4), 42–46, 1994. https://doi.org/10.1109/35.275334

[23] R. Toghraee, Learning OpenDaylight - The Art of Deploying Successful Networks, Packt Publishing Ltd., 2017.

[24] J. W. Suurballe, R. E. Tarjan "A quick method for finding shortest pairs of disjoint paths," in Networks, **14**(2), 325–336, 1984. https://doi.org/10.1002/net.3230140209

[25] L. Guillen, S. Izumi, T. Abe, T. Suganuma, H. Muraoka "SDN-based hybrid server and link load balancing in multipath distributed storage systems" in 2018 IEEE/IFIP Network Operations and Management Symposium, Taipei Taiwan, 2018. https://doi.org/10.1109/NOMS.2018.8406286

[26] M. Kuźniar, P. Perešíni, D Kostić "What You Need to Know About SDN Flow Tables," in: Mirkovic J., Liu Y. (eds) Passive and Active Measurement. PAM 2015. Lecture Notes in Computer Science, 347–359, 2015. https://doi.org/10.1007/978-3-319-15509-8_26

[27] A. Langley, A. Riddoch, A. Wilk, A. Vicente, C. Krasic, et al "The QUIC Transport Protocol: Design and Internet-Scale Deployment." in Proc. of the ACM Conference of Special Interest Group on Data Communication, 183–196, New York USA, 2017. https://doi.org/10.1145/3098822.3098842

Evolving AL-FEC Application on 5G NGMN-Edge Computing Systems

Christos Bouras[*1,2], Nikolaos Kanakis[2]

[1]*Computer Technology Institute and Press "Diophantus", Patras, Greece*

[2]*Computer Engineering and Informatics Department, University of Patras, Greece*

A B S T R A C T

Fifth generation of mobile networks (5G) comes to cope with the context of the new mobile telecommunications era. Edge computing is a new collaborative technology under standardization utilizing end-user devices or near-end-user edge devices to operate processing, communication, or control operations. The achievement of efficient error control is a very critical aspect in the successful development of 5G Next Generation Mobile Networks (NGMN). Especially utilizing Forward Error Correction (FEC) codes on the application level is an efficient approach on improving error control on NGMN since it is adopted in several mobile multicast standards. FEC is a feedback free error control technique where redundant data are introduced with the source data to enable the recipient recovering from packet losses with an interesting approach on applying AL-FEC error protection introduces deterministic or randomized online algorithms. Based on this we present a novel online scheme on applying AL-FEC application on the context of the AL-FEC policy online problem. In this work we present an online algorithm based on feedback received from mobile end-users which adapts the introduced protection based on this. Another aspect of this work is the utilization of AL-FEC protection on the edge level based on RaptorQ FEC codes.

1 Introduction

The design of an algorithm capable to adapt the transmission redundancy that will be introduced by AL-FEC protection can be addressed with online algorithms. Online algorithms [1] are utilized on problems where the algorithm does not have knowledge of the input in advance. Hence, online algorithms should be able to make a decision without a priori knowledge of the input since it is not available at present. On online algorithms design it is assumed that algorithm's input arrives in steps and the online algorithm has to make a decision upon the arrival of each new input. Moreover, in problems where deterministic algoithms are not suitable or cannot even apply the simplest and several times the most efficient approach is utilizing randomized online algorithms [2]. Regarding the performance of online algorithms it is analyzed utilizing competitive analysis where the output produced by the evaluated online algorithm is compared with the performance defined for the optimal offline algorithm of the problem. The baseline for the concept of the optimal offline algorithm of an online problem is that this algorithm has a priori knowledge of the entire input and thus is able to confront with it with minimum cost. The metric used to describe the performance of an online algorithm A is the competitive ratio. The competitive ratio is always computed against an adversary knowing in advance the operation of the online algorithm and generating a sequence σ, based on this knowledge, with the online algorithm serving it. To compute the competitiveness of an online algorithm A, if $A(\sigma)$ stands for the cost of the online algorithm and $OPT(\sigma)$ stands for the cost of the optimal offline algorithm the online algorithm A is c-competitive if a constant α exists that $A(\sigma) - c \cdot OPT(\sigma) \leq \alpha$.

Edge computing [3] is a virtualization technology providing computation, and networking operations between devices located at the edge of the network deployment and cloud computing. The baseline con-

text of Edge computing is having light-weight devices at the proximity of end users enabling the provision of custom and location-aware services. Such micro clouds can be easily integrated to Enhanced Packet Core (EPC) of mobile networks enabling on EPC new capabilities on deploying evolutionary architectures or service solutions via virtualization or not.

This paper is an extension of work originally presented in [4] focused on the application of AL-FEC error protection on NGMN multicast services where the novel MTO adaptive online algorithm aiming to efficiently apply AL-FEC protection is analyzed. In this work, we moreover concentrate on the migration of AL-FEC protection to the edge of mobile edge networking architecture as presented in [5]. We provide an analysis on such an architecture and the operations required for such integration. Furthermore, we evaluate the AL-FEC application performance on the edge revealing, apart from the benefits introduced from such a choice, the performance bounds on its deployment.

The context of this work has the following structure: In Section 2 we discuss relevant work of this paper. In Section 3 we describe the concept of Mobile Edge computing and in Section 4 we present the AL-FEC reliability control. In Section 5 we present the proposed online algorithm and we analyze its performance. In Section 6 we analyze the migration of the AL-FEC error protection on the edge and we evaluate this design performance under several perspectives. Concluding in Section 7, we discuss the advantages introduced from the proposed online error protection scheme and the AL-FEC application on the edge and we refer to some possible future directions that could follow in order to extend this work.

2   Related Work

Many active research fields of communication networks utilize online algorithms. In the work of [6] the authors cope with routing alogorothms under energy efficiency constraints and they present an online algorithm improving the throughput on multihop networks, The authors of [7] utilize online to address energy-constrained limitations on ad-hoc networks in the context of multicast routing problems. Moreover, the authors of the work [8] provide distributed online algorithms for the frequency assignment online problem. In the work presented in [9] the authors cope again with energy efficiency in conjunction with delay constraints on wireless multicast deployments presenting an algorithm for online scheduling problems. The authors of [10] investigate the decision sending a source packet, retransmiting a lost packet or sending a redundant packet with competitive analysis, providing an online algorithm capable to pick the appropriate decision. Finally, the authors of this manuscript in [11] stated the online AL-FEC policy problem referring to reduction of the efficient selection of the introduced transmission overhead in an AL-FEC protected delivery to online problems for mobile multicast networks

and they provided a first attempt on this newly introduced online problem with a randomized online algorithm. Moreover, the same authors in the work [12] designed a deterministic online algorithm utilizing weight assignments and afterwards the same authors provided an enhanced adaptive version of this deterministic online algorithm in [13].

On the edge computing field, the authors of [14] provide a thorough definition of this techonology, analyzing several case studies and discussing the challenges and opportunities of edge computing. Moreover, the authors of the owrk [15] present the edge technology in the context of sensor networks, peer-to-peer networks and network virtualisation functions. Finally, in the work [16] the authors investigate the faesibility of utilizing constrained devices as edge gateways in edge arihitectures. From the results of this work they conclude that under appropriate configuration and fine-tuning, scalability and reduced overhead can be achieved.

3   Mobile Edge Computing on 5G Multicast NGMN

The increased demands on network capacity and user experience drive the next generation mobile systems. The efficient utilization of the radio spectrum over common resources is one of the most major aspects for broadcast and multicast delivery towards 5G NGMN. Apart from the development of novel technologies there is the path where multicast services will cooperate and be enhanced by several existing ecosystems, like Internet of Things (IoT) or Device-to-Device (D2D) communication. To achieve this cooperation development is required on all layers of the protocol stack, from modulation and coding to transport protocols, and applications.

Mobile Edge Computing (MEC) [3] is a technology under standardization from the ETSI MEC Industry Specification Group (ISG). Mobile Edge Computing describes a service environment with cloud-enabled capabilities located at the edge of the mobile network in proximity to mobile end-users. The main goals of MEC is to achieve reduced latency, increased efficiency for network operation and services delivery, and as a result an enhanced user experience. MEC side by side with Functions Virtualization (NFV) and Software-Defined Networking (SDN)) [17] is one of the main focuses on the development of 5G evolved systems. One of the main advantages of MEC in conjunction with NFV and SDN is that will enable the transformation of traditional deployments to a programmable interface for various operations contributing on the achievement of the demanding requirements of 5G standard. MEC is itself based on virtualization, and comes to complement NFV, while NVF focuses on network functions while, MEC is responsible to enable applications running at the edge of the network. Another advantage of MEC is the fact that allows operators to reuse their infrastructure deployments since the infrastructure

hosting MEC operations is similar with the equired NFV infrastructure hence, it is possibl to host both VNFs (Virtual Network Functions) and MEC applications on the same infrastructure. Combining all of the above, MEC technology will allow 5G environments to provide improved latency, bandwidth, and location awareness and so new or existing services/operations could be hosted on the edge.

4 AL-FEC Error Protection

Fountain Codes are codes developed to provide reliability control on data delivery over unreliable channels characterized by unspecified probability of data corruptions [18]. Fountain code name comes from the fact that operate like water fountains infinite supplying water drops and anyone collecting those drops in a bucket under the fountain with the bucket removed when enough drops are collected. In the same context but with a digital source instead of a water fountain, a recipient receives encoded symbols from one or more transmitters and collects them till enough are obtained. Then the recipient is able to reconstruct the original transmitted object, and which symbols are received does not matter. Since the source can theoretically generate unbounded amount of encoded symbols, Fountain Codes are rateless codes.

Since random linear codes cannot be applied in proactice due to their complexity, Luby Transform (LT) codes presented in [19] which are able to provide applicable encoding and decoding and at the same time they are able to maintain low transmission overhead. In more details, LT codes can operate close enough to channel encoding under appropriate selecting the distributions of the edges in the Tanner graph. Moreover, if the reliability of the decoding process is decreased the complexity can be further reduced providing linear time for encoding and decoding complexity. Since all of the input symbols cannot be decoded with the lower degree distribution for the same overhead if an erasure correcting pre-code is utilized the erasures arising from the weakened decoder could be recovered. Raptor codes provide exceptional encoding and decoding performance if a linear time block code like LDPC is used for the precoding process.

RaptorQ codes [20] are the predecessor of Raptor codes providing much more efficiency on AL-FEC encoding and decoding process. They also present by far improved flexibility and error recovery. Encoding of RaptorQ codes is almost the same with Raptor code encoding. Of course RaptorQ codes introduced an enhanced design providing enhanced performance compared with their predecessor. The main difference on the operation of RaptorQ and Raptor codes is that the newly introduced RaptorQ codes utilize larger finite fields overcoming the performance limitations of Raptor code and achieving recovery from lost symbols with decreased transmission overhead. Apart from this RaptorQ codes are capable to operate over higher number of source symbols and thus increased number of generated encoding symbols. This expansion on the number of symbols RaptorQ codes can handle enables the simplification and increased flexibility of applying AL-FEC protection.

5 Online AL-FEC Protection

5.1 Online Model

In this work we assume a common mobile networks environment where multicast delivery of streaming flows to endusers is simulated. In more details, a multicast source injecting bunches of packets in the network which are handled as a continuous object and are delivered to endusers via multicast bearers encapsulated in RTP/UDP streaming flows.
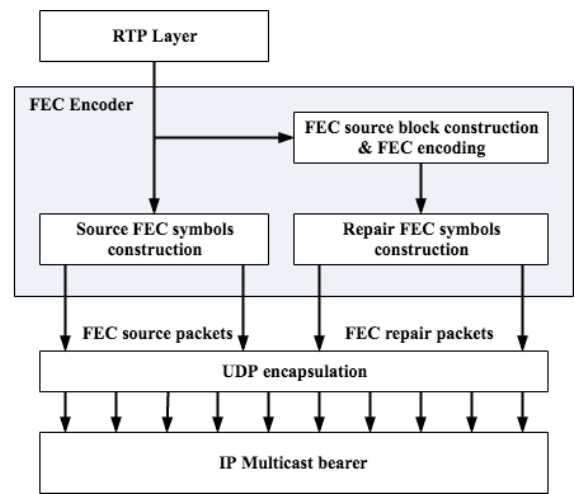
Figure 1: AL-FEC encoding process

Regarding the application of AL-FEC protection in the transmitted object we utilize the newly introduced RaptorQ FEC codes. Fig. 1 presents the encoding process of AL-FEC protection The first step is to partition the transmitted object into one or more FEC source blocks with each one of the FEC source blocks having size k symbols, a value which is an encoding parameter and is called source block length (sbl). During the RaptorQ AL-FEC encoding process a number of redundact FEC symbols are generated from each FEC source block. Those redundant symbols are called repair symbols. The amount of generated repair FEC symbols depends on the desired AL-FEC protection introduced by the multicast source. The outcome of the RaptorQ AL-FEC encoding is a certain amount of FEC encoding symbols which can be either source or repair symbols. In the simulated environment each produced FEC encoding symbol receives a number which is unique for each one.

The decoding failure probability which is the metric characterizing the decoding performance of RaptorQ AL-FEC is described as (1) [21]:

$$p_{f_{RQ}}(n,k) = \begin{cases} 1, & \text{if } n < k \\ 0.01 \times 0.01^{n-k}, & \text{if } n \geq k \end{cases} \qquad (1)$$

For a FEC source block of size k with an enduser received n encoding symbols, the quantity $p_{f_{RQ}}(n,k)$ is the probability the enduser to fail to decode this source block. Regarding the successful reconstruction of a transmitted FEC source block From the receiver perspective with respect to the previous modeled AL-FEC decoding failure probability we assume, forced from the results available in [22], that the threshold for the decoding failure probability of each FEC source block can be 10^{-2} to mark the FEC source block as successfully reconstructed or not.

Finally, regarding the multicast bearer level we model the delivery or not of a transmitted packet with an independent loss transcript. That means that on each packet sequence transmitted on each enduser an independent packet loss mask is applied according to the desired packet loss rate defined in the environment. In this work p denotes the average network packet loss rate of the mobile multicast bearers.

5.2 Optimal Offline Algorithm

To provide the competitive analysis of the proposed online algorithm we need to utilize the optimal offline algorithm of the online problem. Since the optimal offline algorithm has a priori knowledge of the problem input, and according to the work presented in [23], the optimal offline algorithm in the context of a multicast source which has to select the appropriate amount of transmission overhead introduced in a delivery is a multicast source which will select the introduced transmission overhead to a value close to the average packet loss rate of the network.

Moreover, based on the enhanced recovery performance properties of RaptorQ codes as described in (1) and based on the assumption of this work for the failure probability of the decoding process of a received FEC protected source block the optimal offline algorithm can achieve the specific threshold introducing no more FEC repair symbols than the average number of lost encoding symbols computed by the average packet loss rate p of the multicast environment.

According to previous analysis, given that sbl is the number of symbols a source block contains and r is the number of repair symbols the optimal offline algorithm decides to introduce in each FEC source block r can be calculated as: $(sbl + r) \cdot p$. Hence, the cost OPT of the optimal offline algorithm is: $OPT = sbl + r$.

5.3 Mean Transmission Overhead Algorithm

The baseline of the Mean Transmission Overhead (MTO) algorithm is the adaptation of the AL-FEC overhead based on feedback received from UEs of past deliveries. Hence, the operation of the MTO algorithm depends on the previous transmission rounds. On each round the input of the algorithm is the UE coverage attribute. UE coverage denotes a threshold on the number of UEs participating on the delivery and is used to indicate if the transmission round should be considered as sufficient protected or not. The outcome of the MTO algorithm is the computation of the mean AL-FEC overhead after each round and the update of the introduced transmission overhead with the computed value for the current round.

Each UE is able to report the outcome of the AL-FEC decoding process on the transmitted object enabling the algorithm to compute the overall UE coverage of a completed transmission round. Based on this, the MTO algorithm is able to compute the attribute ideal transmission overhead for each past transmission round. This attribute denotes the transmission overhead which should had been introduced in the delivery round to achieve the defined UE coverage for this round. After computing this ideal value transmission overhead of the completed round the algorithm is ready to update the value of the mean transmission with it. Finally, the MTO algorithm will update the current transmission overhead that will be introduced in the current transmission round with the updated value of the mean transmission overhead.

At this point, clarifying the feasibility of a crucial part on the MTO operation, i.e., the feedback received on the reception status of the transmission object from each UE, we have to mention that a various of mobile multicast standards describe a post-delivery process where the UE is able to provide feedback to the multicast source on several aspects i.e., report on delivery status, file repair status, etc. Such kind of capabilities enable a multicast source to track the status of a transmission and maintain extensive report information for past content deliveries.

Having analyzed the operation of the MTO algorithm we can compute the cost of the online algorithm. Denoting with r' the number of source symbols that the online algorithm will introduce to the transmission of a FEC source block with length sbl the value of r' is computed according to the attribute mean transmission overhead of the algorithm and the cost of the online MTO algorithm is: $ALG = sbl + r'$.

5.4 Performance Evaluation

Block Length First part of performance evaluation introduce a comparison on the percentage of UEs that are satisfied. Satisfied UEs are the ones which are able to reconstruct the original object. The comparison is done between the performance of the MTO online algorithm, an approach where AL-FEC overhead is fixed to a specific value approach performance, and lastly a common error recovery approach which is feedback-based and lost packets are retransmitted to individual UEs.

Fig. 2 contains results for a transmitted entity encoded to source blocks with the size of each block varying between 2048 to 65536 symbols. The evaluated packet loss rate is 5%. 100 participating UEs in the multicast delivery and a 65536 symbol sized transmitted object are the other parameters utilized in this evaluation. Moreover, 90% is the target coverage for

---

Algorithm 1 Mean Transmission Overhead

1: **procedure** (*ue_coverage*)
2:     *current_ideal_overhead ← computeIdealOverhead(ue_coverage)*
3:     *mean_overhead ← updateMeanOverhead(current_ideal_overhead)*
4:     *transmission_overhead ← mean_overhead*
5: **end procedure**

---

this evaluation.
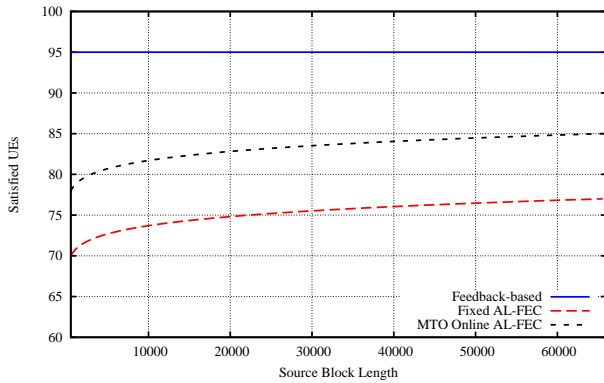
Figure 2: Satisfied UEs vs. Source Block Length

Figure 3: Satisfied UEs vs. Packet Loss Rate

As was anticipated, we cannot have results regarding *sbl* increase for feedback-based packets retransmission case since AL-FEC protection is not used there at all. We can just note that the feedback-based approach is capable to maintain continuous and high volumes of satisfied UEs. However, this performance comes at its own expense, a fact discussed in next subsections of simulation results. Discussing the performance of the approaches utilizing AL-FEC, the performance enhancement by increasing the length of the source block, is clearly noticeable. RaptorQ codes are able to operate more efficiently by increasing the *sbl* since they are able to exploit protection spreading across the whole encoded entity.

In addition, the curve trend for the quantity of the satisfied UEs we notice that is exact same for the MTO algorithm, as well as for the fixed AL-FEC transmission overhead policy, since the evaluated loss rate remains fixed, hence packet delivery conditions are not altered for each simulation snapshot and the MTO algorithm operation is founded on the adaptation of the transmission overhead according to previous reception conditions.

Packet Loss Rate In this section, the network's packet loss rate impact on the performance of the MTO algorithm is analyzed. In Fig.3 there is a presentation of the achieved return on satisfied UEs, for different packet loss rates relative to the approach introducing a constant amount of overhead and the approach where lost data are retransmitted on UEs request. On this evaluation the packet loss rate is gradually increased from 1% to 20% while delivering a 4096 packet length object to 100 UEs, assuming 4096 symbol sized AL-FEC source blocks and one packet per symbol.
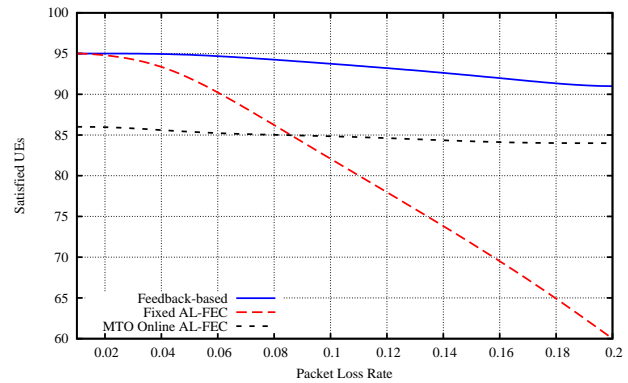
Once more, we can point out that the efficiency of the packet retransmission is superior to the one of the fixed transmission overhead protection. An expected remark is that, for only low average packet loss rates, the efficiency achieved is identical to the constant overhead protection. This is based on the fact that the introduced AL-FEC overhead is of fixed value for the fixed policy approach chosen at a higher value in comparison to the packet loss rate, and causes wasting valuable network resources. Another remark regarding the efficiency of the lost packets retransmission protection is the fact that, even slightly, it is reduced with the packet losses increase. This is an expected behavior, as for high rates of packet losses, the amount of UEs that will enter in bad reception or even dropped will constantly increase.

As shown in the curve trend, even though the performance would be ephemeral, the constant overhead approach can achieve high volumes of satisfied UEs for low average packet loss rates but this is observed just initially, i.e., for packet losses up to 2%. More on this, we can indicate that the constant overhead performance declines fast and the UEs coverage becomes 25% less than the MTO algorithm performance for packet loss 20% since the constant overhead scheme cannot respond to the increase of the packet loss. Regarding the MTO algorithm we observe that now the performance is constant. Of course, the coverage achieved is less than the lost data retransmission approach but, this has nothing to do with the MTO performance and it is just a consequence of the UEs coverage target of the simulation.

Network Resources Those latter simulations of Fig.4, examine the performance of the three evaluated approaches on a network utilization perspective i.e. the amount of exchanged packets during a multi-

cast delivery, against the increasing simulated average packet loss rates is examined. Once more, the packet loss rate ranges from 1% to 20%. For this evaluation we perform a simulation of a 4096 packet object transmission to 100 UEs. As for the AL-FEC protection parameters, we use a 4096 symbol sized source block. It is important to clarify that for the feedback-based protection scheme each UE is able to provide feedback, about the packets that need to be retransmitted, to the multicast source, through unicast bearers. Then, in a following repair transmission phase, the packets will be transmitted to the appropriate UEs from the multicast source, through unicast bearers. Finally, for AL-FEC protection application we assume exactly one FEC symbol per packet.
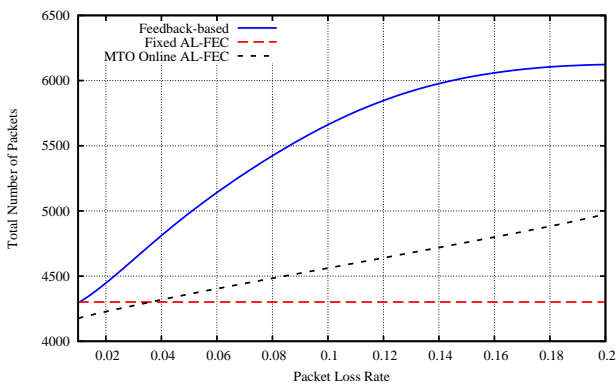
Figure 4: Total Number of Packets vs. Packet Loss Rate

The first immediate remark comes for the constant overhead approach where the amount of packets exchanged remains constant too for any evaluated packet losses value. Of course this is the expected performance for this approach as previously analyzed. The cost regarding network resources to achieve the high performance noted for the retransmission based approach as discussed in previous parts is revealed in this subsection. Compared with MTO, This approach requires a significantly higher number of packets transmitted across the network, and more specifically, we can notice that for packet losses higher than 10% this approach adds on the network over 50% of additional traffic. Finally, on the MTO approach, the curve for the number of exchanged traffic matches the network packet losses trend. This is a direct consequence of the algorithm's nature to adapt the overhead to the reception conditions. Although MTO's approach can not reach the error recovery provided by the retransmission approach as discussed in previous part, its profit on resource utilization is higher than the degradation on error protection.

6 AL-FEC Error Control on the Edge

In this second part we further analyze the AL-FEC Protection Scheme integrated with MEC presented in [5].

We provide performance evaluation results discussing initially the improvements the proposed scheme could introduce in multicast NGMN but apart from this we reveal the bounds on the performance of this scheme regarding deployment parameters.

Fig. 5 presents the basis of the architecture where the AL-FEC protection is migrated to the edge gateways of the network. While in conventional deployments the AL-FEC is applied on the multicast gateway level, in this work the AL-FEC application is also handled by the edge gateways responsible for a cluster of UEs in cooperation with the multicast gateway.
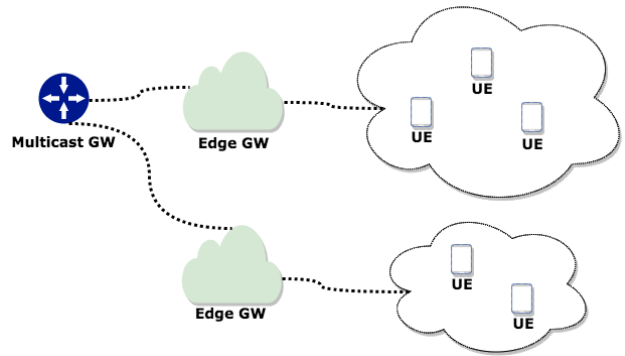
Figure 5: AL-FEC error control on the edge

The delegation of the decision on the transmission overhead amount to the level of the edge gateways enables a coarse-grained policy on the process computing the most appropriate AL-FEC transmission overhead for a delivery. Hence, this architecture provides enhanced efficiency for AL-FEC since the number of repair packets will be reduced resulting also on network load decrease always with respect to the reception conditions of the UEs cluster. Another important fact that will also provide enhanced scalability and efficiency is that any monitoring functions required for an AL-FEC scheme can also be implemented on the edge gateways.

In more details, the Multicast GW is responsible to select, according to the Multicast GW and Edge GW connection type, if it will introduce AL-FEC redundancy to the content forwarded to the edge layer. On the edge GW layer when the transmitted content arrives the GW should decide if the content will be forwarded as it is to the endusers in case of an already encoded content. In case of the arrival of a non AL-FEC protected content, the edge GW will AL-FEC encode it introducing the appropriate overhead based on networking functions implementation on the receptions conditions of the UEs cluster served by the specific Edge GW.

For the sake of clarity, we have to clarify that in the application layer, apart from the AL-FEC protection applied, a file-repair post-session is also utilized [24] where a UE that fails to decode a FEC protected source block is able to determine which source symbols are missing for the source block and achieve the retransmission of them via dedicated unicast bearers.

6.1 Performance Evaluation

In this part we present theoretical performance results on the potential improvements such a scheme has on the transmitted redundancy and in general in the overall network traffic.

AL-FEC Transmission Overhead In the presented results we investigate the improvement on the volume of the data redundancy introduced. The percentage results refer to the comparison of the proposed method with conventional architectures with the protection redundancy computed from the multicast source. In both cases we assume that both examined approaches require the same delivery success rate in the endusers.

In Fig. 6 we investigate the decrease on the average introduced error protection overhead with the increase of the number of UEs per each Edge GW for three different values of average packet loss rate 5%, 10% and 20%.
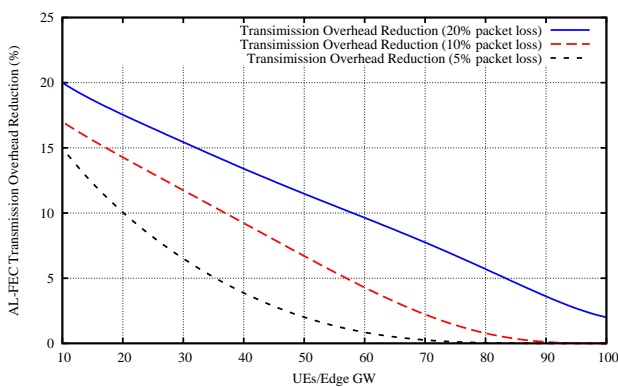
Figure 6: AL-FEC Transmission Overhead vs UEs/Edge GW

It is anticipated that as the number of UEs served by an Edge GW increase the gains on the transmission overhead are decreased since the required average transmission overhead is increased. In more details when evaluating 10 UEs per Edge GW with 20% packets lost the transmission overhead is reduced around 20% and for the same number of UEs but for 10% and 5% of average packet loss rate the improvement is about 17% and 15% respectively. For the case of 20% of average packet loss rate and for 50 UEs per Edge GW we notice a decrease of almost 10%. While, for 100 UEs the decrease is shortened to around 5%. This performance is expected since if the number of endusers assigned to an Edge GW increases then the number of UEs with heterogeneous packet losses under the same Edge GW is increased too resulting to the increase of the average transmission overhead amount that should be introduced to the delivery. However, the improvement on the introduced AL-FEC transmission overhead is still noticeable even in the case of the 100 UEs per Edge GW and 20% packet losses. Another interesting remark is that for the cases of evaluated packet losses of 5% and 10% a threshold for the efficiency of the proposed scheme is revealed with respect to the num-

ber of UEs served by an Edge GW since for 5% packet losses we have no gains for more than around 50 UEs and for 10% packet losses this threshold is less than 70 UEs.

Network Traffic In this last part we investigate the impacts of the edge error control architecture on the network traffic of the mobile network. We evaluate the total amount of traffic towards the UEs but also on the backwards channels too since a UE is able to request the retransmission of lost symbols from the multicast source via unicast bearers in case of corrupted source content.

Fig. 7 presents simulation results evaluating the impact on the network traffic in terms of packets exchanged on the network with the average packet loss rate increase from 5% up to 20% and the number of UEs served by an Edge GW increased from 10 UEs to 100 UEs.
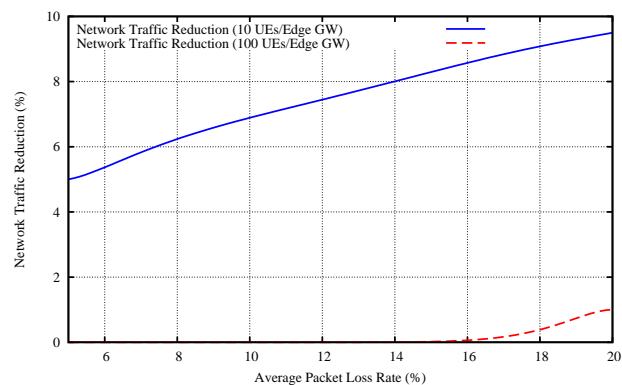
Figure 7: Network Traffic vs Packet Loss Rate

The first remark from the trends of the 10 UEs curve is that the network traffic decrease is enhanced as the packet losses increase. For 10 UEs served by an Edge GW with 5% of packet losses the network traffic decrease is more than 5% and for 20% of losses the decrease overcomes 9%. This is a direct consequence from the results presented in previous section since as the packet losses are increased the decrease achieved from the proposed scheme on the average transmission overhead required is increased too. Hence, less redundant repair symbols are transmitted required in average to achieve a successful decoding. On the other hand, as in the previous results presented in Fig. 6 the increase of the UEs delegated to an Edge GW to 100 UEs reveals again the scalability bound of the proposed scheme since we observe that in this case the decrease on the network case is eliminated for average packet loss rate values till around 18% and only extremely high packet losses of 20% denote a network traffic decrease of around 1%. This is a direct consequence of the fact that as the UEs delegated to an Edge GW increase the coarse-grained tuning on the required AL-FEC transmission overhead gets eliminated since the number of UEs with high deviations on reception conditions is constantly increased too.

7 Conclusions and Future Work

The MTO online algorithm is presented in this manuscript on the online AL-FEC policy problem. The baseline of the proposed algorithm is the adaptation of the introduced AL-FEC transmission redundancy based on the feedback received from previous transmissions. We have presented and stated the introduced MTO AL-FEC online algorithm based on the online AL-FEC policy and we have provided an analysis of the impacts of this algorithm on the application of AL-FEC error control comparing with feedback-based error recovery and the common approach of introducing a fixed amount of AL-FEC overhead in the transmission and providing simulation results on the achieved performance of the MTO online algorithm. According to the results, the online algorithm achieves close performance to that of the retransmission based approach and sometimes it overcomes it. Another aspect we have investigated is the impacts of the MTO algorithm on network resource utilization. The evaluated results verified that the online approach is the most efficient compared with the other evaluated approaches.

The second part investigates the migration of AL-FEC protection on the edge of a MEC-enabled network. Apart from this, performance evaluation results were presented indicating the gains that such a scheme can introduce on the efficiency of the AL-FEC protection application since, such a choice is able to introduce coarse-grained error recovery capabilities on a multicast network. This design choice was analyzed in terms of network traffic reduction but also we have investigated the scalability bounds of the proposed scheme under different reception conditions and deployment parameters. We have verified that, under the condition that the parameters of such an approach are carefully selected, such a design choice id capable to reduce the AL-FEC overhead and hence increase network resources utilization.

Future work of this manuscript could concentrate on the design of more sophisticated and advanced online algorithms on the problem investigated in this work i.e., advanced approaches to apply AL-FEC efficiently on NGMN. Another potential field that could extend the work presented in this manuscript is device-to-device (D2D) communication over MEC architectures. IoT architectures, apart from the common communication channels providing connectivity between the cloud and IoT devices, defines also connectivity capabilities between the IoT devices. Based on this, the proposed architecture could be enhanced enabling D2D capabilities where, during the AL-FEC source block reconstruction, missing repair symbols could be also received from another device in proximity and not from the Edge GW.

References

[1] A. Borodin and R. El-Yaniv, *Online computation and competitive analysis.* New York, NY, USA: Cambridge University Press, 1998.

[2] R. Motwani and P. Raghavan, "Algorithms and theory of computation handbook," M. J. Atallah and M. Blanton, Eds. Chapman & Hall/CRC, 2010, ch. Randomized algorithms. [Online]. Available: http://dl.acm.org/citation.cfm?id=1882757.1882769

[3] Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, and V. Young, "Mobile edge computing?a key technology towards 5g."

[4] C. Bouras and N. Kanakis, "Evolving al-fec application towards 5g ngmn," in *2018 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, Feb 2018, pp. 1–5.

[5] C. Bouras and N. Kanakis, "Al-fec application on ngmn-edge computing integrated systems," in *2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN)*, July 2018, pp. 364–369.

[6] L. Lin, N. Shroff, and R. Srikant, "Asymptotically optimal energy-aware routing for multihop wireless networks with renewable energy sources," *Networking, IEEE/ACM Transactions on*, vol. 15, no. 5, pp. 1021 –1034, oct. 2007.

[7] W. Liang and X. Quo, "Online multicasting for network capacity maximization in energy-constrained ad hoc networks," *Mobile Computing, IEEE Transactions on*, vol. 5, no. 9, pp. 1215 –1227, sept. 2006.

[8] J. Janssen, D. Krizanc, L. Narayanan, and S. Shende, "Distributed Online Frequency Assignment in Cellular Networks," *Journal of Algorithms*, vol. 36, no. 2, pp. 119 – 151, 2000. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0196677499910684

[9] A. El Gamal, C. Nair, B. Prabhakar, E. Uysal-Biyikoglu, and S. Zahedi, "Energy-efficient scheduling of packet transmissions over wireless networks," in *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 3, 2002, pp. 1773 – 1782 vol.3.

[10] Y. Bartal, J. Byers, M. Luby, and D. Raz, "Feedback-free multicast prefix protocols," in *Computers and Communications, 1998. ISCC '98. Proceedings. Third IEEE Symposium on*, jun-2 jul 1998, pp. 135–141.

[11] C. Bouras, N. Kanakis, V. Kokkinos, and A. Papazois, "Deploying AL-FEC with Online Algorithms," in *Next Generation Mobile Apps, Services and Technologies (NGMAST), 2013 Seventh International Conference on.* IEEE, 2013, pp. 175–180.

[12] C. Bouras and N. Kanakis, "A Competitive AL-FEC Framework over Mobile Multicast Delivery," in *Wireless Communications and Mobile Computing Conference (IWCMC), 2013 9th International*, jul. 2013.

[13] C. Bouras and N. Kanakis, "An Adaptive Weighted Online AL-FEC Algorithm over Mobile Multicast Networks," in *Wireless Communications and Networking Conference (WCNC), 2014 IEEE.* IEEE, 2014.

[14] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 637–646, Oct 2016.

[15] L. M. Vaquero and L. Rodero-Merino, "Finding your way in the fog: Towards a comprehensive definition of fog computing," *SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 5, pp. 27–32, Oct. 2014. [Online]. Available: http://doi.acm.org/10.1145/2677046.2677052

[16] P. Bellavista and A. Zanni, "Feasibility of fog computing deployment based on docker containerization over raspberrypi," in *Proceedings of the 18th International Conference on Distributed Computing and Networking*, ser. ICDCN '17. New York, NY, USA: ACM, 2017, pp. 16:1–16:10. [Online]. Available: http://doi.acm.org/10.1145/3007748.3007777

[17] 5G Infrastructure Public Private Partnership, "5G Vision: The 5G Infrastructure Public Private Partnership: the next generation of communication networks and services," 5G Infrastructure Public Private Partnership, https://5g-ppp.eu/wp-content/uploads/2015/02/5G-Vision-Brochure-v1.pdf, Tech. Rep., 2015.

[18] D. J. MacKay, "Fountain codes," in *Communications, IEE Proceedings-*, vol. 152, no. 6. IET, 2005, pp. 1062–1068.

[19] M. Luby, "LT codes," in *Foundations of Computer Science, 2002. Proceedings. The 43rd Annual IEEE Symposium on*, 2002, pp. 271–280.

[20] M. Luby, A. Shokrollahi, M. Watson, T. Stockhammer, and L. Minder, "RaptorQ Forward Error Correction Scheme for Object Delivery," RFC 6330, Internet Engineering Task Force, Aug. 2011. [Online]. Available: `http://tools.ietf.org/rfc/rfc6330.txt`

[21] 3GPP, "Rationale for MBMS AL-FEC Enhancements," 3rd Generation Partnership Project (3GPP), Tdoc S4-110449, 2011.

[22] 3GPP, "Simulation results for the performance and complexity of RS codes for MBMS FEC," 3rd Generation Partnership Project (3GPP), Tdoc S4-050107, 2005.

[23] C. Bouras, N. Kanakis, V. Kokkinos, and A. Papazois, "Application layer forward error correction for multicast streaming over LTE networks," *International Journal of Communication Systems*, 2012. [Online]. Available: `http://dx.doi.org/10.1002/dac.2321`

[24] 3GPP, "Multimedia Broadcast/Multicast Service (MBMS); Protocols and codecs (Release 10)," 3rd Generation Partnership Project (3GPP), TS 26.346, 2011. [Online]. Available: `http://www.3gpp.org/ftp/Specs/html-info/26346.htm`

Implementation of a Robot Contest for Distance Education

Eiken Lübbers[*], Dierk Schoen

Wilhelm Buechner University for Applied Sciences, 64319, Germany

A R T I C L E I N F O

A B S T R A C T

Robot contests are suitable for bachelor courses to gather hands-on experience in engineering diciplines such as mechanics, electronics and embedded software. After summarizing a concept of a robot contest for distance education (robot sumo) as presented in earlier publications this paper describes the implementation and first results of this concept. The contest is seperated into different tasks such as line detection and opponent location. As an example for transferring theoretical knowledge into a working prototype a simple sensor design for robot sumo is described and tested. Additionally a first student concept of a self designed robot is presented.

1 Introduction

This paper adds first implementation experiences to a robot contest concept originally presented at the 2017 International Conference on Research and Education in Mechatronics (REM) [1]. There it is shown that contests play an important role in education and in introduction of higher technology into the bachelor course [2], [3]. The presented concept is suitable for use in distance education where on-site laboratories are not readily available. Out of the evaluation of different contest formats robot sumo is selected as best suited. As an addition to this concept, that will be briefly summarized, this paper will describe the implementation of the contest into the curriculum at the Wilhelm Büchner University of Applied Sciences (WBU). To participate in a contest of robot sumo it is necessary to implement different strategies into a robot. To help students create their own solutions WBU has selected a platform compatible to the Arduino standard. Together with some prototyping parts this platform is given to the students during the project phase in the 7th semester in the module "Project Study for Engineers". The different solution out of these projects will enter into a future contest at the WBH where they will compete against standard implementations that can be purchased. To illustrate the use of creative solutions an innovative sensor implementation is shown to locate and range the opponent. Measurement data and tests are shown to evaluate the performance of the sensors. Finally, a robot is presented that was developed and made by a student during his bachelor thesis. This

robot is entirely made by 3D-printing and shows, what can be achievied within the three months avalible for both the bachelor thesis and the project study.

2 Contest Concept

The original paper [1] evaluated different robot contest formats. The criteria were robot size, arena size, budget, complexity, time to completion and research potentials. Robot Sumo in the mini class has been selected as it seems the most suitable for table top and home development [1].

3 Robot Sumo

The contest is based on the traditional Japanese game of sumo wrestling (jap. 相撲 Sumō) where two contestants try to push each other out of a circular ring (jap. 土俵 Dohyō). The contestants start behind two lines in the middle of the ring (jap. 仕切り線 Shikirisen = Shikiri lines) [4].

There are 28 articles in the unified sumo rules making this a well designed contest [5]. Figure 1 is showing a typical Sumo match situation where one robot is pushed over the edge and touches the ground outside the ring and thus loses this round.

In order to compete in a robot sumo contest the robot must implement different strategies. These are often divided into the different domains of mechanical, electrical/electronical, and software. There are three major strategies to a robot sumo match:

---

[*]Wilhelm Buechner University for Applied Sciences, 64319, Germany, Email: eiken.luebbers@wb-fernstudium.de
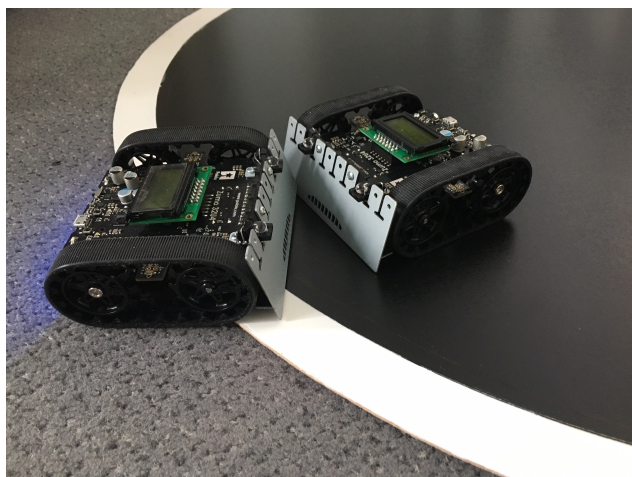
Figure 1: Typical Sumo match situation
(picture: Eiken Luebbers)

3.1 Do not leave the ring

The robot that leaves the ring first will loose the round. Thus it is important to detect the ring border so that the robot will not leave the ring by itself while moving. The rules ask for a black surface of the ring and a 2.54 cm white line around the border (mini class). The most common strategy is to detect the change in reflectance between the black and white surfaces and move backward or turn to avoid leaving the ring. Reflectance sensors are inexpensive and most will provide a digital signal that can be easily processed by microcontrollers.

3.2 Locate your opponent

The most important strategy is to locate the opposing robot so that it can be targeted with the actors. The solutions vary from moving randomly around the ring until an acceleration sensor registers a collision to LiDAR-Sensors that will determine the exact location and range. All strategies aim to put the opponent in front of the own robot so it can be pushed out of the ring. A simple but efficient implementation of this strategy will be described later in section 6.

3.3 Push the opponent out of the ring

This strategy seems to be the most straightforward implementation: Move in the direction of the located opponent. But it can be subdivided into mechanical parts such as blades, that can lift the opponent so it will loose traction and motors that could combine traction and momentum. The mechanical and electromechanical design will be part of the outlook, a first implementation can be seen in section 7.

4 Robot platform

Not all students have the experience, time, or knowledge to design a complete robot system. To concentrate on the chosen strategy in the project a platform is provided by the WBH to help the students with an easy start into development. First experiences show that there are some groups that concentrate on software, some on sensors and others on mechaical design.

The platform is a commercially available robot specifically designed for the robot sumo mini class. It is an electromechanical platform that includes the power supply out of four AA-batteries, downward-oriented reflectance sensors for the ring edge detection, general purpose inertial sensors, the motor drivers including metal-gear motors, and a differential rubber chain drive. It also includes an Arduino-compatible interface to add a microcontroller and/or sensors.

When students sign up for the project study (see section 5), the WBH will provide two of these platforms together with two Arduino compatible microcontroller-boards, two mini-breadboards for rapid prototyping and a number of wires with headers for quick installation of sensors. No soldering is necessary to get started with the platform and first implementations.

5 Project Study for Engineers

The project study has been identified as the proper format for the robot sumo development in [1]. The students learn project management principles and team work. And it must have a measurable outcome within a 3 month time frame. When working on a robot during the module "Project Study for Engineers" the students gain practical knowledge in mechatronic systems. The robots need mechanics, electronics and software to function. Students need to research the current technologies and methods and implement them into the contest. Furthermore, the project needs to be managed in time and within a given budget. This adds economic sciences and industrial engineering to the contest. And finally the project needs to be productive for a live contest at a certain interval, adding working under pressure to the picture. All this teaches the students to be productive in future jobs.

6 Sensor Solution

To demonstrate a creative solution for the strategy of "locate your opponent" the implementation in another commercially available platform is analyzed. This platform adds a microcontroller and infrared proximity sensors to the platform described in section 4. It combines an inexpensive standardized electronic hardware with a non-standard control by software to combine detecting and ranging of an object.

6.1 Sensor Electronics

The basis for the hardware is the digital proximity sensor TSSP77038 (Vishay Semiconductors) as it is used e.g. in automatic water fountains [6]. The sensor has a digital output active-low interface for microcontrollers

to signal if an object is detected or not. The sensor is sensitive for infrared light of a certain bandwidth, the typical wavelength is 940 nm. The sensor eliminates interferences from natural or artificial light by detecting only modulated infrared light with a frequency of 38 kHz. To illuminate an object that can be detected, an infrared light emitting diode (IR-LED) is driven with a carrier frequency of the same 38 kHz. If enough modulated infrared light is reflected from an object onto the sensor, the output is driven to low to signal an object in range of the sensor.

To add a relative range detection the light amount of the emitting diode is varied using a pulse-width-modulation (PWM) as an envelope function. This allows to control the output power of the diode and thus the amount of reflected light. If an object is detected with the full power of the diode and the power is then stepwise reduced, a relative range of the detected object can be determined. The amount of power when the object is no longer detected is a measure for distance. The less power is needed, the closer the object is. If used with a single emitter and single sensor a one dimensional measurement is possible.

The typical information needed for the strategy "push the opponent out of the ring" is to know, if the robot needs to turn left or right to face the opponent directly with its front so when driving forward, it will hit the opponent an can push it in front of it. To be able to have this twodimensional information the robot needs at least two emitter/sensors combinations.

The implementation actually uses three sets of emitter and sensor. One facing front, one facing left, and one facing right. Each set is controlled individually by the software.

6.2 Control Software

Each infrared light emitting diode can be individually controlled by software via PWM to send out a certain amount of light modulated with 38 kHz. Each sensor gives back the information, if enough modulated light in its field of view has been detected. Sending only light to the left will illuminate objects that are left of the robot. The right sensor should "see" significantly less light than the left or front sensor, depending on the position of the reflecting object.

6.3 Measurement Results

To test the strategy a white object was placed at certain angles and distances to the robot. The relative distances of each sensor where recorded to se if this concept can detect direction and distance of an object.

The PWM can be controlled in 6 levels, where 1 has the highest power output and 6 the lowest. The object was placed in 20 degree intervals around the center of the robot. The results are shown in figures 2, 3, and 4.

Both upper graphs show that the range can be clearly distinguished. Also it can be seen that each sensor has a certain field of view. The bottom graph

shows the overall detection capability of a white object placed at a 10.5 cm distance. But it also shows that there are certain blind spots between 40 and 60 degrees each left and right of the front. The blind spot at the rear is expected as there are no LEDs or sensors here.
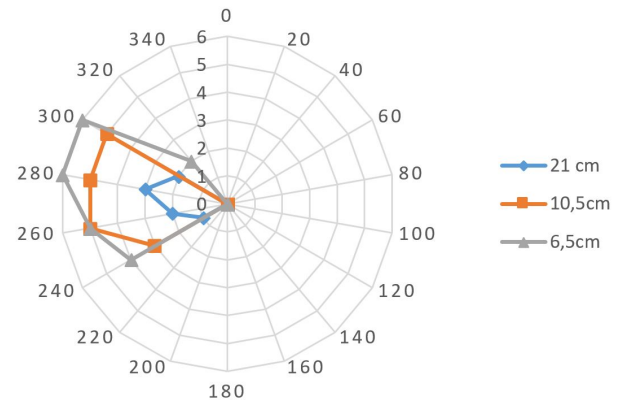
Figure 2: Measurements results left LEDs/left sensor
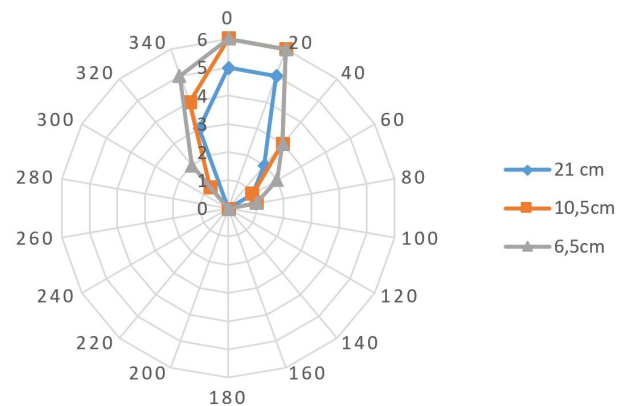
Figure 3: Measurements results middle: right LEDs/front sensor
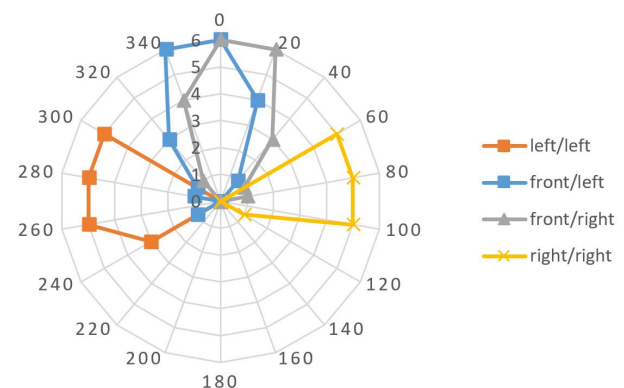
Figure 4: Measurements results bottom: all LEDs/sensors

It was shown, that the sensor concept can detect relative distance and direction of the opponent. When the front sensor shows the same distance when left and right LEDs are used, it can be assumed, that the obejct or opponent is directly in front of the robot (see figure 4 bottom graph at 0 degrees). This information can then be used in the next step where the motors are controlled to move the robot in the right direction. If the detection logic is used continuously the robot can also follow moving opponents.

7   Student Robot Solution

After showing a possible solution with an existing hardware it requires further consideration if it is possible for students to actually complete a robot construction within 3 months that are available for the project study. A student at the WBH has therefore been given the task to design and build a complete robot system with the use of a 3D-printer.

The first task was to analyze potential techniques for sumo robots [7]. The fields of traction (e.g. 2 wheel drives vs. 4 wheel drives), shields and shovels (for protection or to lift the opponent), ropes (to keep other robots at distance), flags (to trick the opponent to follow the flag) or size (keep below the line of sight of the opponent) were analyzed. Additionally the sensor technology was researched. Inertial sensors, several distance sensors, reflection sensors and cameras were described. The research was completed with a look into the Arduino platform and 3D-printing.

For the actual prototype an architecture was developed and the proper components were selected. For the traction a rubber chain differential drive was chosen. This allows the robot to distribute all its weight onto driven wheels or chains respectively to have the maximum traction within the allowed weight range. The force with which the robot can push ("traction") is the frictional force that the driven wheels can generate. This is proportional to the weight of the robot (normal force) at the contact between driven wheel and ring surface (Amontons' First Law). If the robot had only two driven wheels and one or two support wheels then some of the weight rests on the passive wheel(s) and cannot support the friction of the active drives. A chain drive is easier to implement as 4 wheel drive and still distributes all weight onto the driven components.

Out of his research the student chose Time-of-Flight (ToF) laser sensors. These sensors determine the distance to an object by measuring the time a laser pulse needs to travel to an object and back. The sensor used is the "world's smallest ToF sensor" from STMicroelectronics [8]. It is a sensor system in one housing combining a laser diode with a detection circuit using an single photon avalanche diode array and an on-silicon processing unit. It communicates via serial interface with a host controller.

The mechanical parts are 3D-printed. Figure 5 shows the finished robot. Motors and gears were purchased as were the electronic components. In total the price of the robot was about half compared to the commercial product while having more flexibility in design, sensors, and computing power. The shown robot uses two arduino controllers boards that process sensors and mechanics seperately and communicate with each other via an onboard network.

8   Summary and outlook

In addition to the original paper presented at the 2017 International Conference on Research and Education in Mechatronics (REM) [1] where the concept of robot sumo for distance education is explored this paper adds some practical measurements and a student solution to the concept. The first competition will add a major milestone to this concept of learning. Students will have completed a project where the robot is able to compete in robot sumo. Different strategies in terms of mechanics and tactics will be the source of future publications. After showing what is possible for one student within 3 months the project was oficially offered at WBH for the project study in the bachelor courses in Spring of 2018. The project has been promoted by newsletter and YouTube-videos [9]. Since launch the commercial platform has been supplied to a number of teams that are now designing and testing their robots. The first competition will be held at the end of 2018 when at least 4 teams have completed the project. The competition will be supplemented by 4 standard robots that are comercially available. The competition will be recorded and published on youtube.
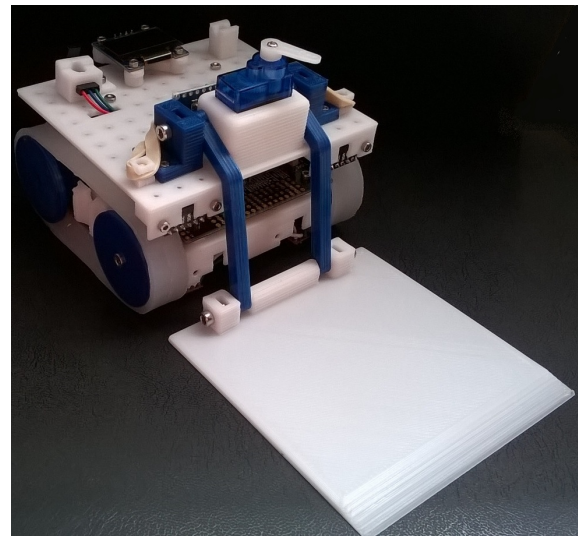
Figure 5: CAD parts and complete 3D-printed robot (source: [7])

References

[1] E. Lübbers, D. Schoen, "Robot contest concept for distance education" in 18th International Conference on Research and Education in Mechatronics (IEEE), Wolfenbüttel Germany, 2017. https://doi.org/10.1109/REM.2017.8075253

[2] T. Verhoeff, "The role of competitions in education" in Future World - International Conference & Exhibition, 1997. [Online]. Available: http://olympiads.win.tue.nl/ioi/ioi97/ffutwrld/competit.pdf

[3] A. Kolmos, "Problem-based and project-based learning" in University Science and Mathematics Education in Transition, O. Skovsmose, P. Valero, and O. R. Christensen, Eds. Berlin: Springer-Verlag, 2009, pp. 261–280.

[4] Wikipedia, "Sumō", 10.05.2017. [Online]. Available: https://de.wikipedia.org/w/index.php?oldid=165000523

[5] Robotics Society of America, Inc, "Unified sumo robot rules", 2015. [Online]. Available: http://robogames.net/rules/all-sumo.php

[6] Vishay Semiconductors, "IR Sensor Module for Reflective Sensor, Light Barrier, and Fast Proximity Applications", 2018. [Online]. Available: https://www.vishay.com/docs/82470/tssp770.pdf

[7] A. Kusmin, "Entwicklung einer autonomen Roboter-Plattform für Roboter-Sumo", Bachelor Thesis, Wilhelm Büchner University of Applied Sciences Pfungstadt Germany, 2017.

[8] STMicroelectronics, "World smallest Time-of-Flight (ToF) ranging sensor" 2018. [Online]. Available: https://www.st.com/content/st_com/en/products/imaging-and-photonics-solutions/proximity-sensors/vl53l0x.html

[9] YouTube, "Roboter-Sumo - Wilhelm Büchner Hochschule", 2018. [Online]. Available: https://youtu.be/8qtDo-sM_0s

Influence of Torrefaction on Gasification of Torrefied Palm Kernel Shell

Razi Ahmad[*,1,4], Mohd Azlan Mohd Ishak[1], Khudzir Ismail[2], Nur Nasulhah Kasim[3]

[1]*Faculty of Applied Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia*

[2]*Faculty of Applied Sciences, Universiti Teknologi MARA, Campus Arau, 02600 Arau, Perlis, Malaysia*

[3]*Coal and Biomass Energy Research Group, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia*

[4]*School of Environmental Engineering, Universiti Malaysia Perlis, 02600, Arau, Perlis, Malaysia*

A R T I C L E   I N F O

A B S T R A C T

In this study, torrefaction pretreatment on palm kernel shell (PKS) was investigated using fixed bed reactor. The PKS was torrefied at the temperatures of 210, 230, 250, 270 and 290 °C. The characteristics between untreated and torrefied PKS were compared. The results showed that, the mass and energy yield lessened, while the calorific value augmented with the increasing torrefaction temperature. Furthermore, with the rise of temperature, the oxygen composition, O/C ratio, oxygenated compounds and volatile matter of torrefied PKS decreased, but, the carbon and fixed carbon content increased. The composition of carbon in torrefied PKS was toward coal where equivalent calorific values was achieved. The gasification of torrefied PKS enhanced the product yield which produced higher gas, lower tar and char yield than the gasification of untreated PKS. Gasification of torrefied PKS increased the gas yield by 16.9 % than the untreated PKS. The tar and char yield of torrefied PKS decreased by 19.4 % and 25.9 %, respectively than the untreated PKS. Therefore, the torrefied PKS, by which their physical and chemical properties have been improved through torrefaction pretreatment is more suitable to be used in gasification and co-gasification as their influences are significant than the untreated PKS.

1. Introduction

Nowadays, the growing utilization of energy, worries on the worldwide environmental difficulties and lessening of fossil fuel, lead the nation to head for clean and renewable energy. This paper is an addition of the work formerly presented in 4th IET Clean Energy and Technology Conference 2016 [1]. Among of the main problems for fossil fuels are the discharge of contaminants such as carbon dioxide, sulfur and nitrogen oxide towards environment [2]. Therefore, biomass is one of the most attractive and broadly used renewable energy source, become important as an alternative energy resource due to little sulfur composition and neutral CO_2 supply [3].

Despite the great prospective of biomass, it has the drawbacks on its properties such as high moisture content, low energy density and hydrophilic characteristics [4-5]. Thus, these characteristics of biomass fuel are connected with some complications in biomass thermal conversion such as in gasification. Previous studied [6-7], revealed that high oxygen compound in biomass lower the gasification productivities compare with less oxygen, for example coal. Thus, alteration the properties of biomass preceding gasification is necessary.

A pretreatment step preceding to thermal conversion is required in the direction to reduce some of the aforementioned problems. Thus, torrefaction appears to be an effective route. Torrefaction involves pretreatment at temperature ranges of 200 to 300 °C in atmospheric surrounding. The pretreated biomass formed a fuel with low moisture and great energy content [2]. Previous studies also show other advantages of this torrefaction pretreatment, such as improving feedstock hydrophobicity, homogeneity and grindability [8-9].

Palm as the highest provider to biomass incomes in Malaysia has appealed huge consideration to achieve the renewable energy demands [10]. In 2016, Malaysia produced 4.19 MnT of PKS, as

*Razi Ahmad, Faculty of Applied Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia, +6049798626, razi@unimap.edu.my

residues from oil palm industry [11]. Thus, transforming PKS to bio-fuel under a thermal conversion offers a greater advantage to substitute fossil fuels, and it minimizes the disposal problems related with the generation of agricultural by-products [12]. PKS exhibited great prospective as fuel to produce gas with enhanced hydrogen and energy content [13]. However, high moisture, low heating value and energy density inhibit the PKS as valuable fuel [14]. Accordingly, these complications can be handled through torrefaction.

Consequently, the research objective was to explore the influence of torrefaction temperature on the characteristic of torrefied PKS. Further, the gasification of torrefied PKS was investigated.

2. Methodology

2.1. Materials

PKS sample was obtained from oil palm factory in Penang, Malaysia. It was crushed and sieved to get particle sizes between 200 to 400 μm. The inherent moisture was removed by drying the sample in an oven for 24 hours at 105 °C.

2.2. Pretreatment

Torrefaction, which is a mild pyrolysis pretreatment was studied thru a fixed-bed reactor at an atmospheric pressure. The reactor has inner diameter and height of 0.06 m and 0.3 m, respectively. The electric furnace surrounding reactor tube was used to heat the reactor. Figure 1 displays a schematic diagram of the pretreatment system.
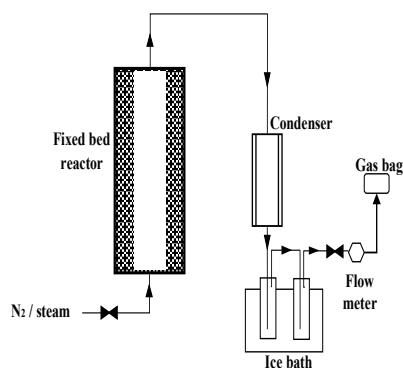
Figure 1. Schematic diagram of the reactor set-up

The sample weight of 5 g was positioned inside the reactor. The reactor was flowed with 500 ml/min nitrogen gas to generate an inert atmosphere. The reactor temperature was fixed at reaction temperatures (210, 230, 250, 270 and 290 °C) with heating rate and retention time of 10 °C/min and 1 hr, correspondingly. The torrefied PKS was weighted when the process was completed and reached the room temperature.

Mass and energy yield are main factors in the pretreatment. Energy yield represented the quantity of energy preserved after the pretreatment and is measured using mass yield of final product. Torrefied biomass has fewer energy yield than untreated biomass, due to reduce in volatile matter of torrefied biomass [15]. The mass yield (Y_m) and energy yield (Y_e) were examined via equations (1) and (2), correspondingly.

$$Y_m = (M_t / M_u) \; x \; 100 \qquad (1)$$

$$Y_e = Y_m \; x \; (HHV_t / HHV_u) \qquad (2)$$

where M_u = mass of untreated biomass, M_t = mass of torrefied biomass, HHV_u = heating value of untreated biomass and HHV_t = heating value of torrefied biomass.

2.3. Characterization

The elemental composition (C, H, N, S and O) was examined using CHNS-O elemental analyser. The proximate analysis was inspected via Mettler Toledo thermogravimetric analyser. The calorific value (CV) was measured using Leco bomb calorimeter. The functional groups were discovered using Perkin Elmer fourier transform infra-red (FTIR) spectroscopy. Table 1 listed the properties of untreated PKS.

Table 1. Properties of untreated PKS

Analysis	Value
Elemental composition (wt. %)	
Carbon	47.7
Hydrogen	5.5
Nitrogen	0.4
Sulfur	0
Oxygen[a]	46.4
Proximate analysis (wt. %)	
Moisture	10.6
Volatile matter	77.5
Ash	0.9
Fixed carbon	11.0
Calorific value (MJ/kg)	18.2

[a] By different

2.4. Gasification Experiment

Figure 1 displays the gasification system of PKS. The sample weight of 5 g was positioned inside the reactor. A nitrogen gas was flowed to the reactor for 10 min formerly the test. The sample was gasified at gasification temperature (800 °C) with heating rate of 50 °C/min. The nitrogen flow of 500 ml/min was continued to generate an inert condition. After the temperature of 800 °C had reached, the steam was streamed into the reactor and the nitrogen flow was stopped. The steam gasification of the sample was held for 45 min.

The volatile product and steam which left the reactor from the upper side were condensed in a tar trap. The solid residue was weighted as char. Tar yield in the tar trap was measured. The gas product was inspected using changed of total mass balances. The gasification was repeated for verification of the outcomes.

3. Result and Discussion

3.1. Mass and Energy Yield

Figure 2 presents the mass and energy yield of torrefied PKS under different torrefaction temperatures. The mass and energy yield reduces by increasing the temperature. The mass yield ranges from 88 to 65 % of torrefied PKS at temperature ranges of 210 to 290 °C. This displays that the conversion of PKS was increased from 12 to 35 %. The slight conversion at the temperature of 210 °C was reflected to the loss of moisture. Thus, the PKS torrefaction was insignificant at low temperature. At upper temperature between 230 to 290 °C, mass reduction was due to the major

hemicelluloses and minor lignin decomposition [4]. Some authors [16-17] established that the main decomposition part during torrefaction was hemicelluloses.

The energy yield of torrefied PKS was considerably reduced to 75 % at pretreatment temperature of 290 °C. This observation was mostly due to the additional decomposition of cellulose. Accordingly, more than 75 % of energy yield was able to be reserved at the pretreatment temperature between 250 to 270 °C. Hence, torrefaction of PKS above 290 °C is not suggested in order to avoid the loss of energy yield below 75 %.
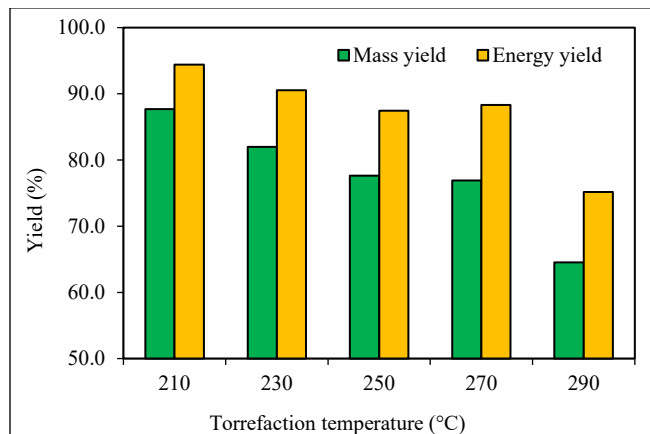
Figure 2. The influence of torrefaction temperature on mass and energy yield

3.2. Fixed Carbon, Volatile Matter and Calorific Value

Figure 3 presents the influence of pretreatment temperature on fixed carbon content and volatile matter of torrefied PKS. The fixed carbon content of torrefied PKS increased while volatile matter decreased notably with the rise in torrefaction temperature. At high temperature (290 °C), the fixed carbon of the torrefied sample improved above 50 % with comparison to the untreated sample. The torrefied sample showed huge reduction of volatile matter with close to 50 % with increasing reaction temperature up to 290 °C. The hemicellulose content in PKS is easy to degrade during torrefaction process. The results on the extensive volatile matter reduction were comparable to the work published by Uemura et al. [18], Matali et al. [19] and Sabil et al. [20] in their study of agricultural wastes.
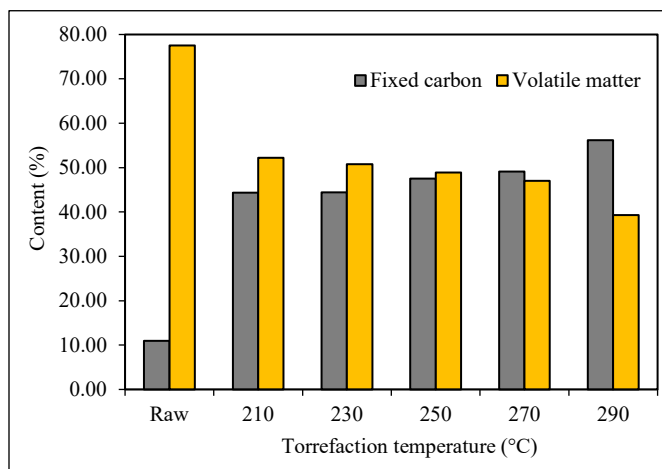
Figure 3. The influence of torrefaction temperature on fixed carbon and volatile matter
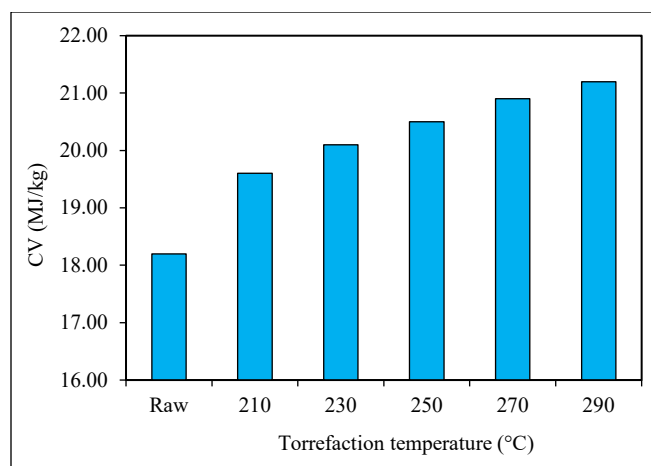
Figure 4. The influence of torrefaction temperature on calorific value

The CV of torrefied PKS is presented in Figure 4. The CV of torrefied PKS increased by rising the temperature. Improvement of CV is related with the rise of fixed carbon component. Accordingly, the PKS energy value enriched with pretreatment.

3.3. Carbon and Oxygen Content

The carbon and oxygen content of torrefied PKS are presented in Figure 5. Overall, the torrefied PKS displayed lower oxygen and higher carbon composition than untreated PKS by increasing the torrefaction temperature. The oxygen was reduced to 39 % and the carbon was increased up to 56 % at the highest pretreatment temperature of 290 °C. These outcomes appear to be in agreement with the earlier reports [21-22]. Moreover, O/C ratio of torrefied PKS reduced by rising the temperature. The reduction of the O/C ratio also indicates the measure of conversion efficiency and oxidation degree of torrefied product [4].
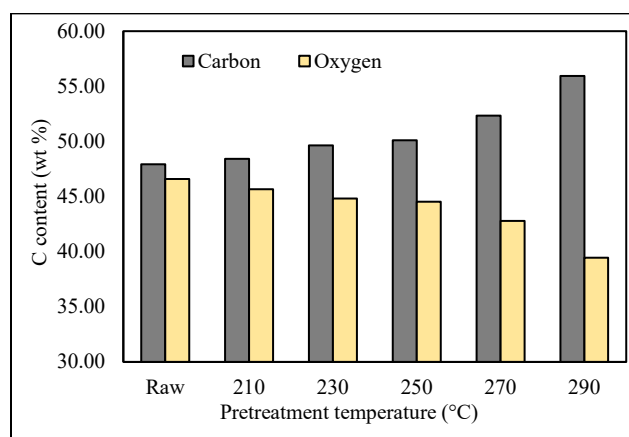
Figure 5. Carbon and oxygen content of torrefied PKS

3.4. Functional Group

Figure 6 shows the chemical structure alteration of PKS samples via FTIR. The spectrum shape was comparable for untreated and torrefied PKS, but the peak strength was dissimilar.

A broad band of 3300 cm^{-1} connected to -OH stretching which related to the alcohols and phenols. The -OH peak reduced

considerably as the pretreatment temperature increased. The aliphatic methylene group was denoted at peak of 2920 cm^{-1}. The C=O bond which is associated with aldehydes, acids and ketones was detected at 1730 cm^{-1}. At greater pretreatment temperature, the peak intensity reduced due to the breakdown of hemicellulose. The C-O stretching and O-H alteration of organic components are assigned in the peak ranges of 1000 - 1500 cm^{-1}. Granados et al. [23] also found the similar trend which the intensity of these peaks were reduced with increasing torrefaction temperature. Peak of 700 cm^{-1} indicated the aromatic groups.
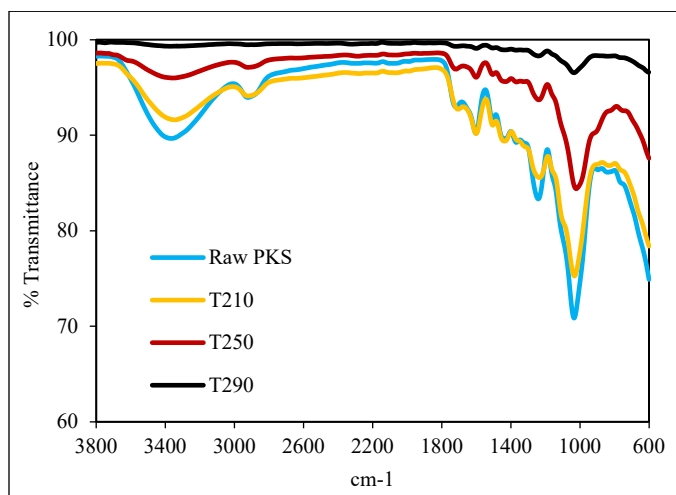
Figure 6. FTIR spectrum of PKS samples

3.5. Effect of Product Yield on Gasification of Torrefied PKS

Figure 7 shows the gasification product yield of untreated and torrefied PKS at gasification temperature of 800 °C. The torrefied PKS produced 16.9 % higher gas yield than the untreated PKS. The torrefied PKS revealed notable influence on the gas production to produce high gas yield compared to untreated PKS. This result was also in agreement with Berrueco et al. [24] which produced higher gas yield using torrefied sample.
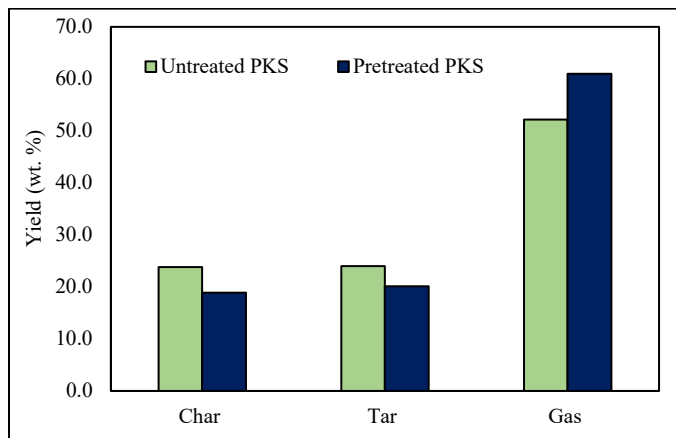
Figure 7. Gasification product yield of untreated and pretreated PKS

The tar yield reduced from 24.0 % to 20.1 % using for the torrefied PKS. The torrefied PKS had markedly reduced the tar yield about 19.4 % as an impact of the partial elimination of oxygenated components and volatiles through the torrefaction. Dudyunki et al. [25] also exhibited a comparable outcome as the effectiveness of using torrefied biomass to produce lower tar than the untreated biomass.

The gasification of torrefied PKS decreased the char yield from 23.8 % to 18.9 %. Low char yield using torrefied PKS was connected with the increased of solid conversion to gas product. Moreover, this event was influenced by the low moisture and oxygenated compound of torrefied PKS.

4. Conclusion

The influences of pretreatment temperature on torrefied PKS was investigated successfully. It was determined that the CV, fixed carbon and carbon content increased, however, mass and energy yield, volatile matter and oxygen content reduced, as the temperature augmented. Furthermore, oxygenated peak intensity in FTIR spectra decreased with increasing temperature. Therefore, PKS revealed a high value biofuel at reaction temperature from 250 °C to 290 °C. PKS torrefied at 250 °C showed significant mass and energy yield around 75 % and 85 %, respectively. The CV also increased more than 10 % compared to untreated PKS. The considerable reduction of oxygenated peak intensity was also found for torrefied PKS at 250 °C.

The gasification with torrefied PKS shows a positive effect in terms of product yield distribution. The torrefied PKS produced 16.9 % higher gas yield than the untreated PKS. The tar yield was reduced from 24.0 % to 20.1 % using the torrefied PKS. Also, the gasification of torrefied PKS decreased the char yield from 23.8 % to 18.9 %. Consequently, the torrefaction pretreatment, which improved the PKS properties, enhanced the gasification performance by producing high gas yield with low tar and char yield.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

References

[1] R. Ahmad, K. Ismail, M. A. M. Ishak, N. N. Kasim, and C. Z. A. Abidin, "Pretreatment of palm kernel shell by torrefaction for co-gasification," in 4th IET Clean Energy and Technology Conference (CEAT 2016), 2016.. https://doi.org/10.1049/cp.2016.1327 IET Digital Library

[2] C. Berrueco, J. Recari, B. M. Güell, and G. del Alamo, "Pressurized gasification of torrefied woody biomass in a lab scale fluidized bed," *Energy*, 70, 68–78, 2014. https://doi.org/ 10.1016/j.energy.2014.03.087

[3] J. S. Brar, K. Singh, J. Wang, and S. Kumar, "Cogasification of Coal and Biomass: A Review," *Int. J. For. Res.*, 2012, 1–10, 2012. https://doi.org/ 10.1155/2012/363058

[4] W. Chen, J. Peng, and X. T. Bi, "A state-of-the-art review of biomass torrefaction , densi fication and applications," *Renew. Sustain. Energy Rev.*, 44, 847–866, 2015. https://doi.org/10.1016/j.rser.2014.12.039

[5] R. Ahmad, M. Azlan, M. Ishak, N. N. Kasim, and K. Ismail, "Effect of Different Pretreatments on Palm Kernel Shell And Low-rank Coal during Co-gasification," *Progress in Petrochemical Science*, 1–7, 2018.

[6] Z. Chen, S. Zhang and Z. Chen."An integrated process for hydrogen-rich gas production from cotton stalks: The simultaneous gasification of pyrolysis gases and char in an entrained flow bed reactor," *Bioresource Technology*,

198, 586-592, 2015. https://doi.org/10.1016/j.biortech.2015.09.015

[7] A. C. C. Chang, H.-F. Chang, F.-J. Lin, K.-H. Lin, and C.-H. Chen, "Biomass gasification for hydrogen production," *Int. J. Hydrogen Energy*, 36(21), 14252–14260, 2011. https://doi.org/10.1016/j.ijhydene.2011.05.105

[8] H. Nam and S. Capareda, "Experimental investigation of torrefaction of two agricultural wastes of different composition using RSM (response surface methodology)," *Energy*, 91, 507–516, 2015. https://doi.org/10.1016/j.energy.2015.08.064

[9] D. Nhuchhen, P. Basu, and B. Acharya, "A Comprehensive Review on Biomass Torrefaction," *Int. J. Renew. Energy Biofuels*, 2014, 1–56, 2014. https://doi.org/10.5171/2014.506376

[10] Z. A. Z. Pooya Lahijani, "Gasification of palm empty fruit bunch in a bubbling fluidized bed: A performance and agglomeration study," *Bioresour. Technol.*, 102(2), 2068–2076, 2011. https://doi.org/10.1016/J.BIORTECH.2010.09.101

[11] D. Ahmad Kushairi, "Malaysian Oil Palm Industry Performance 2016 and Prospects for 2017," in Palm Oil Econ. Outlook Seminar, 2017.

[12] R. Ahmad, N. Hamidin, U. F. M. Ali, and C. Z. A. Abidin, "Characterization of Bio-Oil From Palm Kernel Shell Pyrolysis," *J. Mech. Eng. Sci.*, 7(12), 1134–1140, 2014. https://doi.org/10.15282/jmes.7.2014.12.0110

[13] N. A. Samiran, M. N. M. Jaafar, J. H. Ng, S. S. Lam, and C. T. Chong, "Progress in biomass gasification technique - With focus on Malaysian palm biomass for syngas production," *Renew. Sustain. Energy Rev.*, 62, 1047–1062, 2016. https://doi.org/10.1016/j.rser.2016.04.049

[14] M. A. Sukiran, F. Abnisa, W. M. A. Wan Daud, N. Abu Bakar, and S. K. Loh, "A review of torrefaction of oil palm solid wastes for biofuel production," *Energy Convers. Manag.*, 149(10), 101–120, 2017. https://doi.org/10.1016/j.enconman.2017.07.011

[15] E. M. Gucho, K. Shahzad, E. A. Bramer, N. A. Akhtar, and G. Brem, "Experimental study on dry torrefaction of beech wood and miscanthus," *Energies*, 8(5), 3903–3923, 2015. https://doi.org/10.3390/en8053903

[16] M. Asadullah, A. M. Adi, N. Suhada, N. H. Malek, M. I. Saringat, and A. Azdarpour, "Optimization of palm kernel shell torrefaction to produce energy densified bio-coal," *Energy Convers. Manag.*, 88, 1086–1093, 2014. https://doi.org/10.1016/j.enconman.2014.04.071

[17] M. Wilk, A. Magdziarz, and I. Kalemba, "Characterisation of renewable fuels' torrefaction process with different instrumental techniques," *Energy*, 87, 259–269, 2015. https://doi.org/10.1016/j.energy.2015.04.073

[18] Y. Uemura, W. N. Omar, T. Tsutsui, S. B. Yusup, and S. Bt, "Torrefaction of oil palm wastes," *Fuel*, 90(8), 2585–2591, 2011. https://doi.org/10.1016/j.fuel.2011.03.021

[19] S. Matali, N. A. Rahman, S. S. Idris, N. Yaacob, and A. B. Alias, "Lignocellulosic Biomass Solid Fuel Properties Enhancement via Torrefaction," *Procedia Eng.*, 148, 671–678, 2016. https://doi.org/10.1016/j.proeng.2016.06.550

[20] K. M. Sabil, M. a. Aziz, B. Lal, and Y. Uemura, "Synthetic indicator on the severity of torrefaction of oil palm biomass residues through mass loss measurement," *Appl. Energy*, 111, 821–826, 2013. https://doi.org/10.1016/j.biombioe.2013.05.015

[21] S. Zhang, Q. Dong, L. Zhang, Y. Xiong, X. Liu, and S. Zhu, "Effects of water washing and torrefaction pretreatments on rice husk pyrolysis by microwave heating," *Bioresour. Technol.*, 193, 442–448, 2015. https://doi.org/10.1016/j.biortech.2015.06.142

[22] Y. Uemura, S. Saadon, N. Osman, N. Mansor, and K. Tanoue, "Torrefaction of oil palm kernel shell in the presence of oxygen and carbon dioxide," *Fuel*, 144, 171–179, 2015. https://doi.org/10.1016/j.fuel.2014.12.050

[23] D. A. Granados, R. A. Ruiz, L. Y. Vega, and F. Chejne, "Study of reactivity reduction in sugarcane bagasse as consequence of a torrefaction process," *Energy*, 139, 818–827, 2017. https://doi.org/10.1016/j.energy.2017.08.013

[24] C. Berrueco, D. Montané, B. Matas Güell, and G. del Alamo, "Effect of temperature and dolomite on tar formation during gasification of torrefied biomass in a pressurized fluidized bed," *Energy*, 66, 849–859, 2014. https://doi.org/10.1016/j.energy.2013.12.035

[25] M. Dudyński, J. C. van Dyk, K. Kwiatkowski, and M. Sosnowska, "Biomass gasification: Influence of torrefaction on syngas production and tar formation," *Fuel Process. Technol.*, 131, 203–212, 2015. https://doi.org/10.1016/j.fuproc.2014.11.018

Complexity Drivers in Digitalized Work Systems: Implications for Cooperative Forms of Work

Benedikt Andrew Latos[*], Markus Harlacher[*], Florens Burgert, Verena Nitsch, Philipp Przybysz, Susanne Mütze-Niewöhner

Institute of Industrial Engineering and Ergonomics, Work Organization Department, RWTH Aachen University,

52062 Aachen, Germany

A R T I C L E   I N F O

A B S T R A C T

Digitalization changes the way people work to a considerable extent. It alters business models and process organizations of whole industries. The ensuing market dynamics and faster innovation cycles cause an increase in complexity. In this article, the interconnection of digitalization and complexity in work systems is analyzed. For this purpose, a framework for comparing relevant complexity definitions is developed. Moreover, complexity drivers in digitalized labor systems in six different organizational dimensions (process organization, organizational structure, technology, working conditions, product and personnel) are explored. 23 experts from the academic and industrial sector were interviewed using semi-structured interviews. The results of a qualitative content analysis show that the consideration of complexity and digitalization has extensive impact what becomes evident in interdependent relations amongst the organizational dimensions. Furthermore, complexity drivers in digitalized work systems are determined as a result of the analysis procedure. Finally, the implications of the expert interviews for cooperative forms of work are discussed. The concept of a "task complexity mountain range" is presented to explain the effect of task complexity on performance and motivation in the context of work groups.

1. Introduction

This article is an extension of work originally presented on the IEEE International Conference on Industrial Engineering & Engineering Management in Singapore (December 2017) [1].

Digitalization considerably affects work systems through diverse and quicker communication media as well as through the networking of manufacturing facilities and information systems. Industry 4.0, or the concept of advanced manufacturing, is used to manufacture individualized consumer goods or machines that are progressively equipped with mechatronic components [2]. Also, developments towards dynamic customer requirements change production planning for these products, since personalized customer requirements necessitate real-time and flexible reactions within a company [3]. This causes an increase in complexity in the process organization. According to [4], [5], though, the use of digital systems is able to reduce complexity in production planning, as they enable a more efficient and flexible control and management of company processes. However, the use of digital

systems may cause interface problems or media discontinuities due to missing IT standards, since different IT systems are used for data acquisition that are not compatible with each other [6]. Yet, digitalization also enables the introduction of internal and external company network structures so that distributed and global cooperation across company boundaries is feasible [7]. Another significant change in the course of digitalization, which is discussed in literature, is the increase in the decision-making powers for employees that also partly results from flatter hierarchies [8], [9]. There is a general trend towards decentralized forms of management and control [10]. This implies that team and group work or interaction between individuals (in this article the term "cooperative forms of work" is employed) is relevant – especially in the course of digitalization. The changes result in an increase in the demands on employees. On the one hand, social competences are required. On the other hand, IT-related knowledge and skills in particular are becoming more important because digital technologies are used in almost all areas of companies and therefore must be handled by the employees [6], [9], [11], [12].

[*] Benedikt Andrew Latos, Markus Harlacher; Bergdriesch 27, 52062 Aachen, Germany, b.latos@iaw.rwth-aachen.de; m.harlacher@iaw.rwth-aachen.de

As all these facets might cause levels of higher complexity, it is important to install appropriate complexity management procedures in highly digitalized labor environments [13]. In order to manage complexity, it is necessary to understand the causes of complexity and its relationships, and to keep in mind that complexity may have both positive and negative effects.

This article investigates how the digital transformation relates to complexity in work environments. It examines work systems that use information and communication technologies which are interconnected to a large extent and are increasingly mobile as well as systems that process and supply significant proportions of information in digital forms (cf. [14]). Current complexity definitions from relevant research fields do not entirely incorporate important complexity characteristics with respect to a work system. Therefore, a framework for work systems is employed to systematize and analyze existing complexity definitions (section 2). Moreover, there is a lack of suitable classification systems for complexity drivers in digitalized work systems. In order to systematize complexity drivers that are explored with 23 semi-structured expert interviews (section 3), six organizational dimensions are chosen as a framework to present the results (section 4). Subsequently, complexity management approaches in digitalized work systems are outlined and digitalization-related effects on cooperative forms of work are discussed (section 5). The individual task complexity concept of Dalhöfer and Prieß [15] is extended to consider a work group. The article concludes with a critical discussion of the findings and outlines future research steps (section 6).

2. Complexity definitions from different research disciplines

The expression *complex* is widely employed in everyday language to express the difficulty of comprehending, investigating and forecasting relationships or systems' behavior that is not easy to understand. In research, the use and understanding of complexity depend on the specific research discipline [13], [16], [17], [18]. In order to conceptualize actions to handle complexity in digitalized work systems, it is vital to develop a theoretically grounded approach. Therefore, it is necessary to systematize existing definitions of complexity from relevant research disciplines in a framework that is suitable to represent a work system. As a basis for this framework the general model of a work system according to Schlick et al. [19] was chosen. This model describes the transformation process of input variables inside the work system into outputs. The working person that executes this transformation with work equipment is an essential part of the work system. Relevant research disciplines from the pluri- and interdisciplinary research field of industrial engineering and ergonomics that consider work system design are in particular: engineering, economics, sociology, and industrial and organizational psychology (see e.g. [19]). In order to develop an understanding of complexity for this framework, the understanding of complexity within these research disciplines is examined below:

Although diverse scientific disciplines employ the expression *complexity* in different manners, general system theory is widely used to define the phenomenon of complexity [16]. Since a work system can be regarded as a general system, the characteristics of

complex systems are outlined: systems that are meant to be complex have lots of system elements as well as they exhibit many interactions simultaneously [20][21]. Relevant literature on system theory indicates that certain characteristics of complex systems are the number and type of elements and interactions, rate of change of the system structure and uncertainty of the current and future system status (various complexity definitions can be found e.g. in [16][22]). In fact, von Bertalanffy [23] states that the characteristics of a system are due to the properties of its elements and the connections amongst them. The construct of emergence is special for complex systems. This phenomenon means that the system's behavior is not entirely explainable with the properties of the single system elements [24]. Schuh and Schwenk [25] distinguish between simple, complicated, relatively complex and extremely complex systems, based on their number and diversity of the elements and relations as well as their variability and dynamic (see Figure 1). Whilst simple systems are characterized by few elements, interrelations and behavioral possibilities, extremely complex systems have a multitude of elements with a great variety of interrelations and possible system behaviors with variable impacts of the elements [26].
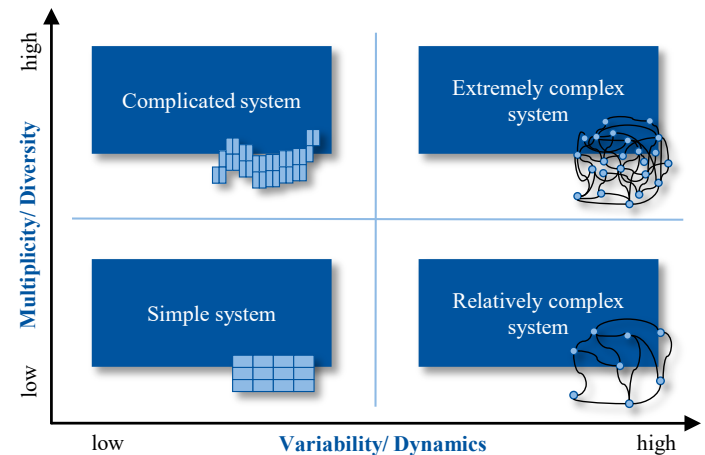
Figure 1: System states according to Schuh and Schwenk [25], Ulrich and Probst [26]

System theory is also frequently used in engineering disciplines to explain the phenomenon of complexity. Therefore, the complexity properties from general system theory also apply to considerations of complexity in an engineering context [27], [28], [29]. For example, system elements in a labor system are produced goods, tools, assignments or personnel. The system's interrelations are, for instance, relationships of the elements during production or communication procedures. Because digitalization in work systems could have an impact on the elements or relationships in the system, this could also change the level of complexity in it.

In the literature, researchers distinguish static or structural complexity from dynamic or operational complexity when manufacturing systems are regarded [27]. The two facets are often generalized as objective complexity, whereas Blockus [22] describes complexity in the field of business administration with reference to system theory and includes perception as a subjective complexity characteristic. This aspect is crucial for a work system framework, since workers operate in these systems. For production systems, subjective complexity reflects how individual persons

experience complexity, e.g. with respect to the design of single work stations in assembly [30]. Zeltzer et al. [31] state that complexity perception in assembly is largely determined by the operator and therefore define complexity as subjectively negative perceived effects of the assembly system on the operator.

Persons in work systems have to make numerous decisions whilst performing their jobs. This is why the complexity understanding in psychology is important for the framework. Within the discipline of psychology, complex situations and problems are described by interrelated variables, a considerable level of dynamics, a lack of transparency concerning a situation and multiple goals that are conflicting (polytely) [32][33][34]. Willke [35] provides a sociological definition that considers the degree of multidimensionality, interconnectedness and relevance of the consequences of a decision problem. Complex systems have diverse options that can be chosen so that operators have to select an alternative [36].

The aforementioned complexity definitions contain central characteristics, which are relevant for a complexity definition with regard to a work system. For a systematic comparative analysis of existing complexity definitions, which relate to the chosen framework, relevant criteria are required:

In this article, work systems are considered with regard to the transformation of work through digitalization. Therefore, a system-theoretic criterion "system size" should be defined that considers the type and number of elements, as well as interactions between them. As second criterion, "dynamics and uncertainty" should be included. This criterion considers the non-static and emergent behavior of complex systems. Concerning the work system framework, the system-theoretical concept should be expanded to consider humans and effects of working conditions on humans. Decision behavior of human operators in complex situations within their work systems should be taken into account. Therefore, "polytely" is defined as a criterion for complex situations with conflicting goals. In addition, complexity may arise due to human information processing and differing semantic perceptions. Following the stress-strain-concept [37][38] and the individual perception of complexity according to Blockus [22], it has to be stated that human perception of complexity is subjective. Thus, the criterion "subjectivity" is introduced. With respect to the transformation of work through digitalization, the aspect "digitalization" is explicitly included. Figure 2 presents a selection of complexity definitions from relevant research disciplines with regard to the derived criteria.

Author(s)	Research Discipline	Definition of Complexity/ Complexity Characteristics	System Size	Dynamics and Uncertainty	Polytely	Subjectivity	Digitalization
von Bertalanffy (1972)	System theory	Number of elements, variety of elements, number of interrelations	●	●	○	○	○
Ulrich and Probst (1988)	System theory	Complex systems are characterized by a high number and diversity of the elements and relations as well as high variability and dynamics. This allows complex systems to be distinguished from simple and complicated systems.	●	●	○	○	○
Deshmukh et al. (1998)	Engineering	Static complexity of a production system is characterized by its structure, variety of subsystems and strength of the interactions.	●	○	○	○	○
Sivadasan et al. (2006)	Engineering	The dynamic or operational complexity is defined as uncertainty regarding the temporal change of the system.	○	●	○	○	○
Blockus (2010)	Business Administration	The main characteristics are the number and type of elements and connections and the variability over time. Beside these objective complexity characteristics, perception is considered as a subjective component.	●	●	○	●	○
Zeltzer et al. (2013)	Engineering (assembly system design)	Complexity of a workstation is the sum of all technical and ergonomical aspects and factors that make the set of tasks to be performed within it by an operator mentally difficult, error-prone, requiring thinking and vigilance, and inducing stress.	○	○	○	●	○
Dörner et al. (1983), Dörner (2011), Kluge (2004)	Psychology	A complex problem is described by four characteristics: dependencies between variables, internal dynamics, non-transparency of the situation, multiple conflicting goals (polytely).	◐	●	●	○	○
Willke (1996)	Sociology	Complexity is described by three characteristics: multidimensionality, interconnectedness and relevance of the consequences of a decision problem.	◐	○	◐	◐	○
Luhmann (1991)	Sociology	A system is complicated with a large number of elements and connections, whereby a complex system simultaneously contains many alternative options and forces selection.	●	○	●	◐	○
Explanation for evaluation criteria: ● fulfilled; ◐ partially fulfilled; ○ not fulfilled							

The results illustrate that none of the considered definitions fully covers all defined criteria. Striving towards a definition containing all criteria, a synthesis of the discipline-specific definitions is required. Additionally, none of the definitions includes the aspect of digitalization. As a conclusion, the findings can be used to derive an integrated comprehension of complexity characteristics in the context of digitalized work systems:

- Complexity in a system refers to a construct of higher order that is composed of single system elements.

- A complex system, as sum of its elements, has the properties of high dynamics caused by multidimensional interactions, unpredictability of the system state and polytely, inducing self-generated characteristics of the system.

- Concerning the digitalization of work systems, objective complexity has to be distinguished from subjective complexity. While subjective complexity is induced by information processing and perception of the individual, objective complexity arises from interactions of internal system relations and external environmental impacts.

Especially from a work design point of view, it is crucial to develop a deep comprehension of variables that induce complexity in a digitalized work environment. On the one hand, labor systems with a too high degree of complexity are avoidable. On the other hand, operators may be qualified to manage complexity, if they are provided with sufficient information via support systems within work procedures. Moreover, complexity is also mentioned in terms of knowledge management as a basis for developing flexible and self-learning organizational structures [39]. All these aspects can only be considered, if there is a fundamental knowledge of complexity drivers in a work system. In this article, the understanding of the term "complexity driver" is based on the general definition of Vogel and Lasch [40], which they developed from the use of the term in the literature:

"Complexity drivers are factors, which influence a system's complexity and company's target achievement. They are responsible for increasing system's *[sic]* complexity level and help to define the characteristics or the phenomenon of a system's complexity. Complexity drivers are influenced by one another, that is by internal or external drivers, and cannot be reduced completely to another one" ([40], p. 18).

As another result of their systematic literature research, Vogel and Lasch [40] show that there is a multitude of different approaches to identify and classify complexity drivers. They identify complexity drivers with a systematic literature research procedure and employ a quantitative frequency analysis of mentioned complexity drivers. They present a general complexity driver classification system, based on existing classification systems, which is shown in Figure 3.

External complexity drivers are distinguished from internal and general complexity. The internal complexity can be actively influenced by the company. It is further divided into correlated complexity drivers, which are directly related to external complexity impacts, and autonomous complexity drivers, which are not directly influenced by external impacts. The general complexity category includes all drivers that cannot directly be

assigned to other specific categories (e.g. uncertainty). The rectangles in Figure 3 represent one complexity driver category that is further subdivided into numerous complexity drivers. Detailed descriptions of the assigned complexity drivers are presented in [40].

However, it can be seen that digitalization is not explicitly included as a complexity driving factor. This also applies to alternative complexity driver systematizations (see e.g. [41][42][43][44]). Furthermore, diverse methods exist in order to assess the level of complexity in work systems (see e.g. [41][42][45]). However, these approaches also do not explicitly take changes into account that are caused through digitalization.

Above all, these classification systems regard complexity on a general or company-wide level. They do not focus on complexity drivers that directly relate to work of the persons in a company and their specific work systems. This necessitates that a suitable framework has to be used to explore drivers of complexity in digitalized work systems. This framework is presented in the next section.
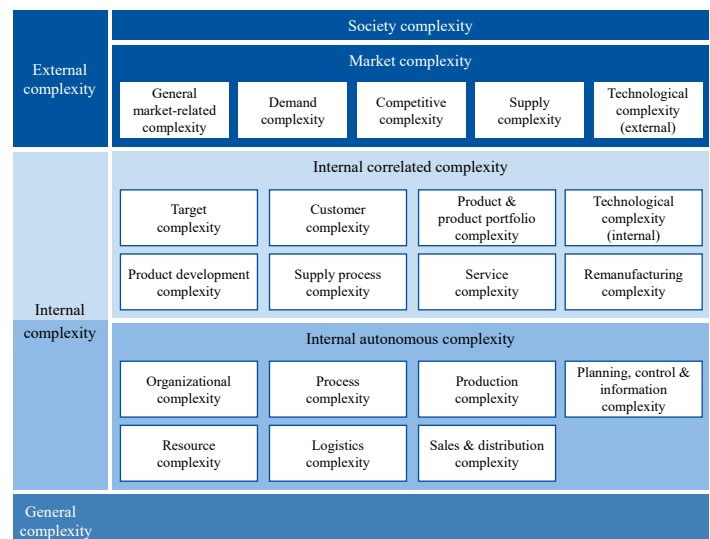
[40]

3. Methodology

As a framework to systematize drivers of complexity in digitalized work systems, six organizational dimensions were chosen. The organizational dimensions were the process organization, organizational structure, technology, working conditions, product and personnel. The authors have employed this framework already in diverse transformation projects in different industry sectors. In these projects, the dimensions have proven to be applicable for holistic organizational transformation projects regardless of industry settings.

To identify drivers of complexity for every organizational dimension in digitalized work systems, semi-structured expert interviews were performed. This methodology was selected, because it provides valuable qualitative insights at an initial exploration stage. Moreover, it can be used to complement ongoing quantitative research steps. The questions consisted of a constant number of open questions. This procedure guaranteed that

all important questions were treated in all interview sessions. However, the interview design made it possible to stress on aspects that are especially important for an individual interviewee at the same time. Also, the aspects could be discussed according to the individual reference framework. The duration of the interviews was about 45 minutes. The interviews were transcribed according to Mayring [46] by using the audio recordings with permission of the interviewees. In one case, no recording was available so that interview notes were used. The interview transcription time took up to 8 hours for one interview. In total, 23 interviews were performed. 22 interviews were conducted via phone. One was a personal interview. The majority was conducted as phone interviews because of interview economics reasons: the interviewees worked in companies that are spread throughout Germany. Figure 4 gives an overview on the experts and their backgrounds that were included in the study. The interviewees were acquired via using the databases and business networks of the research institute. They were requested via e-mail. In total, about 30 interview requests were sent, whereof 23 interviewees accepted the invitation. For being taken into account as an expert, two requirements had to be fulfilled: the experts needed to have privileged access to information and must have been responsible for conceptualizing, implementing or controlling problem solutions with regard to digital strategies [47].

The 23 interviewees held positions in German academia or industry and were requested because of their positions and knowledge of digitalization and complexity management or their expertise of work organization. Moreover, the experts belonged to industries from both production and service. Most worked in the fields of engineering, the management of projects, consignment and logistics, information and communication technology, banking or within the media industry.

Also, the experts were representatives of entrepreneurs or belonged to committees that represent employees. To prevent biased analysis results, the expert acquisition procedure was conducted in a manner to obtain interviewees that cover several domains and stakeholder groups.

Number	Sectors / Domains	Function of the expert	Comment	Number of employees
1	System house	Director of system house	Development and consulting of IT system houses (focus on process organization)	< 50
2	Science	Professor information management	Focus on hybrid services, digital business models	< 50
3	White goods	Head of Department assembly	Cycled assembly system (large quantity)	> 10,000
4	Software company	Customer adviser product lifecycle management	Cloud computing, Big Data analysis	> 1,000
5	White goods	Head of Department industrial engineering	Digital assistance systems with work instructions, error feedback via apps	> 10,000
6	Automotive	Expert industry 4.0 in corporate work council	Focus on digital work, industry 4.0	> 100,000
7	Start-up electro mobility	Chief information officer	Agile, digital product development, smart products in the automotive sector	< 250
8	Science	Director of demonstration factory	Demonstration Factory Industry 4.0	< 250
9	Science	Expert industry 4.0	Holistic production and enterprise system design, project management	< 250
10	Logistics, consignment	Chief executive officer	System solutions for working with smart glasses	< 50
11	Industrial Connectivity	Head of Department global digitalization and intelligence	Data connection of world-wide facilities	> 1,000
12	Automotive	Complexity manager	Focus on the use of digital analysis methods	> 50,000
13	Service, maintenance and repair	Key account manager	Proactive maintenance through the use of digital tools	> 1,000
14	Consulting and research	Director of consulting company	Enterprise organization	< 50
15	Media	Chairperson of work council	Print media	> 1,000
16	Media	Chairperson of work council	Online media	> 250
17	Information and communication technology	Chairperson of work council and member of the supervisory board	Technology and Innovation Committee	> 10,000
18	Logistics	Deputy chairperson of corporate work council	Member of the IT Committee	> 100,000
19	Banking	Member of work council	Member of committees for the adoption of resolutions, e.g. on topics related to digitization	> 1,000
20	Banking	Quality manager and member of work council	Payment service provider	> 10,000
21	Logistics, consignment	IT-expert	System administrator	> 50,000
22	Information and communication technology	Director Human Resources	Operation of IT systems for customers	> 10,000
23	Academia	Director of university campus	Research and teaching on industry 4.0	< 250

The questionnaire guide was divided into four parts. In a first step, demographic information on the interviewees was documented. The second questions asked for aspects that change companies and labor through digitalization; e.g. question were: what are the main digitalization-related changes in your company within the last 10 years? Which important changes do you anticipate for the next 10 years? The third question part focused on whether and how digitalized labor systems result in higher complexity levels. Therefore, the experts should name complexity drivers which are caused by digitalization for the six organizational dimensions respectively. In the final part of the interview, the experts were asked to identify measures to manage complexity in digitalized labor environments. Figure 5 summarizes the content of the interviews. The detailed questionnaire guide is included in Appendix 1.
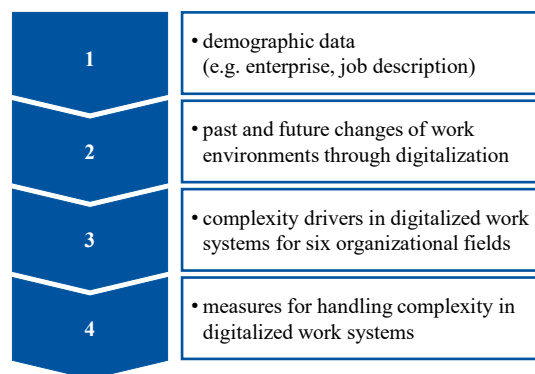
Figure 5: Structure of the interview template

The analysis process initially incorporated a verification of the transcription documents. This was conducted with the available sound recordings of the interviews. Then, the single interviews were qualitatively analyzed via employing the methodology of qualitative content analysis [48][49][50].

A qualitative content analysis is a systematic and intersubjective verifiable text analysis procedure that meets the needs of interpretation and semantic meanings of a linguistical research basis [51]. In contrast to quantitative content analyses, the main purpose is not to discover quantitative connections, but rather to reduce a large amount of analysis material and to extract the main contents. This makes the method especially useful for explorative studies. Mayring [48] describes a systematic and rule-based analysis approach. The structured and transparent procedure enhances the reliability and validity of the analysis. In this case, the general procedure was adapted to the specific research aim.

In a first step, a qualitative-interpretative analysis of every interview was conducted. As code unit, a single word was chosen and a whole sentence structure was defined as context unit. Furthermore, the analysis unit was structured according to the interview template. Within the analysis procedure, the statements of the experts were generalized, paraphrased and reduced to elementary categories. The categories were derived by using a hybrid approach of deductive and inductive category formation. The inductive approach was chosen, since the research did not aim at confirming existing theories. The aim was rather to extract general statements from the interview material. However, deductive analysis elements were included, as statements were

assigned to expressions that are well established in literature. For example, the expression "polytely" (cf. section 2) was chosen as category name for expressions that contained phrases such as "conflicting goals" or "different goals that cannot all be reached at the same time". This simplified a purposeful assignment of the interview phrases.

In a second step, the single analysis results were combined and aggregated with a quantitative analysis of mentioned categories. This made it possible to describe how often a certain category was mentioned by the interviewed persons. By doing so, complexity drivers were extracted via counting the number of mentions for single driver categories within all 6 organizational dimensions. Figure 6 summarizes the whole analysis procedure.
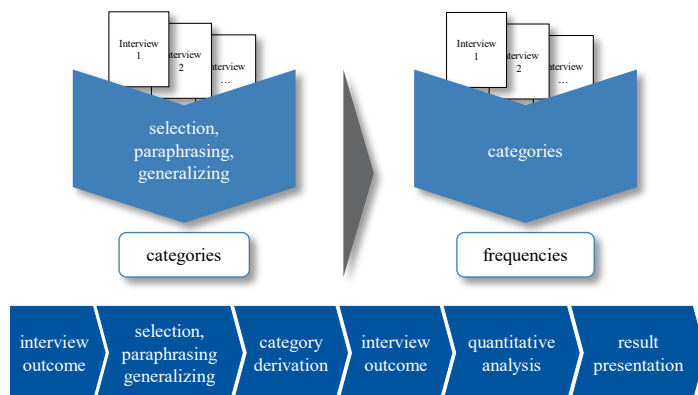
Figure 6: Procedure of qualitative content analysis according to Petz et al. [52], Duckwitz et al. [53]

According to Mayring [48], a qualitative analysis should meet two major quality criteria. On the one hand, intra-code-reliability requires that a repeated analysis procedure of the same coder after a certain time should lead to similar results. On the other hand, inter-code-reliability means that different persons will achieve similar analysis results. As the method is an interpretative approach, there will always be slight differences in the results, e.g. concerning the wording of categories between different coders (see e.g. [48]). With this analysis procedure, one repeated analysis by the initial coder as well as one further analysis with an additional coder was conducted. In both cases, similar results were extracted from the interviews which can be seen as indication for meeting the reliability criteria.

4. Results

In this section, the results of the aforementioned qualitative content analysis procedure are presented. In the first subsection, past and future changes of work environments through digitalization that were extracted from the interviews are outlined. Subsequently, the derived complexity drivers are presented for every organizational dimension, respectively. In the third subsection, the results for measures to handle complexity in digitalized labor systems are displayed.

4.1. Past and future changes of work environments through digitalization

Concerning the last 10 years, the experts pointed out diverse changes within their companies that have been caused through digitalization. The main changes are due to the new opportunities of manufacturing customized products. In addition, a development

towards hybrid products in production industry was named. This means that the product is enhanced with further service options for the customer.

Also, the experts stated that former manually documented procedures have been changed into digital processes. These workflows incorporate several companies so that the dynamics and speed of these procedures have considerably increased as well as the duration for reactions has become shorter. Enterprise software, information and communication tools and cheaper sensor technologies offer the opportunity for companies to track and process big amounts of data. However, the operators within the companies very often have a problem to interpret and mentally capture these masses of information. As a solution for this problem, the use of big data algorithms was proposed.

Employing information and communication technology remarkably altered every day work procedures, since the amounts of interpersonal face-to-face communication are reduced. Nowadays, (digital) work forms of collaboration can be observed increasingly. Portable devices such as mobile phones, tablet computers and netbooks make it possible to perform the job in flexible work forms in terms of location and time. This gives employees more flexibility and enhances family friendliness of enterprises, whereas the persons at the same time have the drawback of possibly being always available. The interviewees concluded that these changes rather relate to administrative departments and service sectors than to manufacturing. Within the service sector, value creation networks have a higher level of digitalization than in manufacturing.

With respect to the next 10 years, the interviewees predicted increasing and quicker rates of process changes. Consequently, data will not be documented manually any more. Planning procedures will be complemented with digital support tools as well as data in real-time will be provided. Concerning manufacturing steps, the interviewees pointed out that planning and controlling would change from an internal enterprise-focused view towards an approach that considers whole value-chains. Furthermore, business models will more and more tend to include services. Therefore, products will include more software and will be a medium to offer services to the users of the products. For instance, a car manufacturer will no longer be only selling cars but rather mobility, since the car offers supplemental services to the customer.

Although experts that represented employees feared big job losses because of digitalization, the majority of the interviewees agreed that operators would not be replaced in labor systems. However, different tasks and qualification requirements will develop so that there will be new job types.

4.2. Complexity drivers in digitalized work systems

Although the interviewees did not all have the same occupational backgrounds, the overall analysis indicates that similar points can be generalized from the single interviews. In total, the analysis results show that a considerable development towards very adaptive procedural and structural organizations is required for meeting the environmental demands and to evolve to customer-focused production procedures. This development is made possible through digitalization of manufacturing and service procedures. Nevertheless, this causes a higher complexity level.

Regarding the organizational dimension of process organization, the experts answered that complexity increases due to more customer involvement into faster work procedures. This calls for reactive and adaptive planning processes. Since digitalization can be used to initiate reactions in a company in real-time via using novel technology, this remarkably raises the dynamics of the whole organization. The change towards worldwide active value creation networks with several levels also increases the overall complexity. Since labor systems are more digitalized, the amount of human machine interfaces and diverse system solution rises. In this context, parallel IT-systems may increase the complexity level because of several maintenance and data processing steps. This may in turn cause data redundancies and problems of lacking data hygiene. Furthermore, media changes during work processes were mentioned as roots of complexity, which is digitally solvable via solutions for integrated systems. Figure 7 summarizes the drivers in the dimension of process organization and their number of mentions. Since every expert could mention several or no complexity drivers for every organizational dimension, the number of mentions does not necessarily have to sum up to the number of experts.
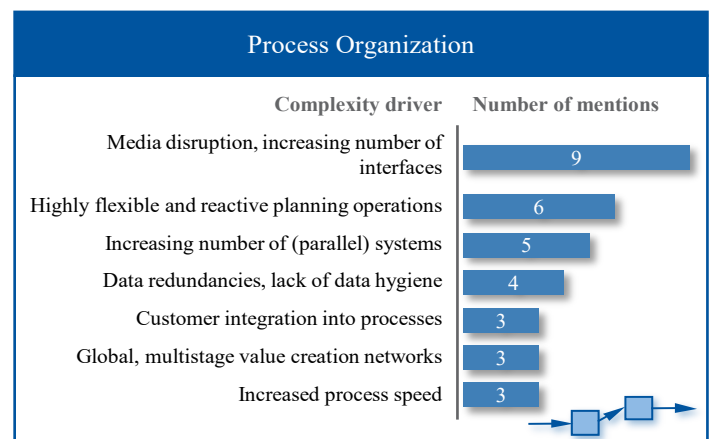
Figure 7: Complexity drivers in digitalized work environments concerning process organization, including their number of mentions

To develop such adaptive enterprise organizations, a shift of traditional hierarchic organizations towards the direction of flatter hierarchies is necessary. Almost all interviewees agreed on the point that agile work methods and project work would be most frequently found in the course of digitalization. In this context, complexity with respect to the organization increases, when a development towards trans-disciplinary work in (worldwide) structures is needed. This causes more coordination procedures. To work in smaller and agile cooperative structures also causes that employees need to be given more autonomy. The targeted final state may make it easier to fulfill market demands. However, the procedure of changing the culture in an organization was seen as raising the complexity level in a first step. If work persons are not empowered and qualified in diverse dimensions, occurrences with conflicting goals (polytely) can imply complexity for single operators. Complexity drivers and their number of mentions regarding the dimension of structural organization are illustrated in Figure 8.
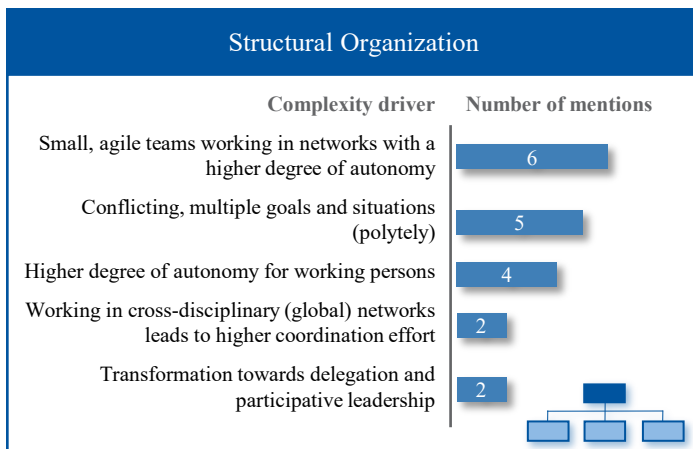
Figure 8: Complexity drivers in digitalized work environments concerning structural organization, including their number of mentions

For the organizational dimension of technology, the experts named rising requirements for IT security as most important aspect in digitalized labor systems. Systems that are not compatible, missing technological standards and lacking common sets of protocols increase complexity during digitalization. These aspects gain importance, since harmonization problems happen because of systems that have evolved over a long time span. Furthermore, much higher information amounts need to be captured in real-time with software to analyze the data. Faster innovation cycles with respect to software and hardware cause more dynamics of introducing novel systems. Figure 9 summarizes the drivers concerning technological aspects and their number of mentions.
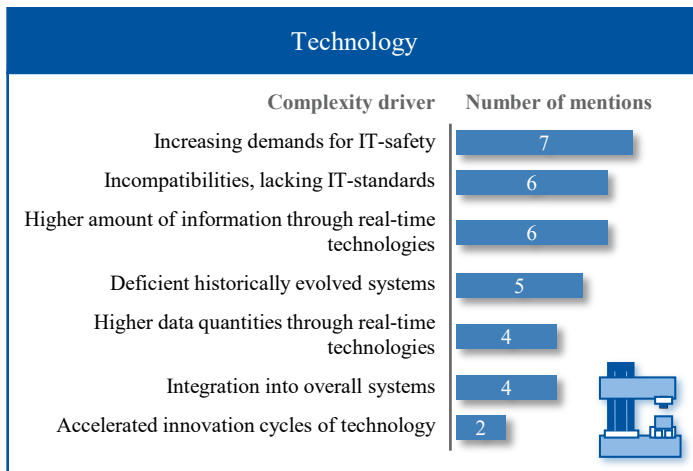
Figure 9: Complexity drivers in digitalized work environments concerning the technology, including their number of mentions

Digitalization will also have an effect on the conditions of work and will make new methods of digital communication possible. However, mentioned by the interviewees, only communicating digitally could not totally replace interpersonal communication and cause communicational complexity, because the semantic meaning of expressed phrases is mainly strengthened with face expressions or gestures. For this reason, hybrid communication ways should be considered. The majority of the interviewees stated that the digital transformation in general changed jobs into the direction of tasks that have to manage more complexity. Although monotonous tasks are likely to be replaced by automation, jobs that

demand analytical information processing, creative thinking and the capability of solving problems combined with specialist skills are more and more required. This causes that employees need transdisciplinary abilities (for instance, qualifications such as self-organization or multidimensional work with respect to diverse disciplines). Task complexity occurs as well, if novel digital systems are used and if more digitalized information has to be captured mentally in a shorter time. Lots of interviewees stressed on the point that this facet already produced trouble for employees in many cases. If the rising amount of operator machine interfaces was not created in a way that they can be used easily, a higher level of complexity would be a consequence for the employees. Nevertheless, single interviewees mentioned that the use of digitalization could at the same time be employed to make work easier. For example, work instructions can be provided during the task fulfillment. A further point is that the digital transformation could cause more task quantities. In this context, scan devices facilitate to introduce inventory documentation jobs into work procedures. All in all, a major complexity cause in digitalized labor systems was identified in the point that digitalization frequently is not very well established in management hierarchies. Particularly for worldwide operating organizations, the significance of different laws and agreements for data protection in diverse nations has extensive consequences. Figure 10 illustrates complexity drivers in the dimension of working conditions.
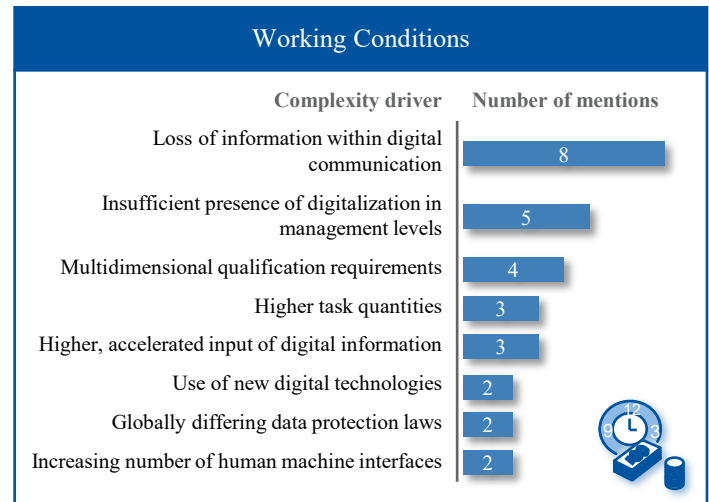
Figure 10: Complexity drivers in digitalized work environments concerning working conditions, including their number of mentions

With respect to the organizational dimension of products, the interviewees named individualized products and more product variants as main complexity drivers. In general, digitalization causes product and service portfolio complexity because sensor devices, software and connecting elements are important parts of a product. This implies that designers of goods and machines will have to take into account that the final product is going to be used in a web-based context. These changes remarkably enhance feasible functions of the product but also cause more software dependency. The radical development from typical products towards hybrid production necessitates that the whole life-cycle of a produced good will need be regarded during the development of the product. Since producers and customers will not only interact once when selling a smart product, they will likely be in close contact during the whole life-cycle when processing data packages

are transferred. This provides numerous novel options and makes a product more complex. For instance, individualized services or upgrades may be suggested as well as novel business models that are based on data analyses will come into focus.

In addition, very iterative and accelerated product development procedures in which clients work with the contractor together were mentioned as complexity drivers. This process of co-development between companies is essential, since a considerable gain in IT-skills and system thinking develops on the client side. In particular within the service sector, digital opportunities are employed to integrate the customer (business-to-customer relation) into the development of new services and into the service delivery. In the context of self-service, digital procedures make it possible to outsource tasks to the customers and to let them directly take part in the service fulfillment. In analogy to the organizational dimension of technology, the interviewees mentioned that digitalization fastens the release cycles for new innovations or products so that the complexity level of the products increases. Figure 11 summarizes the complexity drivers concerning the product.
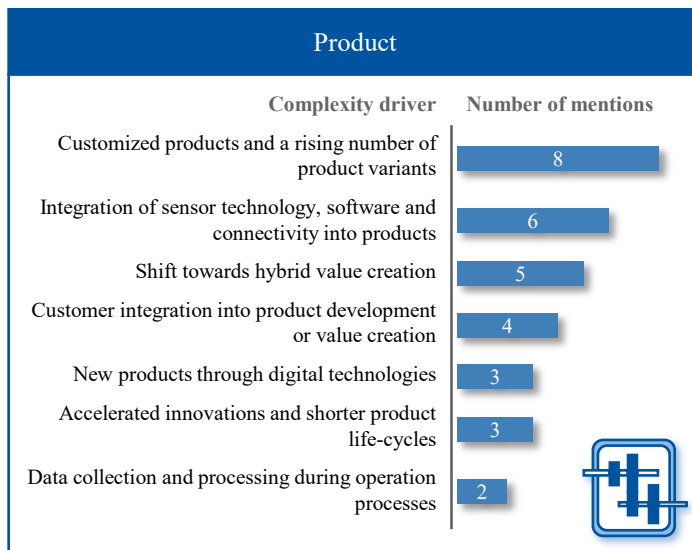
Figure 11: Complexity drivers in digitalized work environments concerning the product, including their number of mentions

Concerning the organizational dimension of personnel, the interviewees said that qualification requirements for more digital understanding and skills are the most important aspect for handling the complexity in digitalized labor systems. However, the operators very frequently do not have the needed competencies. Moreover, a major gap of IT understanding between employees can be detected. Also, digitalization offers the possibility to work in mobile and flexible work conditions. This allows for balancing the private life and the job. Nevertheless, this also can cause very complex situations for the staff, because contrary requirements could evolve from the trend that makes the separation between the job and free time more difficult. The employment of mobile gadgets also implies that an employee is always available. This can be the cause of a dependency on technology to some extent. Additionally, severe problems were mentioned that concern the aspects measuring and controlling the performance of the personnel in labor environments. If new places of work are conceptualized for digital work systems, the complexity of

planning increases due to missing approaches for strain measurement. Furthermore, there is a lack of rules and relief concepts for digital work. Ultimately, many experts stated that digitalization offers novel job forms, such as crowdworking. This may have major implications for single persons or whole organizations. In addition, this can also imply complex problems for societies. These could be related to topics of social exclusion or regulatory problems for whole systems due to lacking contributions to social insurances. Figure 12 summarizes the complexity drivers concerning the personnel.
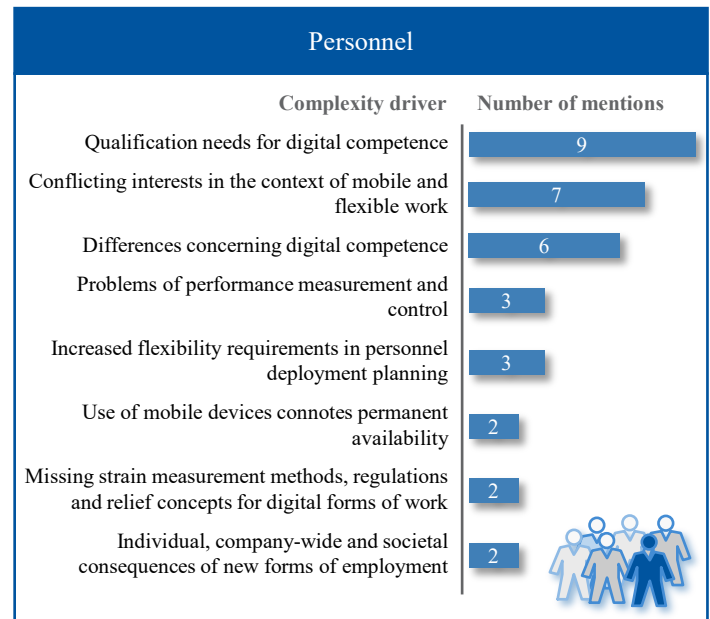
Figure 12: Complexity drivers in digitalized work environments concerning the personnel, including their number of mentions

4.3. Measures to handle complexity in digitalized work systems

The interviews were also used to explore measures to manage complexity in digitalized work systems. Adequate training of employees was most frequently mentioned by the experts. On the one hand, this implies qualifications for imparting methodical and expert knowledge on new technologies, processes and work equipment. On the other hand, training includes enhancement of social and communicative competences, which are targeted e.g. at avoiding or solving communication problems and conflicts in cooperative decision-making processes. Training was also considered as necessary to support employees coping with mental and physical strain as a consequence of changing or increasing job requirements.

The second most frequently mentioned measure was participation. Employees that are affected by digitalization-induced change should not only receive adequate training, but also should be involved in change-related development and decision-making processes at an early stage. Participation in pilot projects for implementation of new technologies and communication tools was mentioned as an example in order to enhance acceptance for digitalization-related innovations and to ensure economic efficiency. Additionally, participation-oriented pilot projects were recommended to design user-centered support systems and identify policies or required formalities on the following aspects: workload, working time, occupational safety and data protection.

In this context, work in agile teams was, for example, seen as form of work that especially requires policies concerning these aspects.

Development and use of algorithms for automatic data analyses were proposed as further measures to handle complexity. Thus, increasing data volume and diversity in the course of digitalization could be limited to the information and key figures required for the specific application. These and similar measures aim to relieve employees from analyzing and processing large amounts of data. The resources saved can be used for the accurate interpretations of data, which still need to be carried out by employees.

To harmonize different systems and their interfaces, the development and introduction of reference architectures for IT systems were mentioned. A modular structure of IT systems enables further development and easier linking of new technologies to the overall IT system. The definition of IT standards and the standardization of the IT structure help to avoid interface problems and to ensure the interoperability of the systems used. According to the experts, simple operation of products and systems is an equally important measure, e.g. through intuitive user guidance and ergonomic interface design. While this results in potentially more demanding programming, there are various benefits: avoidance of handling problems and errors, reduction of operating effort, and reduction of qualification effort during introduction and application of newly introduced technologies and systems.

The experts also mentioned the development of more transparent business processes to clarify responsibilities and avoid misunderstandings in work instructions due to inaccurate communication or misleading information. This also aims at a solid understanding of processes of all employees and managers as well as better overall communication.
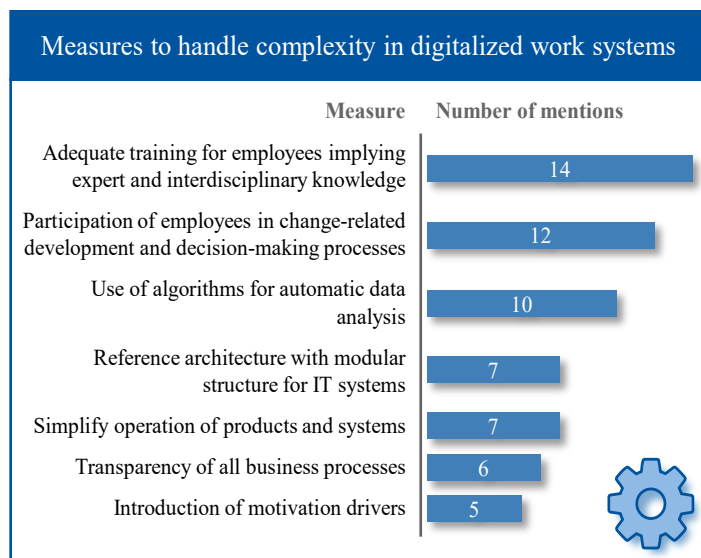
Figure 13: Measures to handle complexity in digitalized work environments including their number of mentions

Finally, the introduction of motivation drivers includes measures that can or should contribute indirectly to mastering the resulting complexity by promoting the willingness of employees to change. Motivation drivers could be e.g. suitable break concepts, more varied activities or "gamification" approaches

combined with monetary incentives. Figure 13 illustrates the categorized measures to manage complexity in digitalized work systems as a result of the interview analysis.

5. Implications for cooperative forms of work

The results of the expert interviews have far-reaching implications for cooperative forms of work. In this context, the term cooperative form of work is employed to express that people work in a team or group structure and thus exhibit interaction amongst each other whilst performing their job. This includes a project team as well as a work group in a production environment.

In the literature, there is already extensive research on team or group work that describes the interrelations of relevant variables, e.g. in team models (e.g. [54][55][56][57]; overviews can be found in [19][58][59]). These models often focus on teamwork in the context of innovation and do not consider complexity parameters. Only a few conceptual and non-empirically validated models (e.g. [60]) contain "task complexity" in the context of cooperative forms of work as a single complexity-related variable. The expert interviews revealed that the ratio of cooperative forms of work is expected to grow in the field of production. A literature review of existing models for cooperative forms of work shows that currently there is no interrelated and empirically assessed model for group work in production and thus no consideration of complexity in a superordinate model (see e.g. [59]). Therefore, further research is necessary in this field. Such a model could be developed conceptually as an input-process-output model (see e.g. [19]). This approach enables to investigate the effects of digitalization and complexity as input factors on process and output dimensions. The interview results offer diverse aspects that directly refer to cooperative work. These aspects could be integrated into the development of such a new model. By doing so, the explorative results of the interviews, which originally did not aim at confirming existing theories, could be used to develop hypotheses, new theories and an overall framework for analyzing cooperative forms of work.

When regarding the dimension of structural organization in the interviews, a transformation towards cooperative forms of work was the most frequently mentioned complexity driver for digitalized work environments. Small and agile teams working in network structures with a high degree of autonomy were considered by the experts as a predominant form of work in the late stage of digital transformation processes. This can be seen as a continuation of the trend observed by the experts, which already describes a change towards cooperative forms of work over the past 10 years. These results strengthen that team and group research will increasingly come into focus – particularly in the course of digitalization. This conclusion is also reached by Trompisch [61].

Hence, there is a need for further research on group processes on the one hand and to explicitly focus on the effects of digitalization on cooperative forms of work on the other hand. In addition, the results of the expert interviews showed that digitalization does effectively imply complexity-driving aspects that have an impact on working persons. For this reason, it is also necessary to apply complexity considerations to cooperative forms of work and to investigate complexity in greater depth in this context.

Assuming that humans do not have a stable input-output relationship [62] supports the consideration of human behavior and perception in complexity studies, e.g. with regard to emergent events [63]. For example, a user of a new digital system could use the system for a completely different purpose compared to the original goal of the system. If this individual view was transferred to a work group consisting of several individuals, it seems trivial that the interaction of several people can especially be classified as complex. From a system theory point of view, a work group, e.g. a team of workers in production, can be seen as a complex structure or subsystem within the complex company system. In this context, the characteristics of complexity from system theory that were derived in the first section of this article can be considered and transferred to the system "work group".

On the whole, it is essential to consider factors such as digitalization and complexity in a differentiated manner. Complexity may contain positive and negative aspects concerning both organizational levels (e.g. [17] [64]) and individual levels. On the company level for instance, a larger variety within a product portfolio – and thus greater complexity – can be used to respond to changing market requirements and thereby to achieve competitive advantages. However, an excessive broad product range also bears the risk of strong cross-subsidization effects and thus economic losses (see e.g. [65] [66]). As the focus of this article lies on work design, the individual level concerning work persons will be discussed in more detail.

For an individual person and in analogy to the Yerkes-Dodson law, Dalhöfer and Prieß [15] conceptually describe the influence of task complexity on the individual performance and motivation in the context of indirect business processes with an inverted u-shape parabola. The decline in performance for tasks that are less complex than the optimum task complexity is explained with demotivation due to highly monotonous activities. The area near the parabola's apex describes an optimum of task complexity in which the worker is challenged in a positive sense by the work task and thereby is highly motivated. The decline in performance to the right of the complexity optimum is caused by the employee being overloaded by the work task. Regarding the individual skills of the working person, the task becomes too complex so that work performance and motivation decrease. Since complexity is not clearly measurable due to the subjectivity of its perception, and a slight change in complexity does not necessarily lead to immense changes in performance, an interval for acceptable task complexity should be defined. Dalhöfer and Prieß [15] use the term "Eu-Stress" (good stress) to define the term "Eu-complexity" (good complexity). This term describes the range around the complexity optimum in which a high performance is achieved. In analogy, they describe the area of overload as "Dys-complexity". Figure 14 illustrates these conceptual considerations.

It is important to note that the curve trace of the parabola and therefore also the interval boundaries vary in analogy to the stress-strain concept [37][38] between different persons.

In general, the individual parabolic curve can be explained either by constitutional characteristics, disposition characteristics, qualification and competence characteristics or by adaptation characteristics of a worker [19]. Constitutional characteristics (e.g. gender, body type, nationality etc.) are invariable in a person's life-

cycle so that it is difficult to change a person's ability to cope with the required task complexity. This also applies to disposition characteristics (e.g. personality, age, intelligence, body weight etc.) which can be changed over time – even though not directly by the employee. However, qualification and competence characteristics (e.g. experience, knowledge, skills, education, competence etc.) can be varied through long, medium and short-term work structuring processes. For instance, dedicated training can help to ensure that a work task is no longer perceived as too complex, i.e. more complex than the complexity optimum of the parabolic curve, by the individual worker. In this context, Dalhöfer and Prieß [15] regard the line manager as key figure in complexity management, as the complexity optimum could individually be designed for every employee. Finally, short-term work structuring measures may lead to a variation of adaptation characteristics (reaction to energetic-effector, informational-mental or emotional stress; e.g. strain, fatigue, satisfaction etc.) and to a reduction of subjectively perceived task complexity, e.g. through supporting information systems.
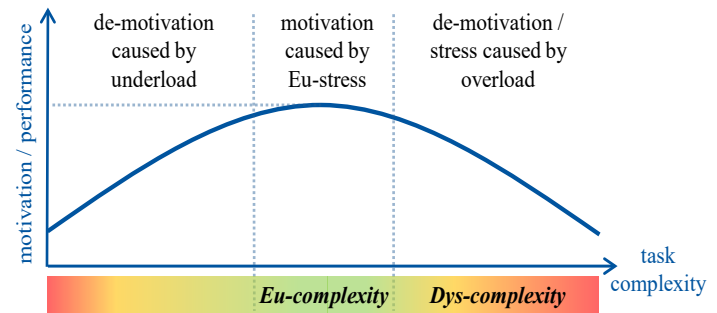
Figure 14: Performance, motivation and task complexity according to Dalhöfer and Prieß [15]

This article makes the first attempt to extend these conceptual considerations towards cooperative forms of work and hence further developing the concept of Dalhöfer and Prieß [15]. For the example of a work group, it is assumed that the original concept for the individual worker still is valid. Thus, each work group has individual complexity parabolas for each person in this group. The individual complexity parabolas are now transferred into a three-dimensional diagram, which contains the individual persons of the work group as a third axis. This creates a three-dimensional graph that can be covered with an envelope curve. In this article, the concept of a "task complexity mountain range" of a work group is introduced as a metaphor. This concept is illustrated in Figure 15.

It is evident that although the task complexity mountain range is composed of the individual graphs, the interval of a complexity optimum also exists for the work group as a whole. The interval appears through the envelope curve, which covers the individual graphs. However, this optimum cannot clearly be determined due to the different apex points of the individual parabolas and can only be described as an interval in which the abilities of the group members complement each other. Within this interval, the overall complexity of the group's work task needs to be distributed adequately according to the individual complexity parabolas of the particular group members. However, identifying the optimum is already a complex task, as the optimum may shift dynamically depending on the group structure and internal group processes.
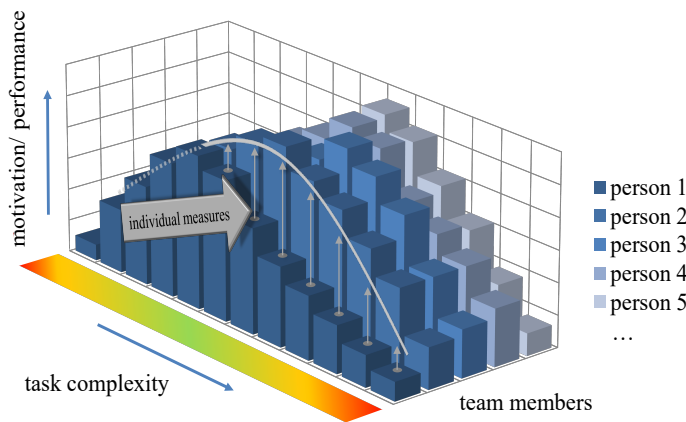
Figure 15: Development of the individual task complexity concept of Dalhöfer and Prieß [15] towards a "task complexity mountain range" for cooperative forms of work; including individual coaching measures to adjust the complexity curve of person 1

In the context of cooperative forms of work, the concept of leadership in complexity management of Dalhöfer and Prieß [15] is even more important: within the group, the line manager has to respond to individual working persons and individually adjust the complexity optimum. In Figure 15, the arrow labelled "individual measures" indicates that through dedicated training the complexity optimum of a particular person – and in a reverse conclusion of the entire work group – can be shifted towards more complex work tasks. For instance, this is achievable by systematically enhancing the individual qualification and competence characteristics. In addition, the work group as a whole can be enabled by the line manager to handle more complex tasks independently and to react adaptively to unpredictable situations. Transformational leadership, for example, can achieve this by strengthening the intrinsic motivation of employees through transforming their values and ambitions as well as supporting the individual development [67]. ElMaraghy et al. [17] point out that collaboration is able to lead to adaptive and creative forms of work that can handle complexity very effectively.

However, the expert interviews have shown that the transformation towards delegating and cooperative management initially increases complexity – and that sufficient complexity handling can only be achieved in the target state of the transformation process. Also, participation was identified in the expert interviews as most frequently mentioned measure for managing complexity in digitalized work systems. This should consequently be a subject of research in the context of management of work groups. With regard to management processes, goal-setting theory [68] should also be examined in greater depth. For instance, the effect of conflicting goals on work processes of groups with a high degree of autonomy should be inquired. The expert interviews have indicated that a higher degree of autonomy and conflicting objectives increase complexity for working persons and that these aspects are intensified in the course of digitalization. Therefore, further research is needed to determine how a good target quality can be achieved for work groups. Overall, especially in the context of digitalization, an intensive investigation of cooperative leadership processes in group work is particularly relevant in order to develop strategies for complexity management for work groups.

According to Schuh and Riesener [65], corporate complexity management strategies for optimizing the product portfolio not only include measures to handle complexity, but also measures to avoid and reduce complexity. These could also be applied to complexity management of group tasks. Thus, there are two further strategies for designing the "optimal" task complexity for work groups: on the one hand, task complexity can be reduced via redesigning the work process (corrective approach), if the level of task complexity lies within the interval that decreases work performance and motivation due to overstraining conditions. On the other hand, task complexity is initially avoidable in terms of preventive work design.

The degree of optimal task complexity of a work group or team also always depends on the operational conditions. As an example, resources provided, support from other organizational units or digital support tools determine how well complex tasks can be solved by the work group. However, the use of digital support systems also requires a differentiated consideration: the expert interviews have shown that digital assistance systems are able to reduce complexity if they are designed in a user-centered manner. In addition to participatory development processes, which have the potential to increase acceptance of new systems, there are instructions for ergonomic designs of such systems in norms, e.g. in EN ISO 9241 [69]. Nevertheless, the interview results also showed that perceived complexity of a work task may increase due to user-unfriendly designs of digital work tools. In addition, the perceived complexity depends on the life-cycle phase of the support system. It is important to distinguish whether the system is currently being introduced in the company, adapted or already established. Even though a new digital support system may be intended to make complexity manageable in a work system, the introduction of a new system will initially generate a (subjective) increase in complexity. Initially, the system is unfamiliar and requires training. This phase is followed by an iterative phase in which the employees become familiar with the use of the system. At the same time, possible system errors are eliminated and the system is adjusted to the needs of the employees. In the third phase, the "steady state", the intended subjective complexity reduction of the work task is sustainably realized. The interviews indicated that a participatory development and testing procedure of such systems facilitates to reach this target state in a shorter period of time. To accompany this, appropriate training for the use of the systems should support the introduction process. Figure 16 illustrates the settling process of digital support systems with regard to perceived task complexity.

Overall, it is apparent that considerations relating to complexity and digitalization in the context of cooperative forms of work offer multifaceted research questions. In order to contribute to bridging the identified research gaps, a holistic empirical analysis of the interrelationships in a structural equation model is suggested (see e.g. [59]). Further research steps and a general discussion are considered in the following section.

6.   General discussion and future research steps

Diverse disciplines conduct research on the construct of complexity with specific points of interest and with different methods. This article explored complexity drivers that arise through digitalization. The qualitative content analysis of the

interviews identified various complexity drivers in digitalized labor systems in all six organizational dimensions. They may be employed to enhance current complexity driver systematizations that already exist in literature.
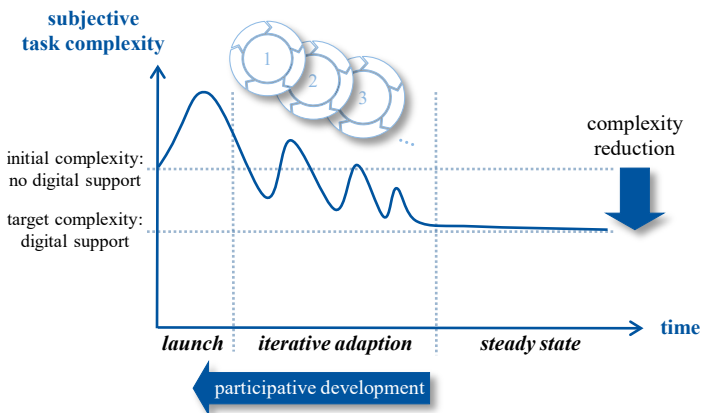
Figure 16: Settling process of digital support systems with respect to subjective task complexity

Since digitalization changes labor systems and whole industrial sectors, it is crucial to gain a deep understanding of which factors raise the level of complexity not only from a technical but at the same time from a work design perspective. The findings show that the supply of lots of information as well as diverse options of displaying them may imply complexity for employees in a negative connotation. This is why it is crucial to agree on guidelines for company-internal communication. A lower level of complexity is obtainable by reducing the number of communication systems. Moreover, information should be organized and supplied in a manner that makes is easy for operators to spot the information that is needed by them to perform their jobs.

However, it is frequently based on the viewpoint whether a certain topic should be seen as a driver of complexity. In this connection, the complexity drivers belong to various points of view which relate to whole systems as well as to subjectively perceived complex incidences from the perspective of an employee. Regarding the operator's perspective, a new digital support system could reduce complexity for an operator, if it was well designed. Yet, complexity rises on a system level, as the amount of system elements and system maintenance procedures increases.

Furthermore, it should be mentioned that the organizational dimensions cannot be examined separately, since they are interdependent in diverse ways. This is why it is especially complicated to assign drivers to particular dimensions. The validity of the results of the analysis procedure can be increased by enlarging the sample size of interviewed experts. This could enable cross-references with the results of the interviews as well as complexity drivers could be identified that are specific for certain industrial sectors. All in all, the conducted method of semi-structured interviews with experts enabled to qualitatively analyze the answers of the interviewees. Therefore, this method was adequate for the first research step of exploration. If conducted with sufficient many persons for the coding procedure, the method may offer reliable results. However, the inter-code-reliability of

the results can be enhanced and quantitatively judged by repeating the analysis procedure with more than two coders. Further research steps will use the findings to develop theories and to conduct ongoing quantitative empirical studies.

This article discussed the findings in the context of cooperative forms of work and offers initial concepts that will be examined with empirical research methods in future. The studies will have to show whether the concept of the task complexity mountain range should be retained or adapted. Furthermore, the effects of digitalization on output and process measures of cooperative forms will be assessed.

Eventually, complexity should not be seen unequivocally as a thing that should be avoided at all costs. Actually, complexity is also the basis for market leadership, if enterprises have an expertise in reproducing the demanded market complexity. However, the important question is how much capacity is required to fulfill these requirements so that too complex organizations can be prevented. This especially leads to the overall challenge of how work systems should be designed to provide rather challenging than overstraining work conditions for employees.

7. Conclusion

This article examined which complexity drivers in labor systems are caused by digitalization. Initially, diverse complexity definitions from relevant research disciplines were presented. General properties of complexity in theory were identified. Moreover, the used method of semi-structured interviews with experts for exploring the complexity drivers was outlined. The group of interviewees was made up of 23 experts with positions in German economy and science. The findings show that the digital transformation could remarkably alter work processes and whole industrial sectors. In this connection, diverse complexity drivers in digitalized labor environments were explored and clustered into six organizational dimensions. The findings can be employed to point out far-reaching consequences of digitalization, because the six organizational dimensions and the topics of complexity and digitalization are mutually interrelated. Further, measures for handling complexity in digitalized work systems that were extracted from the interviews. The findings were discussed in the context of cooperative forms of work as well as existing concepts for explaining complexity impacts were transferred to this topic. In this context, the concept of the task complexity mountain range for cooperative forms of work was developed. Future research steps will attempt to empirically assess the complexity in digitalized labor systems. In this context, it will be analyzed which effects complexity and digitalization have on cooperative forms of work in production systems. The results may be used for developing additional measures to handle complexity in digitalized work systems to those that were presented in this article.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

Appendix

Appendix 1: Questionnaire guide for the semi-structured interviews

1. Name, company, company size, function and tasks in company

2.1 How has digitalization changed your company and business in the last 10 years?

2.2 How will digitalization change your company and business in the next 10 years?

3. How does digitalization influence the level of complexity in the 6 dimensions of work organization? Which aspects of digitalization cause a higher level of complexity?

a) Process organization (e.g. production planning, material flow, information flow, innovative assembly concepts)

b) Organizational structure (e.g. hierarchy levels, delegation of responsibility), decision-making powers)

c) Technology (e.g. production technology, information technology)

d) Working conditions / incentive systems (e.g. employment contracts, work safety, working hours, remuneration)

e) Product (e.g. product design, product structure, new business models)

f) Personnel (e.g. cooperation, communication, leadership, personnel planning, qualification)

4. What measures can be developed to manage the complexity caused by digitalization?

References

[1] B. A. Latos, M. Harlacher, P. M. Przybysz, S. Mütze-Niewöhner, "Transformation of Working Environments Through Digitalization: Exploration and Systematization of Complexity Drivers" in IEEE International Conference on Industrial Engineering and Engineering Management, Singapore, Singapore, 2017. http://dx.doi.org/10.1109/IEEM.2017.8290059

[2] G. Schuh, M. Riesener, J. Kantelberg, N. Steireif, "Transmission of software-related agile mechanisms of action towards product development processes for technical products" in IEEE International Conference on Industrial Engineering and Engineering Management, Singapore, Singapore, 2017. https://doi.org/10.1109/IEEM.2017.8290192

[3] C. Châlons, "Die Rolle der IT als Enabler für Digitalisierung", In: F. Abolhassan (Ed.), Was treibt die Digitalisierung? Warum an der Cloud kein Weg vorbeiführt, Wiesbaden, Germany: Springer, 27-37, 2016. http://dx.doi.org/10.1007/978-3-658-10640-9

[4] C. Aichele, M.. Schöneberger, Die Digitalisierung der Energiewirtschaft: Potenziale und Herausforderungen der IKT-Branche für Utility 4.0, In: O. D. Doleski (Ed.), Herausforderung Utility 4.0. Wie sich die Energiewirtschaft im Zeitalter der Digitalisierung verändert, Wiesbaden, Germany: Springer, 501-530, 2017. https://doi.org/10.1007/978-3-658-15737-1

[5] W. Widuckel, K. de Molina, M. J. Ringlstetter, D. Frey, Arbeitskultur 2020. Herausforderungen und Best Practices der Arbeitswelt der Zukunft, Wiesbaden, Germany: Springer, 2015. https://doi.org/10.1007/978-3-658-06092-3

[6] W. M. Walter, "Der Einfluss der Digitalisierung auf die Organisation eines Unternehmens", In: O. D. Doleski (Ed.), Herausforderung Utility 4.0. Wie sich die Energiewirtschaft im Zeitalter der Digitalisierung verändert, Wiesbaden, Germany: Springer, 227-248, 2017. https://doi.org/10.1007/978-3-658-15737-1

[7] W. Bauer, S. Schlund, "Wandel der Arbeit in indirekten Bereichen – Planung und Engineering", In: H. Hirsch-Kreinsen, P. Ittermann, J. Niehaus (Ed.), Digitalisierung industrieller Arbeit. Die Vision Industrie 4.0 und ihre sozialen Herausforderungen, 1st ed., Berlin, Germany: Nomos, 53-70, 2015. https://doi.org/10.5771/9783845263205-1

[8] Institut DGB-Index Gute Arbeit, DGB-Index Gute Arbeit. Der Report 2016. Wie die Beschäftigten die Arbeitsbedingungen in Deutschland beurteilen, Berlin, Germany: PrintNetwork pn / ASTOV Vertriebsgesellschaft mbH, 2016.

[9] O. Stettes, Arbeitswelt der Zukunft. Wie die Digitalisierung den Arbeitsmarkt verändert, Cologne, Germany: IW Medien, 2016.

[10] H. Hirsch-Kreinsen, "Digitization of industrial work. Development paths and prospects" Journal for Labour Market Research, 49(1), 1-14, 2016. https://doi.org/10.1007/s12651-016-0200-6

[11] L. Bertschek, B. Dworschak, P. Meil, T. Niebel, J. Ohnemus, T. Vetter, H. Zaiser, Arbeitsmarkt 2030 - Digitalisierung der Arbeitswelt. Fachexpertisen zur Prognose 2016, Munich, Germany: Economix Research & Consulting, 2016. https://doi.org/10.3278/6004559w

[12] S. Schlund, B. Pokorni, Industrie 4.0 - Wo steht die Revolution der Arbeitsgestaltung? Ergebnisse einer Befragung von Produktions-verantwortlichen deutscher Unternehmen, Ulm, Germany: Ingenics AG, 2016.

[13] B. A. Latos, M. Harlacher, M. El-Mahgary, D. Götzelmann, P. M. Przybysz, S. Mütze-Niewöhner, C. Schlick, "Komplexität in Arbeitssystemen: Analyse und Ordnung von Beschreibungsansätzen aus unterschiedlichen Disziplinen" in 63th Frühjahrskongress der Gesellschaft für Arbeitswissenschaft FHNW, Brugg-Windisch, Switzerland, 2017. https://doi.org/10.18154/RWTH-2017-04217

[14] M. Schwemmle, P. Wedde, Digitale Arbeit in Deutschland: Potenziale und Problemlagen, Bonn; Germany: bub Bonner Universitäts-Buchdruckerei, 2012

[15] J. Dalhöfer, M. Prieß, "Führung im Komplexitätsmanagement: Zusammenhang zwischen Komplexität, Motivation und Stress" ZWF Zeitschrift für Wirtschaftlichen Fabrikbetrieb, 107(1-2), 87-93, 2012. https://doi.org/10.3139/104.110666

[16] A. Hoeschen, "Complexity-based distribution of value added in multinational production systems" (Thesis in German) Ph.D. dissertation, RWTH Aachen University, Germany, 2015.

[17] W. ElMaraghy, H. ElMaraghy, T. Tomiyama, L. Monostori, "Complexity in engineering design and manufacturing" CIRP Annals – Manufacturing Technology 61(2), 793-814, 2012. https://doi.org/10.1016/j.cirp.2012.05.001

[18] R. Riedl, Strukturen der Komplexität: Eine Morphologie des Erkennens und Erklärens, Berlin, Germany: Springer, 2000. https://doi.org/10.1007/978-3-642-56946-3

[19] C. Schlick, R. Bruder, H. Luczak, Arbeitswissenschaft, 4th ed., Heidelberg, Germany: Springer, 2018. https://doi.org/10.1007/978-3-662-56037-2

[20] C. R. Shalizi, "Methods and Techniques of Complex Systems Science", In: T. S. Deisboeck, J. Y. Kresh (Ed.) Complex systems science in biomedicine, New York, USA: Springer, 49-131, 2006.

[21] H. A. Simon, The sciences of the artificial, 3rd ed., Cambridge, USA: MIT Press, 1996.

[22] M.-O. Blockus, "Komplexität in Dienstleistungsunternehmen: Komplexitätsformen, Kosten- und Nutzenwirkungen, empirische Befunde und Managementimplikationen", Ph.D. dissertation, Universität Basel, Switzerland, 2010, 1st ed., Wiesbaden, Germany: Gabler Research, 2010. https://doi.org/10.1007/978-3-8349-8958-1

[23] L. von Bertalanffy, "Vorläufer und Begründer der Systemtheorie" in System Theory Research and Information, (Systemtheorie Forschung und Information), R. Kurzrock, Berlin, Germany: Colloquium-Verlag, 17-28, 1972.

[24] J. B. Rosser, "Computational and Dynamic Complexity in Economics", In: J. B. Rosser (Ed.), Handbook of Research on Complexity, Cheltenham, UK: Edward Elgar Publishing, 2009. https://doi.org/10.4337/9781781952665

[25] G. Schuh, U. Schwenk, Produktkomplexität managen. Strategien, Methoden, Tools, München, Germany: Carl Hanser, 2001.

[26] H. Ulrich, G. J. Probst, Anleitung zum ganzheitlichen Denken und Handeln. Ein Brevier für Führungskräfte, Bern, Switzerland: Haupt, 1988.

[27] A. V. Deshmukh, J. J. Talavage, M. M. Barash, "Complexity in manufacturing systems, Part 1: Analysis of static complexity" IIE Transactions, 30(7), 645-655, 1998.

[28] O. Kuzgunkaya, H. A. ElMaraghy, "Assessing the structural complexity of manufacturing systems configurations" International Journal of Flexible Manufacturing Systems, 18(3), 145-171, 2006.

[29] S. Sivadasan, J. Efstathiou, A. Calinescu, L. H. Huatuco, "Advances on measuring the operational complexity of supplier–customer systems"

European Journal of Operational Research, 171(1), 208-226, 2006. https://doi.org/10.1016/j.ejor.2004.08.032

[30] S. Mattsson, P. Gullander, U. Harlin, G. Bäckstrand, Å. Fasth, A. Davidsson, "Testing Complexity Index – a Method for Measuring Perceived Production Complexity" in 45th CIRP Conference on Manufacturing Systems, Athens, Greece, 394-399, 2012. https://doi.org/10.1016/j.procir.2012.07.068

[31] L. Zeltzer, V. Limère, H. van Landeghem, E.-H. Aghezzaf, J. Stahre, "Measuring complexity in mixed-model assembly workstations" International Journal of Production Research, 51(15), 4630-4643, 2013. https://doi.org/10.1080/00207543.2013.783246

[32] D. Dörner, H. W. Kreuzig, F. Reither, T. Stäudel, Lohhausen. Vom Umgang mit Unbestimmtheit und Komplexität, Bern, Switzerland: Hans Huber Verlag, 1983.

[33] D. Dörner, Die Logik des Misslingens, Strategisches Denken in komplexen Situationen, 10th ed., Hamburg, Germany: Rowohlt Verlag, 2011.

[34] A. Kluge, Wissenserwerb für das Steuern komplexer Systeme, Lengerich, Germany: Wolfgang Pabst Science Publishers, 2004.

[35] H. Willke, Systemtheorie 1: Grundlagen, 5th ed., Stuttgart, Germany: UTB, 1996.

[36] N. Luhmann, Soziale Systeme. Grundriß einer allgemeinen Theorie, Frankfurt a. M., Germany: Suhrkamp Verlag, 1987.

[37] W. Rohmert, "Das Belastungs-Beanspruchungs-Konzept" Zeitschrift für Arbeitswissenschaft, 38(4), 193-200, 1984.

[38] H. Luczak, "Untersuchungen informatorischer Belastung und Beanspruchung des Menschen" Fortschrittsberichte der VDI-Zeitschriften, 10(2), Düsseldorf, Germany: VDI-Verlag, 1975.

[39] M. W. McElroy, "Integrating complexity theory, knowledge management and organizational learning" Journal of Knowledge Management, 4 (3), 195-203, 2000. https://doi.org/10.1108/13673270010377652

[40] W. Vogel, R. Lasch, "Complexity drivers in manufacturing companies: A literature review" Logistics Research, 9(1), 1-66, 2016. https://doi.org/10.1007/s12159-016-0152-9, licensed under: http://creativecommons.org/licenses/by/4.0/

[41] K.-P. Schoeneberg, Komplexitätsmanagement in Unternehmen. Herausforderungen im Umgang mit Dynamik, Unsicherheit und Komplexität meistern, Wiesbaden, Germany: Springer Gabler, 2014. https://doi.org/10.1007/978-3-658-01284-7

[42] S. Bednar, V. Modrak, "Mass Customization and its Impact on Assembly Process' Complexity" International Journal for Quality Research, 8(3), 417-430, 2014

[43] N. Asadi M. Jackson, A. Fundin, "Drivers of Complexity in a Flexible Assembly System - A Case Study" Procedia CIRP, 41, 189-194, 2016. https://doi.org/10.1016/j.procir.2015.12.082

[44] N. Asadi, S. Javadi, "Key complexity dimensions in assembly systems with mixed-model assembly lines - a multiple case study" in 5th World Conference on Production and Operations Management, Havana, Kuba, 2016.

[45] S. Mattsson, M. Karlsson, P. Gullander, H. van Landeghem, L. Zeltzer, V. Limère, E.-H. Aghezzaf, "Comparing quantifiable methods to measure complexity in assembly" International Journal of Manufacturing Research, 9(1), 112-130, 2014. https://doi.org/10.1504/IJMR.2014.059602

[46] P. Mayring, Einführung in die qualitative Sozialforschung, 5th ed., Weinheim, Germany: Beltz, 2002.

[47] M. Pfadenhauer, Professionelles Handeln, Wiesbaden, Germany: VS Verlag für Sozialwissenschaften, 2005.

[48] P. Mayring, Qualitative Inhaltsanalyse: Grundlagen und Techniken, Weinheim, Germany: Beltz, 2010.

[49] P. Mayring, Qualitative content analysis: theoretical foundation, basic procedures and software solution. Klagenfurt, Germany: Beltz, 2014.

[50] J. Gläser, G. Laudel, Experteninterviews und qualitative Inhaltsanalyse, 3rd ed., Wiesbaden, Germany: VS Verlag für Sozialwissenschaften, 2009.

[51] A. Petz, "Produktivitätsbewertung und simulationsbasierte Gestaltung von wissensintensiven Dienstleistungen" Ph.D. dissertation, RWTH Aachen University, Germany, 2017. https://doi.org/10.2370/9783844057058

[52] A. Petz, S. Duckwitz, C. Schmalz, "Productivity of Services: An Explorative Study in the Electrical and Chemical Engineering Sector" Amfiteatru Economic, 14(Special No. 6), 635-652, 2012.

[53] S. Duckwitz, A. Petz, C. Schmalz, R. Jahnel, P. Lehmacher, C. M. Schlick, "Dienstleistungsproduktivität aus Unternehmensperspektive - Ergebnisse semistrukturierter Leitfadeninterviews", In: C. M. Schlick, M. Schenk, D. Spath, W. Ganz (Ed.), Produktivitätsmanagement von Dienstleistungen: Modelle, Methoden und Werkzeuge, Berlin, Germany: Springer Vieweg, 13-27, 2016. https://doi.org/10.1007/978-3-642-45071-6

[54] J. Wegge, F. Jungmann, S. C. Liebermann, K.-H. Schmidt, B. C. Ries, "Altersgemischte Teamarbeit kann erfolgreich sein: Empfehlungen für eine

ausgewogene betriebliche Altersstruktur" Sozialrecht + Praxis, 7, 433-442, 2011.

[55] M. Högl, "Teamarbeit in innovativen Projekten: Einflussgrößen und Wirkungen" Ph.D. dissertation, Universität Karlsruhe, Germany, 1998.

[56] M. Högl, H. G. Gemünden, "Teamwork Quality and the Success of Innovative Projects: A Theoretical Concept and Empirical Evidence" Organization Science, 12(4), 435-449, 2001. https://doi.org/10.1287/orsc.12.4.435.10635

[57] T.-M. Myskovszky von Myrow, "Teamwork during innovation projects: The effects of diversity, participation, and the organizational context" Ph.D. dissertation, RWTH Aachen University, Germany, 2018.

[58] P. M. Przybysz, "Empirische Untersuchungen der Auswirkung von Altersdiversität in Innovationsteams" Ph.D. dissertation, Ernst-Moritz-Arndt-Universität Greifswald, Germany, 2016.

[59] R. A. Stranzenbach, "Analysis and Design of Group Work in Integrated Production Systems" (text in German) Ph.D. dissertation, RWTH Aachen University, Germany, 2018. in press

[60] S. Tannenbaum, R. Beard, E. Salas, "Team building and its influence on team effectiveness. An examination of conceptual and empirical developments" Advances in Psychology, 82, 117-153, 1992. https://doi.org/10.1016/S0166-4115(08)62601-1

[61] P. Trompisch, "The implications of Industry 4.0 on the future of work" (article in German with an abstract in English) Elektrotechnik & Informationstechnik 134(7), 370-373, 2017. https://doi.org/10.1007/s00502-017-0531-1

[62] J. Rasmussen, A. M. Pejtersen, L. P. Goodstein, Cognitive Systems Engineering, New York, USA: Wiley, 1994.

[63] S. Mattsson, "What is perceived as complex in final assembly? To define, measure and manage production complexity" Ph.D. Thesis, Chalmers University of Technology Gothenburg, Sweden, 2013.

[64] T. Fässberg, U. Harlin, K. Garmer, P. Gullander, Å. Fasth, S. Mattsson, K. Dencker, A. Davidsson, J. Stahre, "An Empirical Study Towards a Definition of Production Complexity" in 21st International Conference on Production Research, Stuttgart, Germany, 2011.

[65] G. Schuh, M. Riesener, Produktkomplexität managen. Strategien –Methoden – Tools, 3rd ed., Munich, Germany: Hanser, 2017. https://doi.org/10.3139/9783446453340

[66] A. Kaiser, "Integriertes Variantenmanagement mit Hilfe der Prozesskostenrechnung" Ph.D. Thesis, University of St. Gallen, Switzerland, 1995.

[67] S. D. Dionne, F. J. Yammarino, L. E. Atwater, W. D. Spangler, "Transformational leadership and team performance" Journal of Organizational Change Management, 17(2), 177-193, 2004. https://doi.org/10.1108/09534810410530601

[68] E. A. Locke, G. P. Latham, A Theory of Goal-Setting and Task Performance, New Jersey, USA: Englewood Cliffs, Prentice Hall, 1990.

[69] EN ISO 9241 Part 110 (September 2008): Ergonomics of human-system interaction - Part 110: Dialogue principles

NemoMap: Improved Motif-centric Network Motif Discovery Algorithm

Tien Huynh[*,1], Somadina Mbadiwe[2], Wooyoung Kim[1]

[1]*Computing and Software Systems, School of STEM, University of Washington Bothell, Bothell, WA 98011, USA*

[2]*Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV 26506, USA*

A B S T R A C T

Network motif analysis has several applications in many different fields such as biological study and social network modeling, yet motif detection tools are still limited by the intensive computation. Currently, there are two categories for network motif detection method: network-centric and motif-centric approach. While most network-centric algorithms excel in enumerating all potential motifs of a given size, the runtime is infeasible for larger size of motifs. Researchers who are interested in larger motifs and have established a set of potential motif patterns could utilize motif-centric tools to check whether such patterns are truly network motifs by mapping them to the target network and counting their frequency. In the paper, we present NemoMap (Network Motif Mapping algorithm) which is an improvement of the motif-centric algorithm, GK (by Grochow and Kellis) and MODA (Motif Detection Algorithm). Experimental results on three different protein-protein interaction networks show that NemoMap is more efficient in mapping complex motif patterns, while GK and MODA is much faster in analyzing simpler patterns with fewer edges. We also compare the performance of NemoMap and ParaMODA (introduced previously to improve MODA), and the result shows that NemoMap yields better runtime due to the implementation of Grochow-Kellis' symmetry-breaking technique and the better node selection process.

1   Introduction

Advancements in modern computer processing power and storage capability have enabled the collection of large-scale networks such as social networks, computer networks, and biological networks. These networks hold thousands or even millions of nodes (vertices), and they could be modelled using the graph structure which allows researchers from different fields to study under a common framework. One of the biggest challenges in network analysis is to determine network motifs, defined as the subgraph patterns that occur more frequently in the original network than in similar randomized networks. As an extension of work originally presented in IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2017 [1],

we provide an improved network motif detection algorithm in this paper.

The motivation for network motif study comes from the desire to gain more insight on the structural foundation of complex networks because those substructures might be the key component of a specific functionality. Such knowledge could be used to classify networks into "superfamilies" [2], or to select appropriate network model to study a real network [3]. In biological networks, network motif analysis has been used for discovering basic functional foundation in Transcriptional Regulation Networks (TRN) [4, 5], predicting interaction in Protein-Protein Interaction (PPI) networks [6], and studying breast-cancer related genes [7].

Although network motif analysis has significant ap-

---

[*]Corresponding author, Computing and Software Systems, School of STEM, University of Washington Bothell, WA 98011, USA, huynhti@uw.edu

plications in the real world, in practice it is infeasible to find motif of larger size because the process involves a nondeterministic polynomial time (NP) isomorphic testing and statistical testing. Two graphs which have the same number of nodes connecting in the same way are considered isomorphic, and most network motif analysis includes a huge number of isomorphic testing to classify motif and to avoid double counting. Additionally, statistical testing is used to determine the uniqueness of motifs in a number of similar randomly generated networks and the same process of isomorphic testing runs for each newly created network (or ensemble of networks), so time complexity increases dramatically. Typically, more than $1,000$ networks are generated, and the frequency of each motif is statistically compared using P-value or Z-score. A subgraph pattern with P-value of less than 0.01 or a Z-score of higher than 2 is identified as a network motif [8]. Such operations are very time consuming and usually are the main bottleneck of network motif analysis.

There have been several motif detection algorithms developed over the years, and they generally fall into two categories: network-centric and motif-centric [9]. The network-centric approach would search all possible enumerations of subgraphs of a given size in the target network, while the motif-centric method would predetermine one or more query graph patterns then count the occurrence of each query graph in the original network. Both approaches utilize statistical testing to check the significance of each potential motif.

Here, we propose NemoMap (Network Motif Mapping algorithm) which is an extension of ParaMODA in [1]. NemoMap improves the performance of ParaMODA which is an improvement of the motif-centric algorithms developed by Grochow-Kellis (GK) [10]. NemoMap is implemented in C++ so that it can be included in NemoLib [11], which is currently developed with Java, C++ and Python and includes network-centric method only. With the addition of NemoMap in NemoLib, we expect to have both choices available in a single library. In this paper, we also compare the performance of NemoMap against GK, as well as ParaMODA.

The rest of the paper is organized as the following: Section 2 describes background information of network motif; Section 3 discusses NemoMap algorithm and its implementation; Section 4 reports experimental results and performance comparison, followed by conclusion and future works in Section 5.

2 Network Motif

2.1 Network Motif Detection

Network motifs were introduced in 2002 by Milo et al. as "patterns of interconnections occurring in complex networks at numbers that are significantly higher than those in randomized networks" [4]. Suppose we have a network G with $|G|$ vertices, and a number k such that $3 \leq k << |G|$. A connected subgraph M of size k which

appears more frequently in G than a predetermined threshold value is a network motif. Typically, we use the statistical P-value or Z-score of the frequency of M in thousands of randomly generated similar networks as the threshold value.

$$\text{P-value}(M) = \frac{1}{N} \sum_{n=1}^{N} c(n), \qquad (1)$$

$$\text{where } c(n) = \begin{cases} 1, \text{if } f_R(M) \geq f_G(M). \\ 0, \text{otherwise} \end{cases}$$

$$\text{Z-score}(M) = \frac{f_G(M) - \mu(f_R(M))}{\sigma(f_R(M))} \qquad (2)$$

Here, N is the number of random graphs, and $f_G(M)$ is the frequency of M in the target network while $f_R(M)$ is the frequency of M in the random network. $\mu(f_R(M))$ is the mean of frequencies of M in the random networks and $\sigma(f_R(M))$ is the standard deviation of frequencies of M in the random networks.

Generally, a subgraph with P-value < 0.01 or Z-score > 2.0 is considered a network motif after analyzing in more than $1,000$ random networks [8].

2.2 Isomorphic Testing
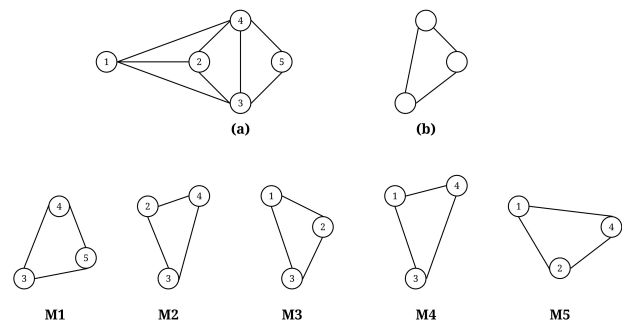
Figure 1: (a) A sample undirected graph; (b) A subgraph pattern; (c) M1-M5 are all isomorphic to (b)

Two graphs are isomorphic if they have the same number of nodes and there is "a one-to-one mapping between their nodes such that each edge in one graph can be mapped to an edge in the other graph" [12]. Figure 1 shows isomorphic subgraph patterns as an example. Since isomorphic testing is known to be an NP-complete problem, a common practice for categorizing graphs is to use a program called *Nauty* [13] to canonically label subgraph patterns; if two graphs have the same canonical label, they are isomorphic. One drawback of using Nauty is its dependency on the Linux operating system.

Another technique for analyzing isomorphism is comparing certain invariants of the graphs such as the degree of each node or the degrees of each node's neighbors [10]. Those properties are inherent to the structure of the graph, so two graphs are isomorphic if

they have the same properties. Here, we will utilize the later approach because we want the implementation without Nauty, to be operating system (OS) independent.

2.3 Network-Centric Approach

Many researchers have developed algorithms and software implementations of network motif detection focusing on the network-centric approach which searches all possible enumerations of subgraphs of a given size in the original target network. Some of the most distinguished applications include MAVisto [14], MFinder [15], FANMOD [16, 17], Kavosh [18], NemoFinder [19], and NetMODE [20], most of which involve the typical process of enumerating different k-size subgraphs in the original network. Unfortunately, even with modern computer architecture such solutions are still limited in the size of detectable motifs. Network-centric algorithms using exhaustive exact enumeration of k-size subgraphs in the network can often find only motifs with up to 6 nodes [10] due to the time-consuming process of scanning for every nodes and edges of the network.

One method to substantially reduce such processing time is sampling, employed by applications like MFinder and FANMOD. MFinder uses an edge-sampling strategy introduced by Kashtan et al. [15] which randomly selects an edge in the target network then picks a k-size subgraph around that edge to compare. However, edge-sampling technique seems to be biased toward subgraphs that have more edges resulting in higher count for such subgraphs [12]. Additionally, there is also the node-sampling strategy used by FANMOD which assigns a probability to the node to determine the chance that such node would be explore further [17]. The later sampling technique proves to be more efficient in reducing runtime, yet it could only increase FANMOD maximum discoverable motif size to 8 nodes.

2.4 Motif-Centric Approach

Due to the limitation of network-centric algorithms, Grochow and Kellis (GK) introduced a different approach for larger motif detection by first determining a query list of subgraph patterns of size k, then finding the frequency of each pattern in the target network by mapping the query graphs to all possible location in the network [10].

This strategy is described as "motif-centric", and the application can find motifs of up to 15 nodes. Moreover, the algorithm employs an efficient enumeration process called symmetry-breaking technique to significantly reduce isomorphic testing which is one of the most intensive tasks in network motif detection. The symmetry-breaking condition ensures that each mapping of the query graph to the target network is unique so that unnecessary isomorphic testing is avoided. However, one disadvantage is that it requires a set of query subgraphs to start with. Therefore, if

the researcher has no method to determine an appropriately concise k-size query list, run-time can suffer heavily with increasing subgraph size since full enumeration of all possible variations of k-size subgraphs is needed. Table 1 shows the exponential growth of pattern variations of k-size subgraphs up to 10 nodes, and in the worst-case scenario the application would have to iterate over $341,247,400,399,400,000,000$ of 10-node patterns many of which might not even exist in the target graph [9].

Table 1: Number of non-isomorphic subgraphs for undirected and directed graphs with up to 10 vertices. Courtesy of [8]

Vertices	Number of non-isomorphic subgraphs	
	Undirected	Directed
1	1	1
2	1	2
3	2	13
4	6	199
5	21	9,364
6	112	1,530,843
7	853	880,471,142
8	11,117	1,792,473,955,306
9	261,080	13,026,161,682,466,200
10	11,716,571	341,247,400,399,400,000,000

Heavily inspired by Grochow-Kellis' (GK) algorithm, MODA (Motif Discovery Algorithm) [21] improves upon the former by the introduction of a query graph hierarchy expansion treeas illustrated in Figure 2. MODA chooses a subgroup of the query graphs which might appear more frequently in the target network to start mapping with as the top level of the tree; then it adds one edge to each graph of the previous level while ensuring that no duplicate is present to build the subsequent level of the tree. Because each level is only extended by one edge, the mapping information of the previous level can be saved and reused to extend mapping by one node, so computation time would be reduced substantially. It also uses a node-sampling method during the edge-expansion based on the linearly proportional distribution of nodes and edges [9].

3 NemoMap

We propose NemoMap (Network Motif Mapping algorithm) as an extension of ParaMODA [1] which is an improvement of the motif-centric algorithm introduced by Grochow-Kellis (GK). We call it NemoMap since it is an improvement of the mapping function introduced by GK and we will include it in the NemoLib library [11]. Since this is a direct improvement of GK, we will firstly explain in detail the main components of GK, then discuss our alteration in NemoMap, and

lastly compare the difference in implementation of NemoMap and ParaMODA.
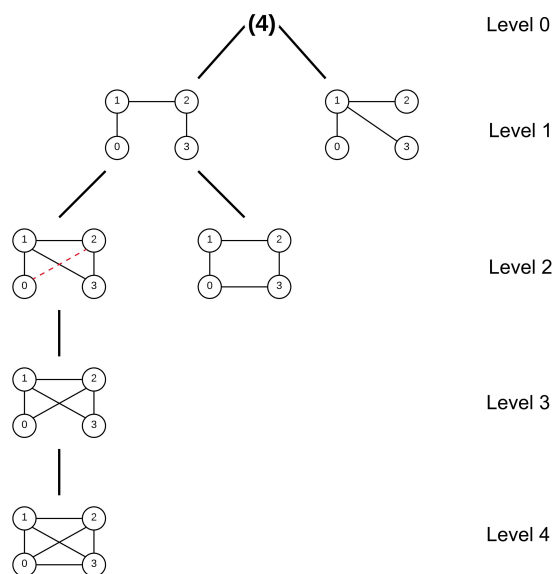
Figure 2: Expansion tree of size-4 subgraph, courtesy of [21]

3.1 Grochow-Kellis (GK)

We will discuss the mapping function of GK which focuses on the mapping of a query graph to the target network to find the frequency of such query graph's appearance in the network because NemoMap and ParaMODA improve upon such functionality.

Suppose we have an input network *G* and a query graph *H*, and the goal is to find and count all the possible mapping *f* from *H* to *G* which *G* can support. GK mapping algorithm involves three functions as shown in Algorithm 1, 2, and 3.

As discussed in Section 2.4, GK utilizes comparison of two invariants between graphs to test for isomorphism: the degree of each node and the degrees of each node's neighbors. It starts with an arbitrarily partial map of one node *h* of *H* to one node *g* of *G*, then extends that map by a single node for every recursive call of IsomorphicExtensions (Algorithm 2) until all nodes of H are mapped to appropriate nodes of G. Additionally, IsomorphicExtensions guarantees that any newly mapped node is connected to the already-mapped nodes, so the returned mapping must be an isomorphism [10].

Although the symmetry-breaking condition is optional, it significantly improves runtime and memory usage. Figure 3 shows symmetry-breaking example. Without symmetry-breaking, FindSubgraphInstances (Algorithm 1) would have to maintain a list of all mappings in memory to screen out duplicate mappings resulting in low efficiency as the program has to constantly write and read data from memory or disk. On the other hand, symmetry-breaking ensures that each mapping is unique, so there is no need to maintain the mapping list. Such feature substantially reduces

memory usage and runtime for larger network analysis or bigger size query graphs [10] as illustrated in Table 2.
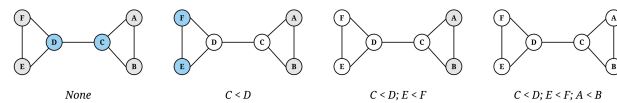
Figure 3: Symmetry-breaking example: Finding conditions that will break all the symmetries of a 6-node graph. White nodes are fixed by any automorphism preserving the indicated conditions, and other nodes are shaded according to their equivalence class under the automorphisms which preserve the indicated conditions. (Courtesy of [10])

3.2 ParaMODA

Compared to GK algorithm, ParaMODA follows closely with a couple of alterations. In FindSubgraphInstances in Algorithm 1, instead of going through all nodes *h* of *H*, we only arbitrarily choose one node *h* to traverse from. The chosen *h* node serves as the fixed starting point for all traversals of the query graph, and since the nodes of *G* are only traversed once with the chosen node *h*, the visited nodes *g* of *G* are not removed from the network [1]. There are no changes to IsomorphicExtensions in Algorithm 2.

ParaMODA improves time complexity by eliminating the inner iteration (loop) in FindSubgraphInstances at the expense of not shrinking the network by keeping visited nodes. The performance comparison will be discussed in Section 4. The logic behind the removal of the inner iteration comes from the fact that if the query graph is connected and the network is unchanged, any node of the network could be used as the starting point for mapping because the set of edges remain unchanged [1].

Another advantage of ParaMODA is the possibility of parallelization. Since each mapping from *H* to *G* is unique and each starting mapping node of the network is independent, we can partition the vertices easily across processors and have the result aggregated. Because the nodes are ordered based on their degrees in descending order, we can achieve better load distribution across processes by using strategies that factor in the uneven nature of that distribution [1].

3.3 NemoMap

As an extension to ParaMODA [1], NemoMap maintains the core structure with two differences: the clarification of node *h* choosing process, and the implementation of symmetry-breaking technique. We implement NemoMap in C++ so that it can be included in the NemoLib library [11].

Instead of arbitrarily picking a node *h* of *H* to start the mapping process as discussed in Section 3.2, NemoMap chooses the most constrained node *h* to traverse from. This selection process starts with picking the node with the highest degree, and if there are more than one nodes with the same highest degree we will

---

Algorithm 1: FindSubgraphInstances (G, H), by courtesy of [10]

Find all instances of query graph H in network G

Input : Graph G and a query graph H

Output: A set of all instances of H in G

Start with an empty set of instances

Find $Aut(H)$. Let H_E be the equivalence representatives of H

Find symmetry-breaking conditions C of H given H_E and $Aut(H)$

Order the nodes of G by increasing degree and then by increasing neighbor degree sequence

for *each node g of G* **do**

    for *each node h of H such that g can support h* **do**

        Let f be the partial map associating $f(h) = g$

        Find all isomorphic extensions of f [up to symmetry] (i.e. call `IsomorphicExtensions` $(f, H, G,$ $C[h])$)

        Add the images of these maps to the set of all instances

    Remove g from G

Return the set of all instances.

Table 2: The number of subgraphs encountered by NemoMap with and without symmetry-breaking (including multiple encounters for the version without symmetry-breaking). The improvement factor is exactly the average number of automorphisms of subgraphs of the associated size. (Courtesy of [10])

Nodes	Undirected PPI Network			Directed Regulatory Network		
	Total Subgraphs Searched	With Symmetry-Breaking	Improvement	Total Subgraphs Searched	With Symmetry-Breaking	Improvement
2	3.7×10^4	1.1×10^4	×3.13	2.6×10^4	1.3×10^4	×2.02
3	4.0×10^5	7.0×10^4	×5.77	9.7×10^5	1.8×10^5	×5.41
4	4.4×10^6	4.1×10^5	×10.9	4.4×10^7	2.5×10^6	×18.0
5	5.1×10^7	2.3×10^6	×22.2	2.3×10^9	3.2×10^7	×73.3
6	5.7×10^8	1.2×10^7	×46.3	1.3×10^{11}	4.0×10^8	×334
7	6.4×10^9	6.6×10^7	×96.2	-	-	-

select among them the node with the highest sum of neighbors degrees. If there is still a tie after the second selection round, we will arbitrarily choose one node among the last filtered group to be node h. Such selection process helps reduce the number of recursive isomorphic testing as the constrained node h has a lower chance of being mapped to nodes in G, resulting in a faster runtime. Algorithm 4 shows the altered process.

Without symmetry-breaking, ParaMODA suffers from memory overhead issue when analyzing query with a large number of mappings because the program has to maintain a list of mappings to avoid duplication. As a result, the program frequently runs out of memory and crashes if the target network is large. Furthermore, it also has slower runtime as redundant recursive calls for isomorphic testing are completed. NemoMap solves this problem by including the GK symmetry-breaking condition with a little tweak. Because only one node h of H is chosen, SymmetryConditions in Algorithm 3 does not need to find the symmetry-breaking conditions for all nodes of H; only the conditions of the selected node h are needed as illustrated in Algorithm 5.

NemoMap implementation also uses the standard

vector data structure instead of the user-defined class Mappings as in ParaMODA program. The change is due to performance optimization, and from the fact that all query graph mappings always start from one fixed node h with the same extension path. Therefore, from the standpoint of a key-value pair of the Mappings class with the keys representing mapped nodes h of query graph H, the key's sequence order will always be the same, so there is no need to maintain such data; we can save it at the beginning for later use if necessary. Instead, we use a vector to store the values, i.e. the corresponding mapping nodes g of network G, in the correct order that the keys are in.

The input file format for the query graph and the target network of NemoMap remains the same with each line contain two integers separated by a white space or a tab. Each line of the input file represents an edge of the graph, and the two integers are the ID number of the nodes which form such edge. Lines start with "#" are comments and will be ignored [1]. Additionally, the program disallows self-edge, i.e. edge from a node pointing to itself, as well as duplicating edge.

As discussed earlier, the program generates a list of mappings of the query graph H in all possible lo-

---

Algorithm 2: ISOMORPHICEXTENSIONS(f,H,G, $C[h]$), by courtesy of [10]

Find all isomorphic extensions of partial map $f : HG$ **[satisfying** $C(h)$**]**
Input   :Partial map $f : H \rightarrow G$, a graph G, query graph H, and a symmetry-breaking condition $C(h)$
Output:All isomorphic extensions of f
Start with an empty list of isomorphisms
Let D be the domain of f
if $D = H$ **then**
    Return a list consisting solely of f. (Or write to disk)
Let m be the most constrained neighbor of any $d \in D$ (constrained by degree, neighbors mapped, etc.)
for *each neighbor n of* $f(D)$ **do**
    if *there is a neighbor* $d \in D$ *of* m *such that* n *is not neighbors with* $f(d)$*, or if there is a non-neighbor* $d \in D$
    of m *such that* n *is neighbors with* $f(d)$*, or if assigning* $f(m) = n$ *would violate a symmetry-breaking*
    condition in $C(h)$*]* **then**
        Continue with the next n
    else
        Let $f' = f$ on D, and $f'(m) = n$
    Find all isomorphic extensions of f'
    Append these maps to the list of isomorphisms
Return the list of isomorphisms.

Algorithm 3: SYMMETRYCONDITIONS(H_E, $Aut(H)$), by courtesy of [10]

Find symmetry-breaking conditions for H **given** H_E**,** $Aut(H)$
Input   :Nodes of H grouped into equivalence classes H_E, automorphisms of H $Aut(H)$
Output:A map from equivalence representatives to sets of symmetry-breaking conditions
Let M be an empty map from equivalence representatives to sets of conditions
for *each* $n \in H_E$ **do**
    Let C be an empty set of conditions
    $n' = n$
    $A = Aut(H)$
    do
        Add "Label(n') <Min{Label(m)|$m \sim An'$ and $m \neq n'$}" to C
        $A = \{f \in A | f(n') = n'\}$
        Find the largest A-equivalence class E
        Pick $n' \in E$ arbitrarily
    while $|A| \neq 1$
    Let $M(n) = C$
Return M

---

cations of the network G. The mappings can be saved to disk if necessary, however the current NemoMAp implementation simply discards them and only count the number of mappings found.

3.4   Summary of Changes

The changes in NemoMap vs. GK algorithm are picking one fixed node h in terms of mapping and finding symmetry-breaking condition of only that node. NemoMap is implemented in C++ and includes the symmetry-breaking technique, while ParaMODA is written in C# and does not have symmetry-breaking; NemoMap also have an improved node h selection process. Table 3 summarizes the changes.

4   Experiments

4.1   Setup

Motif-centric approach is not primarily designed for exhaustive search of network motif, so we will focus on comparing runtime performance between NemoMap and GK on predetermined query graph mapping. The experiment will include two parts: performance comparison between NemoMap and ParaMODA, and between NemoMap and GK. All testing will be conducted on three protein-protein interaction (PPI) networks obtained from the Database of Interacting Proteins (DIP) [22]:

- *Homo sapiens* (Hsapi): $1,715$ nodes with $1,873$ edges (low edge density)

- *Escherichia coli* (E. coli): $1,223$ nodes with $1,654$ edges (average edge density)

- *Saccharomyces cerevisiae* (S. cerevisiae): $2,164$ nodes with $4,303$ edges (high edge density)

As mentioned in [1], the performance difference between ParaMODA and GK algorithm varies widely depending on the query graph used. In this paper, we attempt to elaborate which query graph categories

---

Algorithm 4: FindSubgraphInstances_NemoMap(G, H)

Find all instances of query graph H in network G
Input : Network G and a set of query graph H
Output: A set of all instances of H in G
Start with an empty set of instances
Pick the most constrained node h from H (constrained by degree, sum of neighbors' degrees)
Find $Aut(H)$
Find symmetry-breaking conditions C of h given $Aut(H)$
Order the nodes of G by increasing degree and then by increasing neighbor degree sequence
for *each node g of G such that g can support h* **do**
    Let f be the partial map associating $f(h) = g$
    Find all isomorphic extensions of f [up to symmetry] (i.e. call `IsomorphicExtensions` ($f, H, G, C[h]$))
    Add the images of these maps to the set of all instances
Return the set of all instances.

Algorithm 5: SymmetryConditions_NemoMap($H_E, Aut(H)$)

Find symmetry-breaking conditions for H given H_E, $Aut(H)$
Input : Automorphisms of $H, Aut(H)$
Output: A set of symmetry-breaking conditions for node h
Let C be an empty set of conditions
$n' = n$
$A = Aut(H)$
do
    Add "Label(n') <Min{Label(m)|$m \sim An'$ and $m \neq n'$}" to C
    $A = \{f \in A | f(n') = n'\}$
    Find the largest A-equivalence class E
    Pick $n' \in E$ arbitrarily
while $|A| \neq 1$
Let $M(n) = C$ Return M

---

are more advantageous to each algorithm. Therefore, we break down the query graphs into three pattern groups:

- Group Simple: Low edge-to-node count

- Group Average: Average edge-to-node count

- Group Complex: High edge-to-node count

We choose to use the ratio of edge count to node count as a measure for a graph complexity with higher ratio equating to higher complexity. Although there are other factors in a graphs complexity such as out-degree distribution, we think using the edge-to-node ratio would making the experiment more manageable. Each query graph group consists of four query graphs with four to seven nodes. All tests were conducted on a desktop computer with the following specifications:

- CPU: Intel Core i5-8400 (6-core) 2.80Ghz, Turbo 4.00Ghz, 9MB cache

- RAM: 16GB DDR4 2133Mhz

- Hard Drive: 240GB SSD

- GPU: Intel UHD Graphics 630

- OS: Windows 10 Home Edition

Table 3: Summary of changes of NemoMap compared to ParaMODA and GK

	NemoMap	**ParaMODA**	**GK**
Language	C++	C#	Java
Mapping	Pick the **most constrained node** h of query graph H (constrained by degree, sum of neighbors' degrees), then start all mapping from the h, and do not remove mapped node g	Pick an **arbitrary node** h of query graph H, then start all mapping from the h, and do not remove mapped node g	Iterate over **all nodes of H** (or **all representative node of H** if symmetry-breaking is used), and remove mapped node g after every iteration
Symmetry-breaking	Find symmetry-breaking condition for **only one fixed node h**	Symmetry-breaking condition is **not included**	Find symmetry-breaking condition for **all representative nodes of H**

- IDE: Visual Studio 2017 Community

The source code of Grochow-Kellis' algorithm in Java was provided by the author [10], and all runtimes reported for GK are measured using this implementation. The source code was compiled with the IntelliJ IDEA 2018.5 Community IDE.

4.2 Performance Metrics

The primary performance metric is runtime measured in milliseconds (ms). The comparison metric *Reduction* is computed to represent the relative performance between NemoMap and GK algorithm, and is defined in Equation (3):

$$R_{GK} = \frac{T_{GK} - T_{NemoMap}}{T_{GK}} \times 100 \qquad (3)$$

Here, T_{GK} is the runtime of Grochow-Kellis' algorithm in millisecond (*ms*), and $T_{NemoMap}$ is the runtime of NemoMap in *ms*. R_{GK} represents the relative reduction in runtime from using NemoMap instead of GK expressed in percentage.

Another similar metric $R_{ParaMODA}$ is computed to represent the relative performance between NemoMap and ParaMODA, and is defined in the following equation:

$$R_{ParaMODA} = \frac{T_{ParaMODA} - T_{NemoMap}}{T_{ParaMODA}} \times 100 \qquad (4)$$

$T_{ParaMODA}$, $T_{NemoMap}$ are the runtime of ParaMODA in *ms* and NemoMap in *ms* respectively, and $R_{ParaMODA}$ is the relative reduction in runtime from using NemoMap instead of ParaMODA expressed in percentage.

4.3 Results and Discussion

We show the results comparing the performance of NemoMap against Grochow-Kellis as well as of NemoMap against ParaMODA in Table 4, Table 5, and Table 6.

4.3.1 NemoMap vs. ParaMODA

In all cases, the implementation of NemoMap ran significantly faster than ParaMODA, and the performance gap widens when the complexity of the network increases. Starting with the simpler network of Hsapi (1,715 nodes and 1,873 edges: edge-to-node ratio of 1.09) in Table 4, we have runtime reduction ranging from 61.29% to 99.57% when using NemoMap; the second network of E. coli has more complexity with a higher edge-to-node ratio (1,223 nodes and 1,654 edges: ratio of 1.35) which results in a higher reduction rate of between 93.58% to 99.97% in runtime as recorded in Table 5; lastly, in Table 6, the most complex network S. cerevisiae with a very high edge-to-node ratio of 1.99 (2,164 nodes and 4,303 edges) yields an even higher reduction range from 94.16% to 99.96% in runtime for NemoMap.

ParaMODA ran out of memory and stopped in several runs, most of which are case with large numbers of mappings (more than 9 million mappings). As discussed in earlier sections, this problem could be addressed by introducing symmetry-breaking technique into the program so that maintaining a list of mappings in memory is not needed. Furthermore, symmetry-breaking also reduced runtime by eliminating redundant isomorphic testing which is the most intensive task of subgraph mapping. Other elements that might help improve runtime include language performance difference between NemoMap and ParaMODA and the level of code optimization.

4.3.2 NemoMap vs. Grochow-Kellis

As mentioned in our previous work [1], performance difference between ParaMODA and GK varied depending on the query graph. Therefore, we attempt to figure out the factors affecting performance difference between NemoMap and GK in this paper by dividing query graphs into groups. For illustration purposes, we sketch the performance in runtime (*ms*) of NemoMap and GK with line charts on the networks of Hsapi, E. coli, and S. cerevisiae in Figure 4, 5, and 6 respectively. Each Figure contains three charts for the three different groups of query graphs as described in Section 4.1: (a), (b), and (c) denotes the Simple, Average, and Complex pattern accordingly.

In most scenarios, GK's algorithm was substantially slower than NemoMap. However, in the case of simple query graphs which had a small count of edges (edge count was smaller than node count), NemoMap runtime was slower as illustrated by chart (a) of Figure 4, 5, and 6 with increasing performance gap when the query graph size gets bigger. Interestingly, NemoMap was faster for simple query graphs of size 4, but it dragged behind for size-5 and larger query graphs. The runtime also got slower when the network becomes more complex: on the Hsapi, NemoMap was behind by 40.79% to 72.06%; for the E. coli, the gap increased to up to 89.30% slowdown; lastly, the S. cerevisiae networks show slower performance of between 49.63% to 90.76%.

On the other hand, when the query graphs became complex with more edges, NemoMap ran much faster than GK. Chart (b) and (c) of Figure 4, 5, and 6 all show exponential improvement in runtime when using NemoMap. Although we did not see an upward trend in performance improvement when the query graph size increases or when the network gets more complex, runtime reduction rate was consistently in the 75% to 90% range as recorded in Table 4, 5, and 6. The only exception was in Table 5 for the Average pattern size-6 query graph with improvement of only 48.84%.

Such observation was consistent in all three tested PPI networks. One possible explanation could be from the isomorphic testing phase, which is the most inten-

Table 4: Performance comparison on the PPI network of Hsapi

Homo Sapiens (Hsapi): 1,715 nodes - 1,873 edges				
Group Simple	**4-node**	**5-node**	**6-node**	**7-node**
Number of edges	3	4	5	6
Number of mappings	32,449	195,061	1,110,425	5,480,057
NemoMap (ms)	122	936	6,951	45,263
Grochow-Kellis (ms)	575	1,002	4,937	26,306
Reduction GK	*78.78%*	*6.59%*	*-40.79%*	*-72.06%*
ParaMODA (ms)	1,531	44,000	1,610,718	N/A
Reduction ParaMODA	*92.03%*	*97.87%*	*99.57%*	*N/A*
Group Average	**4-node**	**5-node**	**6-node**	**7-node**
Number of edges	5	6	7	9
Number of mappings	149	1,889	17,686	3,424
NemoMap (ms)	23	50	136	52
Grochow-Kellis (ms)	406	486	914	1,380
Reduction GK	*94.33%*	*89.71%*	*85.12%*	*96.23%*
ParaMODA(ms)	78	484	5,984	12,687
Reduction ParaMODA	*70.51%*	*89.67%*	*97.73%*	*99.59%*
Group Complex	**4-node**	**5-node**	**6-node**	**7-node**
Number of edges	6	9	10	13
Number of mappings	17	4	23	1
NemoMap (ms)	24	23	27	35
Grochow-Kellis (ms)	499	475	961	563
Reduction GK	*95.19%*	*95.16%*	*97.19%*	*93.78%*
ParaMODA (ms)	62	78	187	156
Reduction ParaMODA	*61.29%*	*70.51%*	*85.56%*	*77.56%*

Table 5: Performance comparison on the PPI network of E. coli

Escherichia coli (E. coli): 1,223 Nodes - 1,654 Edges				
Group Simple	**4-node**	**5-node**	**6-node**	**7-node**
Number of edges	3	4	5	6
Number of mappings	91,848	1,072,997	12,494,338	127,746,287
NemoMap (ms)	297	4,531	66,967	879,216
Grochow-Kellis (ms)	830	3,608	31,331	464,460
Reduction GK	*64.22%*	*-25.58%*	*-79.39%*	*-89.30%*
ParaMODA (ms)	5,531	516,593	N/A	N/A
Reduction ParaMODA	*94.63%*	*99.12%*	*N/A*	*N/A*
Group Average	**4-node**	**5-node**	**6-node**	**7-node**
Number of edges	5	6	7	9
Number of mappings	606	10,449	95,052	39,750
NemoMap (ms)	12	46	356	199
Grochow-Kellis (ms)	309	1,122	1,395	4,611
Reduction GK	*96.12%*	*95.90%*	*74.48%*	*95.68%*
ParaMODA(ms)	187	3,421	82,125	673,453
Reduction ParaMODA	*93.58%*	*98.66%*	*99.57%*	*99.97%*
Group Complex	**4-node**	**5-node**	**6-node**	**7-node**
Number of edges	6	9	10	13
Number of mappings	110	125	1,400	48
NemoMap (ms)	10	14	27	37
Grochow-Kellis (ms)	372	549	1,264	966
Reduction GK	*97.31%*	*97.45%*	*97.86%*	*96.17%*
ParaMODA (ms)	171	953	17,031	112,140
Reduction ParaMODA	*94.15%*	*98.53%*	*99.84%*	*99.97%*

Table 6: Performance comparison on the PPI network of S. cerevisiae

Saccharomyces cerevisiae (S. cerevisiae): 2,164 nodes - 4,303 edges				
Group Simple	**4-node**	**5-node**	**6-node**	**7-node**
Number of edges	3	4	5	6
Number of mappings	170,159	1,294,037	9,824,580	67,818,397
NemoMap (ms)	652	6,509	60,693	514,772
Grochow-Kellis (ms)	1,016	4,350	33,507	269,856
Reduction GK	*35.83%*	*-49.63%*	*-81.14%*	*-90.76%*
ParaMODA (ms)	11,171	434,843	N/A	N/A
Reduction ParaMODA	*94.16%*	*98.50%*	*N/A*	*N/A*
Group Average	**4-node**	**5-node**	**6-node**	**7-node**
Number of edges	5	6	7	9
Number of mappings	4,869	53,601	327,121	232,404
NemoMap (ms)	79	263	1,453	1,280
Grochow-Kellis (ms)	810	2,291	2,840	9,270
Reduction GK	*90.25%*	*88.52%*	*48.84%*	*86.19%*
ParaMODA(ms)	1,734	37,640	1,289,328	N/A
Reduction ParaMODA	*95.44%*	*99.30%*	*99.89%*	*N/A*
Group Complex	**4-node**	**5-node**	**6-node**	**7-node**
Number of edges	6	9	10	13
Number of mappings	1,591	1,893	22,809	3,904
NemoMap (ms)	59	81	258	437
Grochow-Kellis (ms)	593	803	3,495	2,547
Reduction GK	*90.05%*	*89.91%*	*92.62%*	*82.84%*
ParaMODA (ms)	1,875	20,301	583,343	N/A
Reduction ParaMODA	*96.85%*	*99.60%*	*99.96%*	*N/A*
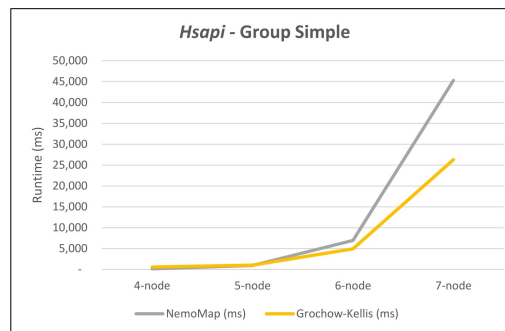
(a)
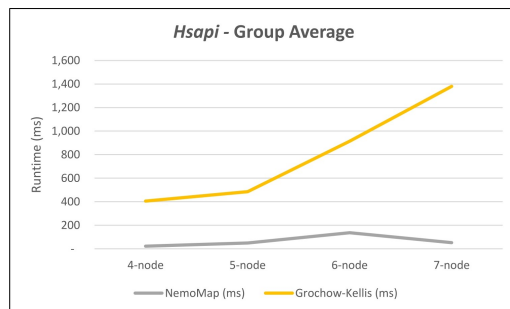
(b)
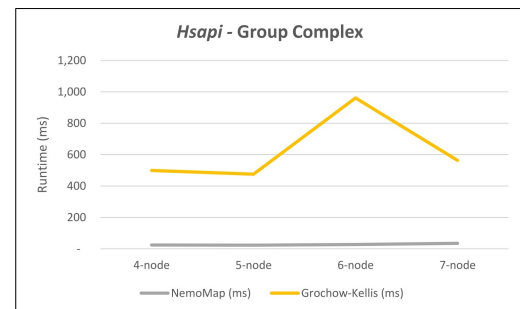
(c)

Figure 4: Graph of runtime against query graph size for (a) Simple graph, (b) Average graph, and (c) Complex graph between NemoMap and Grochow-Kellis for Hsapi. NemoMap outperforms in all cases except for (a).
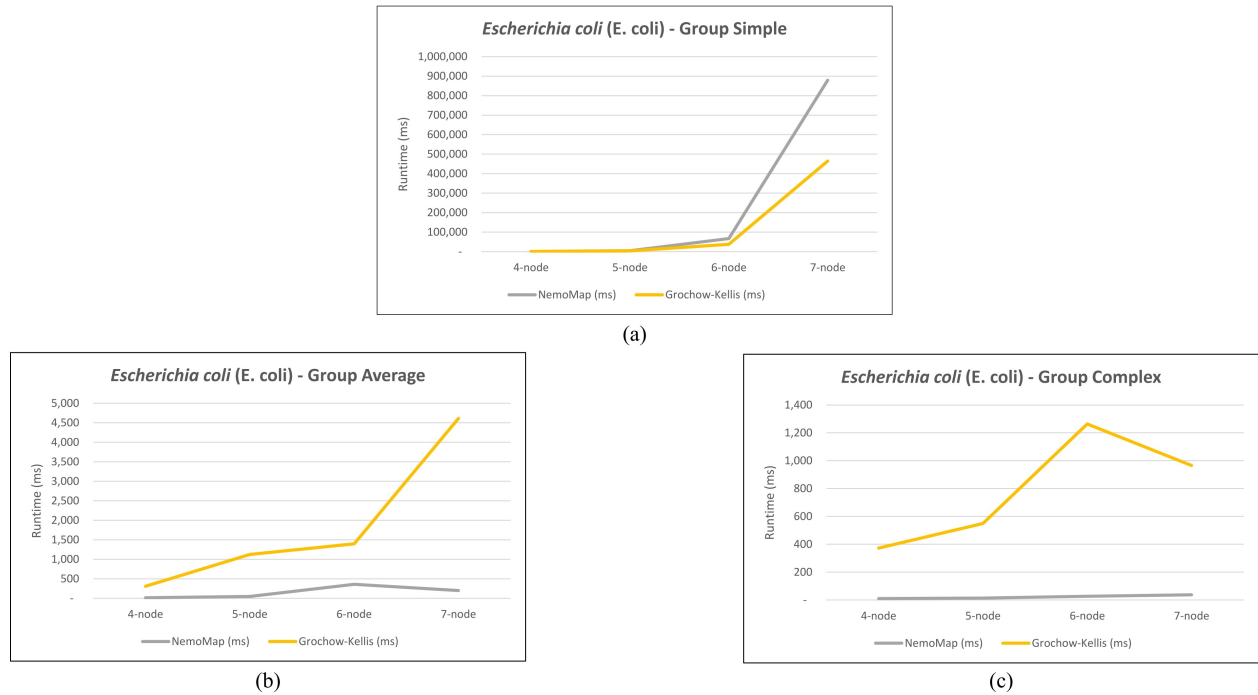
Figure 5: Graph of runtime against query graph size for (a) Simple graph, (b) Average graph, and (c) Complex graph between NemoMap and Grochow-Kellis for E. coli. NemoMap outperforms in all cases except for (a).
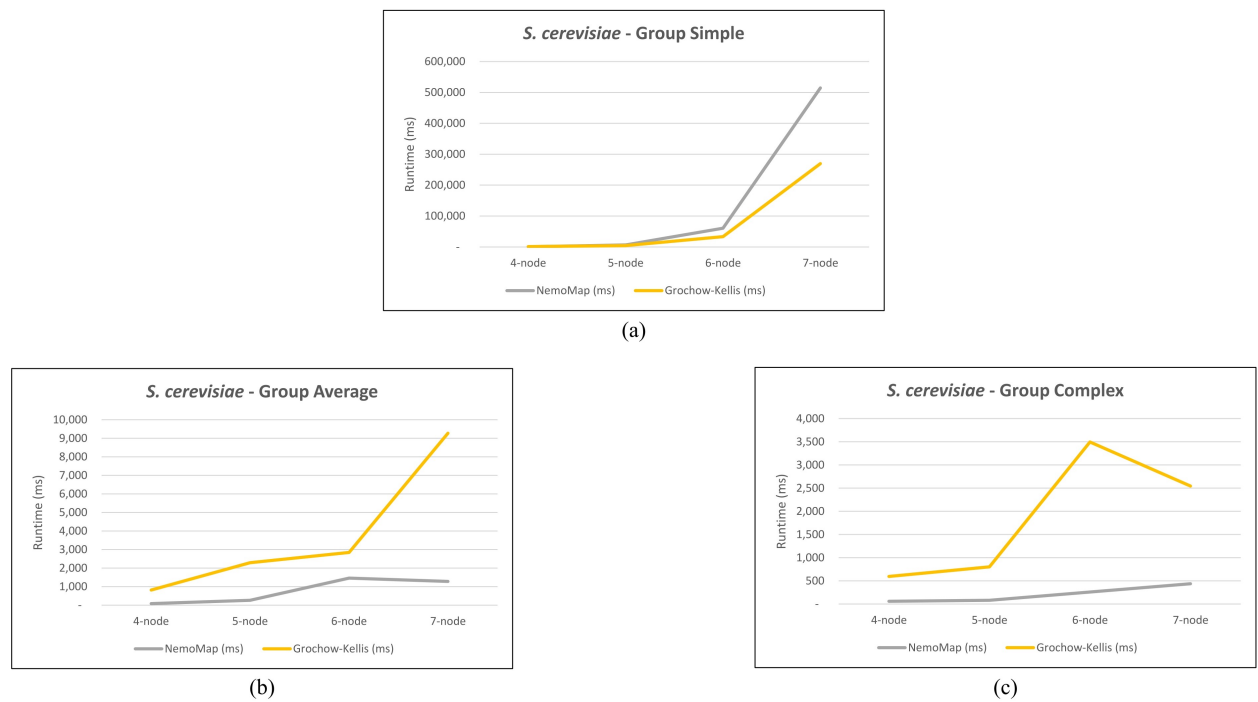
Figure 6: Graph of runtime against query graph size for (a) Simple graph, (b) Average graph, and (c) Complex graph between NemoMap and Grochow-Kellis for S. cerevisiae. NemoMap outperforms in all cases except for (a).

sive task of the subgraph mapping. As discussed in Section 3.1, GK uses two invariants between graphs to test for isomorphism: the degree of each node and the degrees of each nodes neighbors. To clarify the first invariant, when trying to map a node h of query graph H to a node g of network G, if node g's degree is smaller than node h's degree then h cannot be mapped to g and the function must move on to another node g as illustrated in Figure 7.
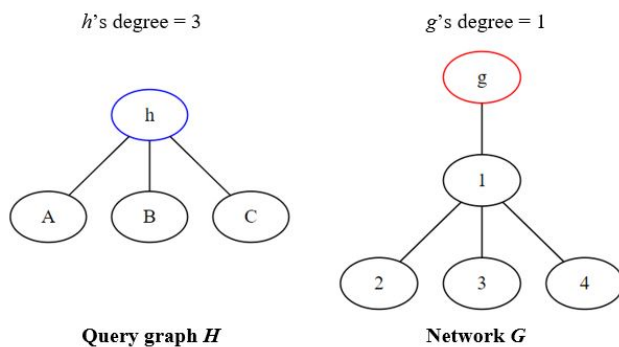
Figure 7: Node h cannot be mapped to node g because g's degree is smaller than h's degree
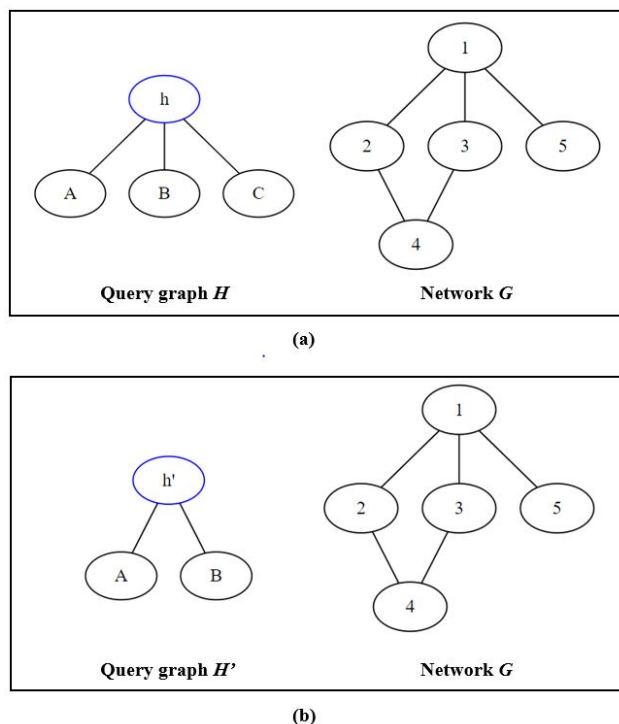
Figure 8: Mapping two different query graphs (a) H, and (b) H' to the same network G

Looking at two different scenarios as illustrated in Figure 8, we have one network G and we want to map two query graphs H and H' to. Assuming that we choose the first mapping node h and h' as highlighted for each query graph, and we examine the possibility of mapping node h to each node of network G.

In scenario (a), node h has a degree of three so only node 1 of G can support it. As a result, the recur-

sive call for isomorphic testing will only be called and branched out when trying to map node h to 1; node 2, 3, 4, and 5 will be aborted early so no isomorphic test is ran. On the other hand, in scenario (b), node h' has a degree of two so node 1, 2, 3, and 4 can all support it, resulting in a lot more of isomorphism recursive tests since only mapping to node 5 can be aborted early.

The larger number of isomorphic tests and the lower chance to abort early in (b) are advantageous for GK because it uses the same isomorphic testing algorithm as NemoMap, but the network is getting smaller and smaller after every iteration because GK removes mapped nodes from the network unlike the fixed network of NemoMap. However, GK needs to iterate over all nodes of H', so the removal effect might get canceled out. If the query graph is highly symmetrical and symmetry-breaking is used, then it will unlikely to be fully even out. Recall that with symmetry-breaking, GK only iterates over representative nodes of equivalence class. In (b), node A, B, and C belong to the same equivalence class as they are symmetrical (for more detail on equivalence class please reference [10]), so instead of iterating over four nodes (h', A, B, and C), Grochow-Kellis would only iterate over two nodes (h' and one arbitrary node from the equivalence class). Such effect might help GK retain better performance on simpler query graphs with lower nodes' degrees and large equivalence classes.

As mentioned in the previous section, other elements that might affect runtime include language performance difference between C++ and Java and the level of code optimization.

5 Conlusion and Future Works

In this paper, we present NemoMap, a motif-centric algorithm inspired by the work of Grochow and Kellis [10], as a tool to determine the frequency of a subgraph's appearance in a target network. Extending from our previous work in ParaMODA [1], the new experimental results show that there is improvement in performance using NemoMap, yet it does not apply in all cases. Specifically, we observe that NemoMap runs much faster when analyzing complex query graphs with more edges, while Grochow-Kellis' solution is more efficient for simpler query graphs with fewer edges (i.e. lower degrees) and high symmetry. As an extension, NemoMap also sees large improvement in runtime and memory usage compared to the predecessor ParaMODA due to the utilization of symmetry-breaking technique and the better node selection process.

Additionally, NemoMap structure allows for an easier parallelization scheme, as multiple query graphs or different nodes of the same query graph can be assigned to different processors running concurrently. Such operation is necessary for future research as the need for finding larger network motif becomes more prominent. Another essential future item is the implementation of a random network generator and a
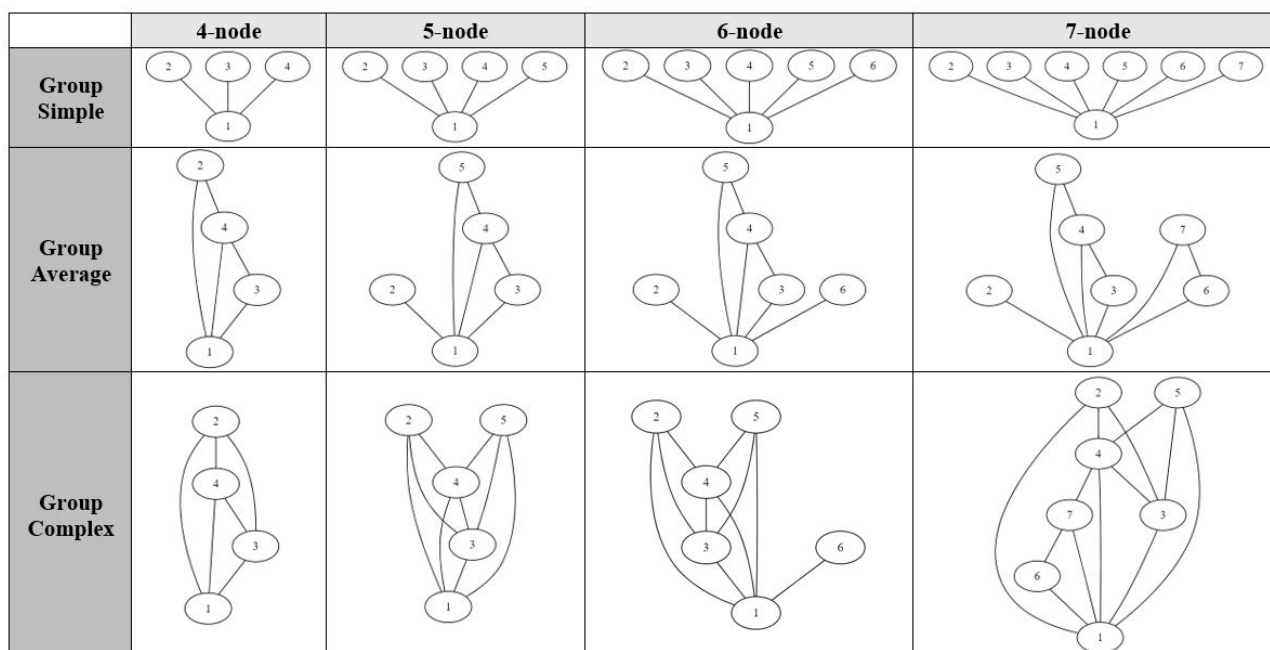
Figure 9: Visualization of groups of query graphs used for testing

statistical analysis system into the mapping function so that the NemoMap program could become a complete network motif detection suite.

The source code of NemoMap and usage instruction are available on https://github.com/tien-huynh/NemoMap. ParaMODA is also available on https://github.com/smbadiwe/ParaMODA.

Conflict of Interest   The authors declare no conflict of interest.

References

[1] S. Mbadiwe and W. Kim. Paramoda: Improving motif-centric subgraph pattern search in ppi networks. In *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1723–1730, Nov 2017.

[2] Ron Milo, Shalev Itzkovitz, Nadav Kashtan, Reuven Levitt, Shai Shen-Orr, Inbal Ayzenshtat, Michal Sheffer, and Uri Alon. Superfamilies of evolved and designed networks. *Science*, 303(5663):1538–1542, March 05 2004.

[3] Manuel Middendorf, Etay Ziv, and Chris H. Wiggins. Inferring network mechanisms: The drosophila melanogaster protein interaction network. *Proceedings of the National Academy of Sciences of the United States of America*, 102(9):3192–3197, 2005.

[4] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: Simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.

[5] R. Milo, N. Kashtan, S. Itzkovitz, M. E. J. Newman, and U. Alon. On the uniform generation of random graphs with prescribed degree sequences. *eprint arXiv:cond-mat/0312028*, December 2003.

[6] I. Albert and R. Albert.   Conserved network motifs allow protein-protein interaction prediction. *Bioinformatics*, 20(18):3346–3352, December 2004.

[7] Yuji Zhang, Jianhua Xuan, Benilo G de los Reyes, Robert Clarke, and Habtom W. Ressom. Network motif-based identification of breast cancer susceptibility genes. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5696–5699. IEEE, aug 2008.

[8] Bjvrn H. Junker and Falk Schreiber. *Analysis of Biological Networks*. Wiley, 2008.

[9] W. Kim, M. Diko, and K. Rawson. Network motif detection: Algorithms, parallel and cloud computing, and related tools. *Tsinghua Science and Technology*, 18(5):469–489, Oct 2013.

[10] Joshua A. Grochow and Manolis Kellis. Network motif discovery using subgraph enumeration and symmetry-breaking. In *Proceedings of the 11th annual international conference on Research in computational molecular biology*, RECOMB'07, pages 92–106, Berlin, Heidelberg, 2007. Springer-Verlag.

[11] Andrew Andersen and Wooyoung Kim. *NemoLib: A Java Library for Efficient Network Motif Detection*, pages 403–407. Bioinformatics Research and Applications: 13th International Symposium, ISBRA 2017, Honolulu, HI, USA, May 29 June 2, 2017, Proceedings. Springer International Publishing, Cham, 2017.

[12] Elisabeth Wong, Brittany Baur, Saad Quader, and Chun-Hsi Huang. Biological network motif detection: principles and practice. *Briefings in Bioinformatics*, 2011.

[13] B. McKay. Practical graph isomorphism. *Congr Numer*, 30:45–87, 1981.

[14] F. Schreiber and H. Schwobbermeyer. Mavisto: a tool for the exploration of network motifs. *Bioinformatics*, 21:3572–3574, 2005.

[15] N. Kashtan, S. Itzkovitz, R. Milo, and U. Alon. Efficient sampling algorithm for estimating sub-graph concentrations and detecting network motifs. *Bioinformatics*, 20:1746–1758, 2004.

[16] S. Wernicke. Efficient detection of network motifs. *IEEE/ACM transactions on computational biology and bioinformatics*, 3(4):347–359, December 01 2006.

[17] S. Wernicke and F. Rasche. Fanmod: a tool for fast network motif detection. *Bioinformatics*, 22:1152–1153, 2006.

[18] Zahra Kashani, Hayedeh Ahrabian, Elahe Elahi, Abbas Nowzari-Dalini, Elnaz Ansari, Sahar Asadi, Shahin Mohammadi, Falk Schreiber, and Ali Masoudi-Nejad. Kavosh: a new algorithm for finding network motifs. *BMC Bioinformatics*, 10(1):318, 2009.

[19] Jin Chen, Wynne Hsu, Mong Li Lee, and See-Kiong Ng. Nemofinder: Dissecting genome-wide protein-protein interactions with meso-scale network motifs. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '06, pages 106–115, New York, NY, USA, 2006. ACM.

[20] Xin Li, Douglas S. Stones, Haidong Wang, Hualiang Deng, Xiaoguang Liu, and Gang Wang. NetMODE: Network motif detection without nauty. *PLoS ONE*, 7(12):e50093, 2012.

[21] Saeed Omidi, Falk Schreiber, and Ali Masoudi-Nejad. Moda: An efficient algorithm for network motif discovery in biological networks. *Genes and Genetic Systems*, 84(5):385–395, 2009.

[22] I. Xenarios, L. Salwinski, X. J. Duan, P. Higney, S. M. Kim, and D. Eisenberg. DIP, the database of interacting proteins: a research tool for studying cellular networks of protein interactions. *Nucleic acids research*, 30(1):303–305, Jan 1 2002.

Spatial Modeling of Flood Risk in Karawang

Bambang Riadi, Yustisi Ardhitasari Lumban-Gaol[*] and Rizka Windiastuti

Geospatial Information Agency, Jalan Raya Jakarta-Bogor Km.46, Cibinong 16911, Indonesia

A B S T R A C T

Conceptually, flood modeling can be done based on process and data availability (data driven model). The implementation of flood mitigation requires accurate data, which can be obtained from high resolution satellite imagery, vertical air photos and high-resolution Digital Elevation Model (DEM). Modeling using high accuracy DEM data can detect landforms in more detail. The objective of this research was to develop flood risk assessment using Topographic Wetness Index (TWI) method. Areas with high humidity indicated that the area was a flood hazard area. Flood spatial risk analysis used hazard mapping methods, vulnerability mapping and capacity assessment. The results obtained indicated that 8,838 ha area was in high flood hazard, while 9,780 ha was in moderate flood hazard and 7,652 ha was in low flood hazard. When incorporating the risk factors, 12,898 ha areas was in high risk, 9,357 ha was in moderate risk, and 4,015 ha was in low risk.

1. Introduction

Land form has an important role when studying the flood, because the shape of the land is one form of cross-section for the process of water flowing into the sea [1]. The areas that are most affected by flood are areas with flat reliefs and ramps. The indicators of flood-prone landscapes are in the form of flood plains, sea terraces, swamps and back swamps. Geomorphological flood areas are characterized by concave morphology or flat land forms associated with rivers, with winding and or meander flow patterns, as one area with potential flooding and this area needs to be mapped [2].

The research located in Karawang Regency was chosen because it met the criteria as a flood-prone area, which is located on a coastal area with slope characteristics < 2%. With an average rainfall of 200 mm per year, this study area is categorized as a flood hazard area [3]. The development of geospatial technology gives good hope because flood-prone areas can be determined more quickly, precisely and accurately. The use of DEM data is expected to provide efficiency in identifying and mapping flood-prone areas [4].

Several flood studies had been conducted using DEM SRTM (Shuttle Radar Topography Mission) data and topography maps at 1:25,000 scale, including [5] and [6]. Some research on floods which is conducted by [7] and [8] use data DEM 15M radar, which is better than the DEM SRTM 30m resolution. The data used in

this flood study, when used as a basis for making contours, would yield the accuracy of ± 10m or in the accuracy category of topography map scale 1: 25,000. Floods that occur in the study area are between 70 cm to 150 cm. If the flood research use DEM data with a precision of 10m, then the results of the research will not meet the standard accuracy.

Flood research in this study used TerraSAR-X DEM data with high spatial resolution and could be an alternative as an accurate source of DEM data [9]. The high spatial resolution of the TerraSAR-X sensor provide access to surface heterogeneity on a better scale [10]. The method used in this study is the Topographic Wetness Index (TWI), which could provide an illustration of an area that have a high flood hazard potential by using the accumulation of flow and slope functions to display the potential flood basin area [11].

Flood risk was analyzed for spatial and temporal aspects of a flood event. Spatial aspects involved the location and coverage of floods, while the temporal aspect included the time of the flood event. Flood modeling could be carried out conceptually based on processes and models according to data availability (data driven models). Flood hazard modeling in paddy fields was assessed from the physical and hydrological conditions of the land supported by statistical data and the results of field surveys [12]. This approach was considered suitable for assessing flooding in paddy fields because it fitted to the facts in the field. Flood risk assessment was carried out by identifying three components, namely hazard,

vulnerability and flood exposure. The objectives to be achieved from this study were: 1. identifying and delineating flood hazard areas using accurate DEM in Karawang regency, and 2. analyzing the risk of flooding in paddy fields.

2. Materials and Method

2.1. Investigated field

The research location was in Karawang Regency, West Java Province (Figure 1), which is geographically located in the north of West Java Province, between longitude 107º 02 '- 107º 40' BT; and latitude 5º 56 '- 6º 34' LS. The study area consisted of 30 districts, 309 villages with land use dominated by technical irrigated paddy field (97,352 ha or 65.44%), and simple irrigated paddy fields (2,204 ha). Other land uses included fields (5,830 ha), yards (44,931 ha) and uncultivated (1,817 ha) (BPS 2016). The land use condition of Karawang Regency reflected the characteristics of the rice farming area and as a rice granary in West Java Province.

Figure 1. The location of Karawang Regency, West Java.

2.2. Materials and data used

This research was carried out in three stages as follows: (a) preparation and data collection of the topography map and TerraSAR-X DSEM (TerraSAR-X DSM was exported to the 16 bit geotiff and then imported into the ILWIS system), (b) data processing for identification of flood hazard areas with TWI (Topographic Wetness Index) method (through the fill sink process, flow direction, flow accumulation and compound index calculation which resulted in a wetness index), and (c) risk analysis and results validation, where risk analysis was preceded by an analysis of land physical vulnerability and vulnerability. Risk analysis was done by overlaying flood hazard maps with physical vulnerability maps and socio-economic vulnerability maps using ArcGIS software and generating risk maps. Validation was carried out by checking the modeling results at 20 points and confirming with the local government or the community at the observation points of flooded areas. The flowchart is shown in Figure 2.

2.3. Flood Hazard

Identification and delineation of flood hazard areas were done by recognizing damp areas using the Topographic Wetness Index (TWI) method. The result was a derivative map from processing DEM data under fixed conditions - steady state - using the flow and slope accumulation functions. The wetness index was a key variable that controls the hydrological process [13]. The TWI model described water trends that accumulated at certain points and local slopes which indicated the influence of gravitational forces on the movement of water [2]. A location that had a gentle slope would have a high wetness index and had a high potential for waterlogging. Conversely a location that had a steep slope would have a low wetness index, so the potential for flooding was also low.

The wetness index was a key variable in the hydrological process. The wetness index value of more than 25 indicated that the area had a gentle slope so it had the potential for waterlogging. The value of the wetness index 12 in this study was assumed to be the inundation area boundary. If the wetness index was lower than 12, it indicated an area with steep slope and had a low potential for flooding.

Table 1. Types of data, data sources, analysis method and output.

No.	Research objevtive	Data	Data source	Analytical method	Output
1.	Identification and delineation of flood hazard areas	Topographical map 1:25000, DSM TerraSAR-X,high resolution satellite imagery,irigation network map	BIG, BPN, PJT2, DPKPP, field survey	Topogra phic Wetness Index, (TWI), digitization,visual interpretation,	Flood hazard map
2.	Flood risk analysis	Flood risk map, high resolution image, flood history,rainfall	Analysis results, BPS, DPKPP, PJT 2, field survey	GIS analysis, land physical vulnerability, socioeconomic vulnerability, overlay vulnerability map	Flood risk map
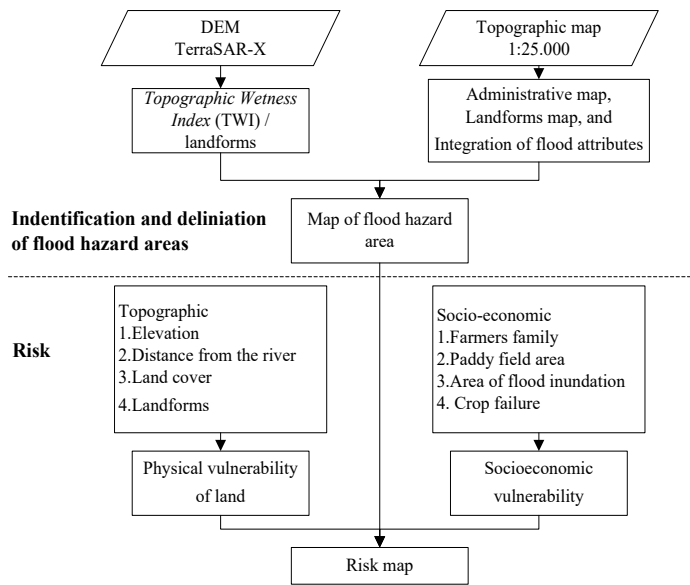
Figure 2. Research flowchart.

Withdrawal of flood inundation boundary areas was carried out through visual interpretation. The flood event attributes in the form of frequency data, inundation height and flood duration in paddy fields [14] were integrated into the flood hazard map [15]. Floods that occurred every year with a puddle duration of more than 7 days with an average height of flood water more than 70 cm, had the potential to become a flood hazard area with high criteria [16], as compiled in Table 2.

Table 2. Criteria for classification of floods in paddy fields.

Frequency/ year	Average flood height (cm)	Duration of inundation/day	Criteria
1 - 2	60 - 150	< 7 days	Low
2 - 3	60 - 150	7 - 14 days	Moderate
> 4	> 70	≥ 15 days	High

2.4. Risk

Vulnerability is a condition of reduced resilience due to the effects of floods that threaten lives, livelihoods, natural resources, infrastructure, economic productivity, and welfare. The relationship between danger and vulnerability, that vulnerability produces a risk or interaction between the level of vulnerability of the region and the threat of danger [17]. The impact of flood has risks that can cause damage and potentially cause losses, loss of security, and disruption of community activities [18]. The overlay method between flood hazard maps and vulnerability maps produced a risk map which was then classified into 3 classes, namely high risk, moderate risk, and low risk [19]. Risk criteria are arranged in a risk level matrix as in table 3.

Table 3. Analysis of risk level matrix.

	Pv1	Pv2	Pv3
Sv (1)	3	4	5
Sv (2)	4	5	6
Sv (3)	5	6	7

Pv = Physical vulnerability of land; Sv = Socioeconomic vulnerability.

Source: analysis result

3. Results and Discussion

3.1. Flood Hazard

The TWI process was through several stages:

(a) Correction of depression / sink due to noise with the fill sink process (Figure 3). The results showed that the northern region had a low elevation value with the number of colors 0 to 50. The number of colors did not reflect the value moisture but represented the height of the area.

(b) Flow direction process, where flow direction (Figure 4) spread throughout the study area which indicated that the study area was a potential area of flooding.

(c) Flow accumulation, where the higher the value of flow accumulation was, the more likely the cell became a stream, regardless of the real conditions on the cell field in the form of rivers, ditches or irrigation channels (Figure 5). Values of flow accumulation were associated with the basin and or direction of gathering water flow. Large values were associated with high flow density. This function modeled the amount of accumulated water flow that occurred in a particular area, the value of water accumulation which was usually also identical to the actual hydrological flow in the field and this area had an indication of a flood-prone area.

(d) The calculation compound index that produced wetness index. Locations that had a gentle slope had a high wetness index that gave a high indication of potential flood areas. Locations that had steep slopes, had low wetness indices that gave a low indication of potential flooding (Figure 6). The limit of the flood potential area was delineated in the wetness index of 12, and the flood hazard area map was the result of the integration of the TWI map by adding historical attributes to flood events (Figure 7).

Figure 8, as a result of overlaying flood hazard maps with landform maps, showed that the majority of flood hazard areas were in the plain estuarine land. Overlaying flood hazard maps with slope maps showed that flood areas were mostly located in areas with slopes < 2% (Figure 9). The analysis showed that the flood hazard area covered 26,270 ha (Figure 10.), where 8,838 ha (part of Jayakerta district, Rawamerta district, Telukjambe Barat district and Telukjambe Timur) was in high flood hazard, 9,780 ha was in moderate flood hazard and 7,652 ha was in low flood hazard.

3.2. Vulnerability

Vulnerability to floods was classified in land physical vulnerability and socio-economic vulnerability. The physical
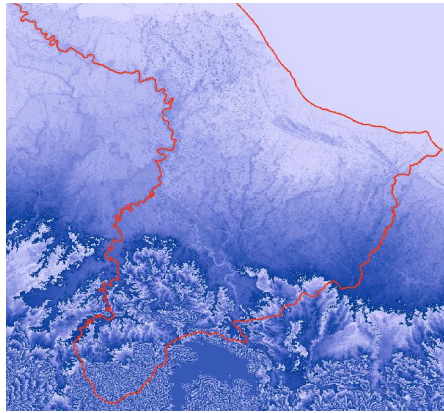
Figure 3. Fill sink.
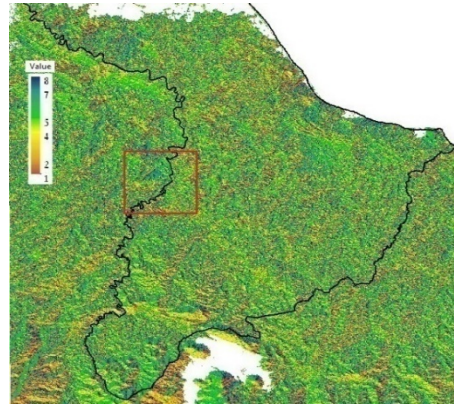
Figure 4. Flow direction.
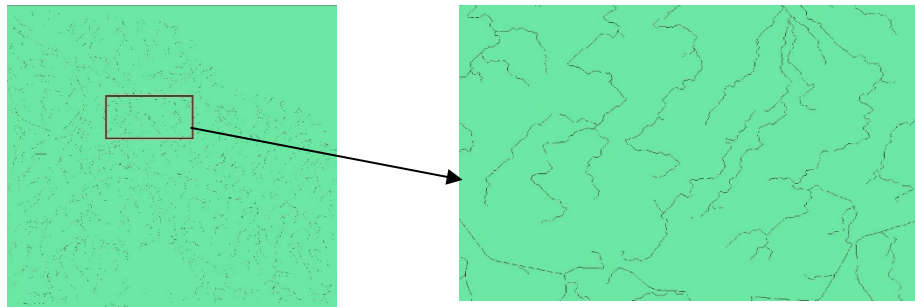
Figure 5. Flow accumulation.
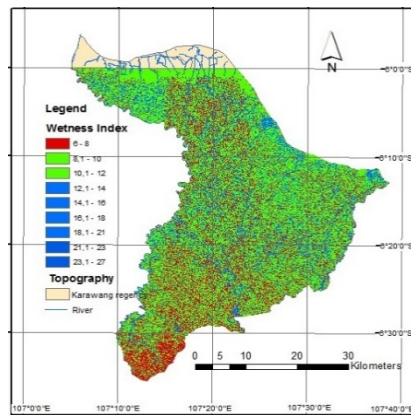
Figure 6. TWI Map.
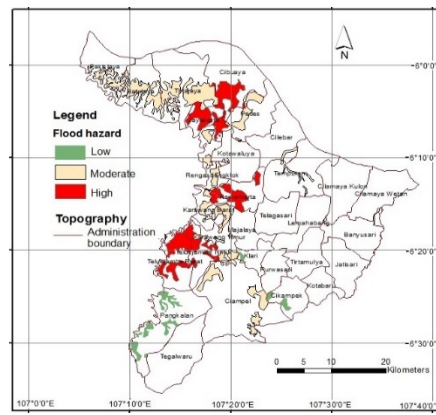
Figure 7. Flood Hazard Map.
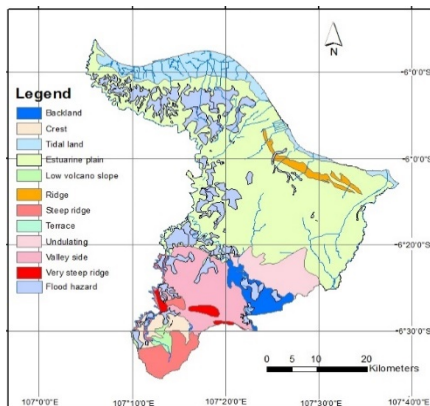
Figure 8. Landforms Map
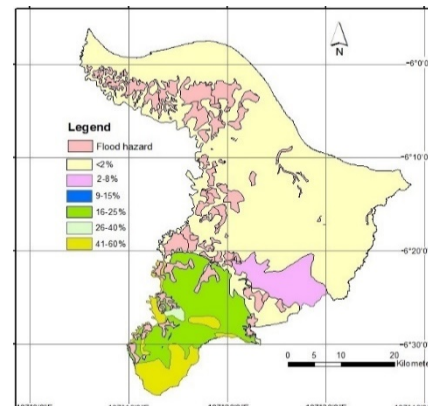
Figure 9. Slope Map
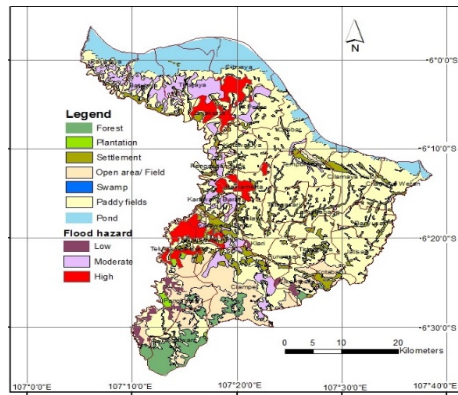
Figure 10. Flood Area.
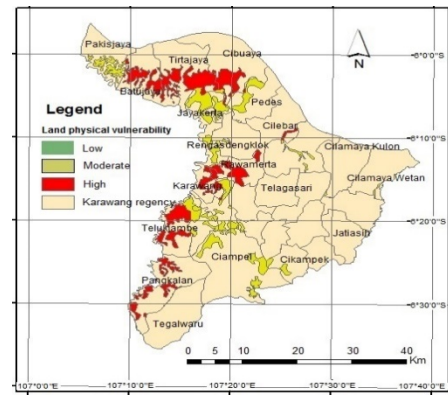
Figure 11. Land Physical Vulnerability.
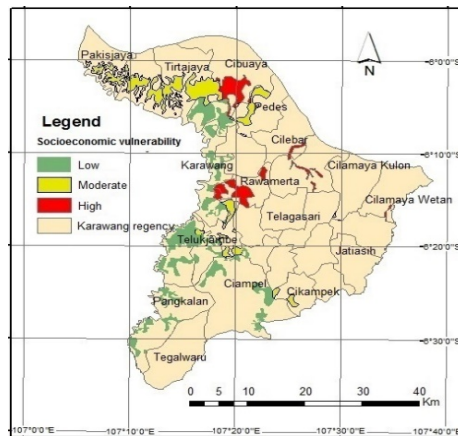
Figure12. Socio-economic Vulnerability.
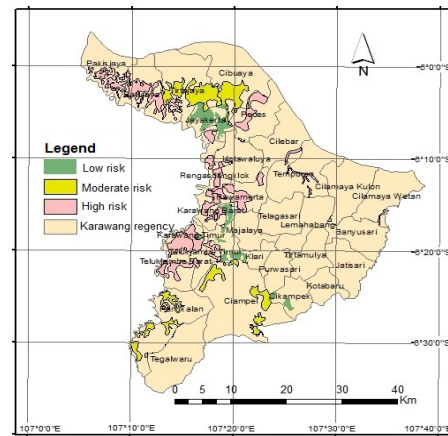
Figure 13. Risk Map.

Table 4.Flood risk level.

No.	Districs	Flood hazard	Land physical	Socio economic	Risk	Classified
1	Batujaya	2.1	3.0	3.0	8.1	High
2	Ciampel	2.2	2.1	2.1	6.4	Moderate
3	Cibuaya	1.2	2.7	2.7	6.6	Moderate
4	Cikampek	0.9	1.8	1.8	4.5	Low
5	Cilamaya Kulon	1.3	2.4	2.4	6.1	Moderate
6	Cilamaya Wetan	1.3	2.7	2.7	6.7	Moderate
7	Cilebar	1.8	2.7	2.4	6.9	High
8	Jayakerta	2.7	2.4	3.0	5.4	Low
9	Karawang Barat	1.8	3.0	2.1	6.9	High
10	Karawang Timur	1.8	2.1	2.1	4.2	Low
11	Klari	1.3	2.1	2.1	5.5	Low
12	Pakisjaya	2.7	2.1	3.0	7.8	High
13	Pangkalan	1.3	3.0	2.4	6.7	Moderate
14	Pedes	2.1	2.4	2.7	7.2	High
15	Rawamerta	2.7	2.7	2.7	8.1	High
16	Rengasdengklok	1.8	2.7	2.4	6.9	High
17	Telukjambe Barat	2.5	2.4	2.4	7.3	High
18	Telukjambe Timur	2.5	2.4	2.4	7.3	High
19	Tempuran	2.2	2.4	2.4	7.0	High
20	Tirtajaya	1.3	2.7	2.7	6.7	Moderate

vulnerability of the land was related to land use, height to sea level and distance to surrounding rivers. Physical susceptibility of high-class paddy fields included Batujaya, Cibuaya, Cilebar, West Karawang, Pangkalan, Rawamerta, West Telukjambe and Tirtajaya districts had high physical vulnerability to land (Figure 11).

The socio-economic vulnerability of the community in the event of a flood was caused by a socio-economic system in which population lived [20]. Socio-economic vulnerability analysis took into account the number of affected farmer families, productivity, production losses, and economic losses incurred. The highest socio-economic vulnerability was in the flood areas of Cibuaya, Cilamaya Wetan, Cilebar, Rawamerta, Tempuran sub-districts (Figure 12). Areas that had high vulnerability could lead to a greater element at risk to be exposed to hazards and could increase the risk of flooding.

3.3. Risk

Floods that often occurred in research areas generally inundated paddy fields, because paddy fields were relatively lower than roads and settlements. For this reason, the flood research used GIS technology (Geographic Information System). Spatial integration with socio-economic data was carried out to create maps of land physical vulnerability and socio-economic vulnerability of farmers. Furthermore, analyzing the area of paddy fields that had a high risk of flood events. This relationship can simply be understood that the risk of flooding (Table 3) would increase if the level of hazard and the level of vulnerability were high.

Based on the results of the analysis 12,898 ha of rice fields in the Batujaya, Cilebar district, Karawang Barat district, Pakisjaya district, Pedes district, Rawamerta district, Rengasdengklok district, Telukjambe Barat district, Telukjambe Timur district and Tempuran district was in high risk of flood, 9,357 ha was in medium risk and 4,015 ha was in low risk (Figure 13).

4. Conclusion

Mapping of flood-hazard areas using high resolution DEM data with the TWI method was successfully implemented and gave good results. This research was intended to encourage the use of TerraSAR-X DEM for various flood research applications. At present, flood research is still using topographic map data of 1: 25,000 scale. Conclusion of flood hazard mapping as follows:

1. The results of the identification of flood hazards indicate that not all lowland estuaries were flooded.
2. Areas with a slope of less than 2% are areas that have the potential to flood, in this study flood hazard areas were also identified in areas with slopes of more than 16%, with this method the basin area can be identified in more detail.
3. Potential flood hazards include 26,270 ha, areas with high flood hazard 8,838 ha, medium flood hazard 9,780 ha and low flood hazard 7,652 ha.

Flood events threaten the lives and livelihoods of the people, one of the livelihood sectors that is threatened by flooding is the agricultural sector. Farmer livelihood vulnerability will present the risk of flooding. Based on the results of flood risk research, it can be concluded that:

1. High risk areas are Batujaya district, Cilebar district, Karawang Barat district, Pakisjaya district, Pedes district, Rawamerta district, Rengasdengklok district, Telukjambe Barat district, Telukjambe Timur district and Tempuran district covering an area of 12,898 ha. Medium risk 9357 ha and risk low of 4,015 ha.
2. The understandable relationship is that the risk of flooding will increase if the level of hazard and vulnerability is high.
3. Analysis of flood risk at the previous district level cannot be avoided, the equation for the accuracy of the basic data geometry used is very possible.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Marfai MA. 2004. Spatial Modelling of Tidal Hazard on East Semarang Coastal Area. (Studi Kasus: Pesisir Timur Semarang). *Jurnal Fakultas Geografi UMS: Forum Geografi* 18(1): 60-69.

[2] Pourali SH, Arrowsmith C, Chrisman N, Matkan AA, Mitchell D. 2014. Topography Wetness Index Application in Flood-Risk-Based Land Use Planning. *Appl. Spatial Analysis* 9(1): 39-54.

[3] SNI. 2015. Standar Nasional Indonesia SNI 8197: 2015 tentang Metode Pemetaan Rawan Banjir Skala 1:50.000 dan 1:25.000. *Badan Standardisasi Nasional*.

[4] Baghdadi N, Cerdan O, Zribi M, Auzet V, Darboux F, El Hajj M. 2008. Operational Performance of Current Synthetic Aperture Radar Sensors in Mapping Soil Surface Characteristics: Application to Hydrological and Erosion Modelling. *Hydrological Processes*. 22(1): 9-20.

[5] Deviana A, Kridasantausa I, Suryadi Y. 2012. *Kajian Pemodelan Spasial Banjir Untuk Mendukung Kebijakan Sempadan Sungai dan Tata Ruang Wilayah* (Studi Kasus Wilayah Pengembangan Baleendah). Tersedia di: https://ftsl.itb.ac.id/wp-content/uploads/sites/8/2012/07/95010011-Aninda-Deviana.pdf. pp. 1-21

[6] Nugraha Al, Haniah. 2013. Kajian Pemanfaatan DEM SRTM dan Google Earth Untuk Parameter Penilaian Potensi Kerugian Ekonomi Akibat Banjir Rob. *Jurnal Ilmiah Bidang Ilmu Kerekayasaan* 34(3): 202-210.

[7] Bates PD and De Roo APJ. 2000. A Simple Raster-Based Model for Flood Inundation Simulation. *Journal of Hydrology*. 236 (2000): 54-77. Elsevier.

[8] Haile AT and Rientjes THM. 2005. Effects of Lidar DEM Resolution in Flood Modelling: A Model Sensitivity Study for The City of Tegucigalpa, Honduras. *ISPRS WG III/3, III/4, V/3 Workshop "Laser scanning 2005"*, pp. 168-173. Enschede, the Netherlands.

[9] Riadi B, Yustisi ALG, Bimo W, Pranadita S. 2018. Vertical Accuracy Assessment of DSM from TerraSAR-X and DTM from Aerial Photogrammetry on Paddy Fields – Karawang, Indonesia. *Advances in Science, Technology and Engineering Systems Journal* 3(4): 187-192.

[10] Somantri L. 2008. Pemanfaatan Teknik Penginderaan Jauh Untuk Mengidentifikasi Kerentanan dan Risiko Banjir. *Jurnal Gea* 8(2): 1-6.

[11] Hojati M. 2016. Determination of A Topographic Wetness Index Using High Resolution Digital Elevation Models. *European Journal of Geography*. 7(4): 41-52. December 2016. Association of European Geographers.

[12] Agus F and Mulyani A. 2006. Judicious Use of Land Resources for Sustaining Indonesian Rice Self Sufficiency. *Proceedings International Rice Conference, 12-14 Sept. Denpasar, Bali*. Indonesian Institute of Rice Research, Sukamandi. pp.121- 133.

[13] Buchanan BP, Fleming M, Schneider RL, Richards BK, Archibald J, Qiu Z, Walter MT. 2014. Evaluating Topographic Wetness Indices Across Central New York Agricultural Landscapes. *Hydrology and Earth System Science* 18:3279–3299.

[14] DPKPP. 2016. Dinas Pertanian Kehutanan Perkebunan dan Peternakan. Laporan Potensi Pertanian di Kabupaten Karawang.

[15] Hartini S, Hadi MP, Sudibyakto S, Poniman A. 2015. Risiko Banjir pada Lahan Sawah di Semarang dan Sekitarnya. *Majalah Ilmiah Globe*. 17(1): 51-58. Badan Informasi Geospasial.

[16] [PSBA] Pusat Studi Bencana Alam UGM dan Bakosurtanal. 2005. Laporan Pengkajian Model Pemetaan Risiko Bencana Alam-Yogyakarta.

[17] Raharjo PD. 2009. Pemodelan Hidrologi Untuk Identifikasi Daerah Rawan Banjir di Wilayah Surakarta dengan SIG. *Jurnal LIPI: Limnotek.* 16(1) :1-9. 2009.

[18] Fajar Y, Marfai MA, Parwati, Suwarsono. 2009. Model Simulasi Luapan Banjir Sungai Ciliwung di Wilayah Kampung Melayu–Bukitduri Jakarta. *Jurnal Penginderaan Jauh.* 6: 43-53.

[19] [BNPB] Peraturan Kepala Badan Nasional Penanggulangan Bencana 2012, tentang Pedoman Umum Pengkajian Risiko Bencana. BNPB, Jakarta.

[20] Rosyidie A. 2013. Banjir Fakta dan Dampaknya, Serta Pengaruh dari Perubahan Guna Lahan. *Jurnal Perencanaan Wilayah dan Kota*. 24(3): 246-257.

An Experimental Investigation of Reconfigurable UWB Modified Octagonal Microstrip Monopole Patch Antenna with Switchable and Tunable Band-Notched Characteristic

Ruchi Paliwal[*,1], Ruchi Paliwal[1], Rajesh Kumar Singh[2], Shiban Kishen Koul[3]

[1]Dept. of Electronics and Comm. Engg., JSS Academy of Technical Education, Noida (UP), 201301, India

[2]Bharti School of Telecomm. Tech. and Management, IIT Delhi, 110016, India

[3]CARE, IIT Delhi, 110016, India

A R T I C L E   I N F O

A B S T R A C T

An experimental investigation of reconfigurable ultra-wideband (UWB) modified octagonal microstrip monopole patch antenna with switchable and tunable band-notched characteristic is presented in this paper. Notched-bands are obtained by connecting or disconnecting the open shunt stubs. The switching is achieved by using PIN diodes. Additionally, more notch-bands are created by adding additional stubs to the original design and tuning is achieved by employing a varactor diode in the circuit. The proposed antenna has more freedom to suppress the interfering bands because of tuning nature. The proposed designs are simulated, fabricated and then tested. The proposed structures are simple in geometry and useful to operate in the rich electromagnetic scattering environment.

1. Introduction

In today's world, wireless communication system becomes an essential need for communication. It requires an antenna that must be effective to change its characteristics in reversible manner and it can be utilized for many applications and functions. With the help of transformable antenna, we can replace multiple antennas by a single antenna element which can utilize the complete unlicensed ultra-wideband (UWB) [1]. On this ground, Planar UWB antennas are in great demand [2-7]. While using the entire bandwidth of UWB, other signals that exist in wireless communication system like Worldwide Interoperability for Microwave Access (WiMAX), wireless local area network (WLAN), X-band downlink, X-band satellite communication link, X-band aeronautical radio navigation, etc., operating around 3.5, 5, 5.8, 7.5, and 9 GHz band, respectively, can cause interference. To mitigate this issue, band-notched UWB antennas have been designed that can filter out interfering frequency and improved selectivity. The band-notched UWB antennas have been presented in the past few years.

By inserting a proper slots or slits with different shapes in a radiating element, notched-band can be achieved [3]. In [4, 5], one notch is created while in [7, 8, 9, 10, 11, 12] two notched-bands

are obtained. In [13], notched-band has been created by using electromagnetic band-gap (EBG) structure. In [14], one notch band is achieved from CPW-fed monopole antenna, but the structure is compact. Wide notch bands have also been achieved in [15]. Besides this, antenna with static feature will become obsolete for next generation wireless systems. Antenna with reconfigurable feature can be a good candidate for these systems. Reconfigurable antennas have been created by connecting and disconnecting different conducting elements to the main geometry with the help of electrical switches like radio frequency micro-electro-mechanical systems (RF MEMS), PIN diodes or by other methods. Reconfigurability can also be achieved with optical switch or by using material change [16]. A reconfigurable UWB antenna with tuning notched-bands could be advantageous in the future wireless communication systems. Variable notched frequency bands have been achieved in [17], [18]. Some compact UWB monopole antennas reported in the literature can be integrated in a compact system and save space [19], [20]. A reconfigurable UWB modified octagonal monopole microstrip patch antenna with switchable and tunable notched-bands is discussed in this paper. The work proposed in this paper is an extended work originally presented in [21]. The proposed antenna has more freedom to suppress the selected bands.

The paper starts with the introduction. Section 2 describes the design of a reconfigurable UWB monopole patch antenna. Section 3 discusses the modified design to create more notched-bands by adding thin stubs. Section 4 presents the tuning in the structure by using a varactor diode and finally concludes the paper with results and discussions.

(a)   (b)
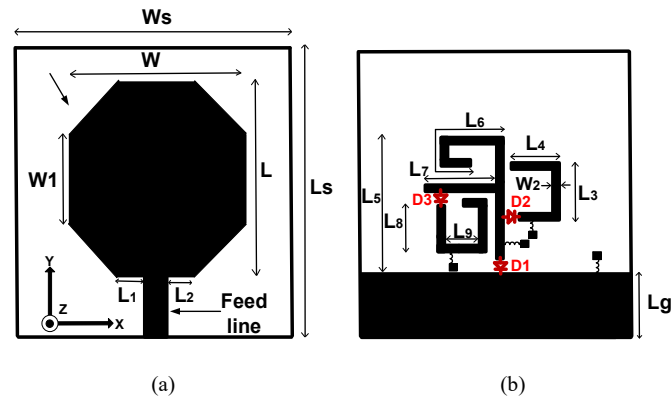
Figure 1. Geometrical model of the modified octagonal patch antenna (a) front side (b) back side.
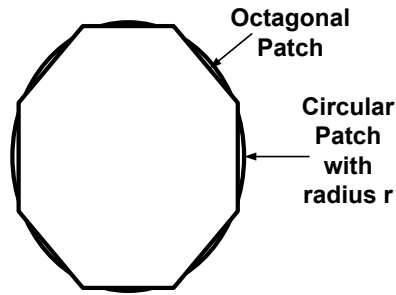
Figure 2. Geometry of a conventional octagonal patch and circular patch

Table 1: Dimensions of the proposed reconfigurable UWB monopole antenna

Parameter	Value (mm)	Parameter	Value (mm)
L	25.50	L8	04.65
L1	04.00	L9	03.30
L2	04.00	Ls	42.00
L3	03.40	W	22.50
L4	04.23	W1	14.50
L5	09.30	W2	00.60
L6	05.50	Ws	40.00
L7	06.60		

2. Reconfigurable UWB monopole patch antenna design and its operation

2.1. UWB antenna design

The geometrical model of the antenna is illustrated in Figure 1. The antenna is designed on FR4 substrate with relative permittivity of 4.4. The height of the substrate is 1.6 mm. The proposed antenna consists of modified octagonal monopole microstrip patch etched on the upper side and partial ground plane on the lower side of the substrate. Thin metal stubs are etched on the back side of the patch to create notches in the reflection coefficient plot and this is

achieved by integrating PIN diodes in the circuit as shown in Figure 1(b). The conventional octagonal (with equal side length) patch is designed by taking its size equivalent to a circular patch as shown in Figure 2. The dominant mode is TM_{110} (TM_{mnp}), the resonant frequency can be calculated as

$$f_r = \frac{1.8412\,c}{2\pi r\sqrt{\varepsilon_r}} \qquad (1)$$

Where

c = Speed of light,

r = Radius of the circular patch,

εr = Relative permittivity.

(a)   (b)
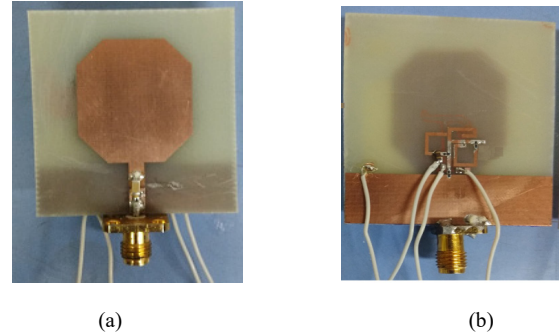
Figure 3. Photograph of a modified octagonal monopole microstrip patch antenna (a) front side (b) back side.

The radius is calculated as 14.05 mm by equating the area of both structures (circular and octagonal patch). The resonant frequency of the dominant mode is 2.94 GHz. The proposed octagonal patch is modified by varying the edges and using partial ground plane and achieved impedance matching throughout the band (3.1-10.6 GHz). The proposed antenna is excited with a 50 Ω line. There are total of three PIN diodes named as D1, D2, and D3 are connected at different places to achieve different notches. PIN diodes are turned ON with the help of biasing circuits at the back side of antenna. The dimensions of the circuit are listed in Table 1. Photograph of the fabricated circuit is shown in Figure 3.

2.2. Operation and results

Three PIN diodes are used to get various states from the circuit. Simulations are done in an EM simulator (Computer Simulation Technology software) [22]. The equivalent model of a PIN diode used in the simulations is shown in Figure 4. The values of Rp, Cp and Rs are 40 kΩ, 0.03 pF and 5 Ω, respectively. The parasitic series inductance is ignored in the simulations. The equivalent model is extracted from the datasheet [23]. Inductors of 56 nH are used to bias the PIN diodes. A capacitor of 100 pF is used in the circuit to block dc. Input impedance of the proposed antenna is plotted in Figure 5. From
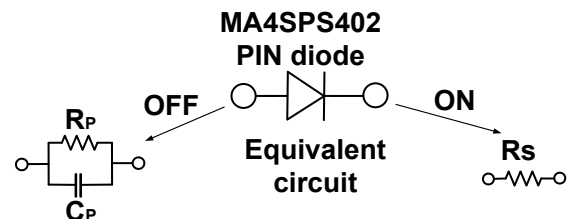
Figure 4. The equivalent circuit model of the PIN diode.

(a)

(b)
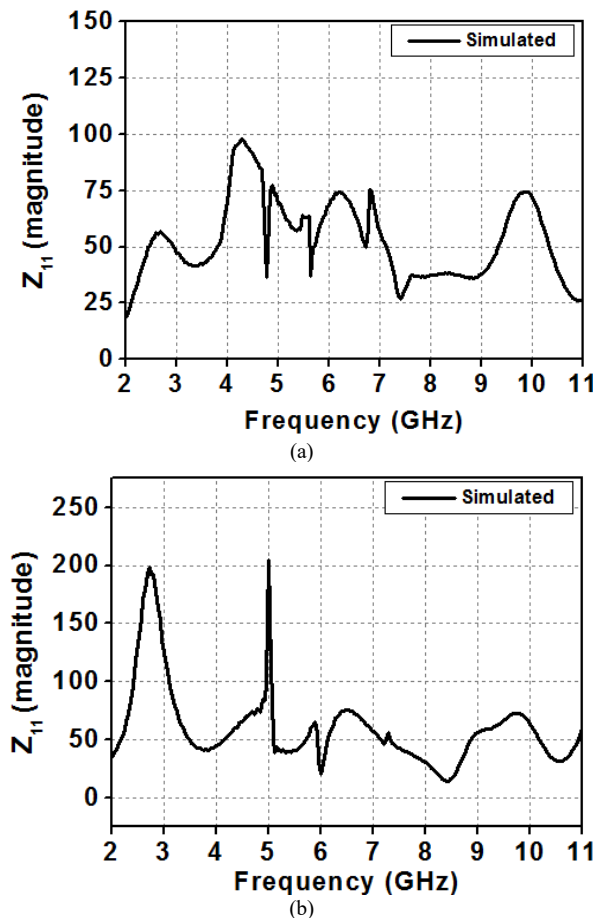
Figure 5. Input impedance of the proposed antenna (a) when all diodes are OFF, (b) diode D1 is ON.

Figure 5(a), it is clear that the input impedance lies in the range from 26 Ω to 96 Ω which corresponds to $S_{11} < -10$ dB. It is matched in the whole UWB band. Figure 5(b) shows the input impedance of 203 Ω ($S_{11} = -4.37$ dB) at 5.05 GHz and 13 Ω ($S_{11} = -4.62$ dB) at 8.25 GHz, which creates two notches. When all diodes are OFF, the impedance bandwidth is achieved in the range from 2.8 to 11.5 GHz as shown in Figure 6(a). In the second state, diode D1 is turned ON (L5 and L6 stubs are connected to the ground plane); two notched-bands are created around at 5.05 and 8.25 GHz as shown in Figure 6(b). In the third state, all three diodes are ON (all stubs are connected), the notched-bands are shifted to some other bands with centre frequencies of 5.5 and 8 GHz as shown in Figure 6(c) but in this case, the impedance bandwidth is reduced. At 3.1 GHz, the magnitude of reflection coefficient is below -10 dB which is not acceptable.

Further the design is modified to get good impedance match across the whole UWB band except the notched-bands.

3. Modified reconfigurable UWB antenna

The modified structure is shown in Figure 7. The geometry of the top side is same as the previous one; the only change is done on the back side (partial ground plane side). Photograph of the printed design is shown in Figure 8. Two additional stubs are connected in this modified geometry to get more notched-bands. There are 4 PIN diodes in the circuit. Here, measurements are done by taking PIN diode as a metal strip of 0.2 mm × 0.2 mm. Input impedance is plotted in Figure 9. It is perfectly matched in the range 2.8–11.4 GHz. When all four diodes are OFF, the reflection

coefficient is below -10 dB across the whole band (3.1-10.6 GHz) as depicted in Figure 10(a). UWB characteristic is achieved from the circuit without turning the diodes ON. In state 2, diode D1 is ON and others are OFF, two notched-bands are created as show in Figure 10(b). In state 3, diodes D1 and D2 are ON, one notched-band is present (around 8.5 GHz) and other notched-band (around 5 GHz) is suppressed because when we connect D2, one more shunt stub (L3 + L4) is connected to the circuit, this open ended stub (length = λ/4) suppress the 5 GHz band as shown in Figure 10(c). In state 4, diodes D1 and D3 are ON, notches are shifted as shown in Figure 10(d), because of change in the effective reactance value. In state 5, diodes D1 and D4 are ON; again, it suppresses the 5 GHz band and obtains the wide notched-band at 7 GHz (center frequency) as shown in Figure 10(e). In the last case (state 6), all diodes are ON, two notched-bands at 3.5 and 7.25 GHz are created as shown in Figure 10(f).

(a)

(b)

(c)
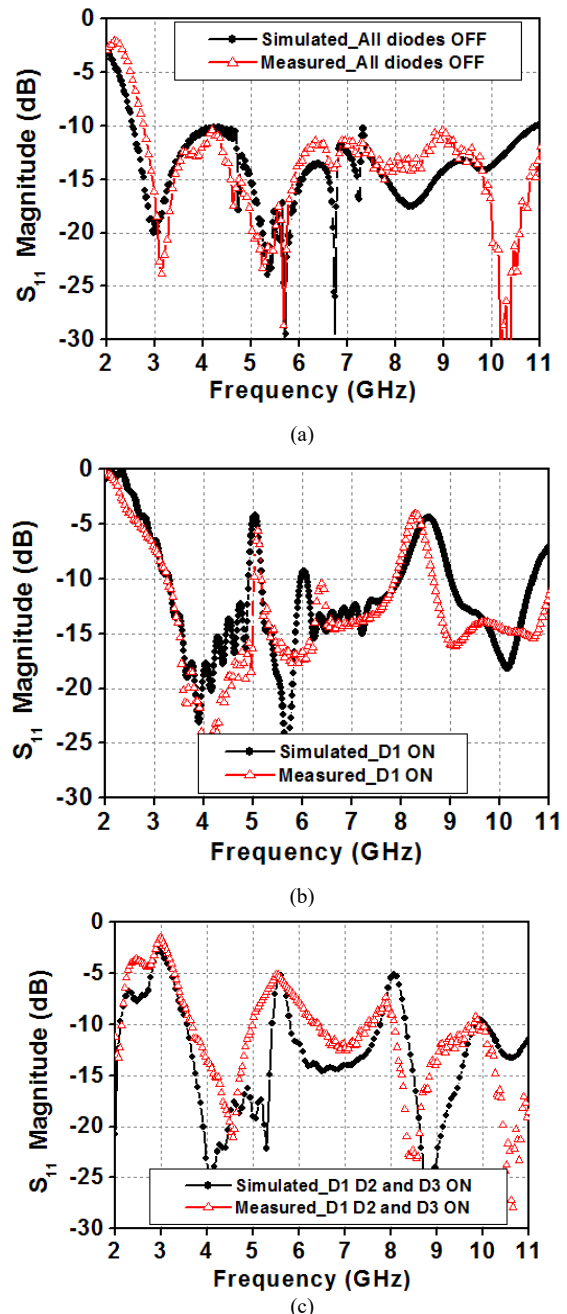
Figure 6. Reflection coefficients of the proposed antenna.
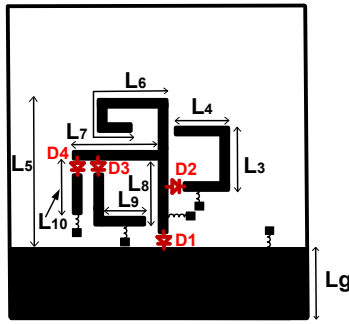
Figure 7. Proposed modified geometry. L3 = 03.40, L4 = 4.23, L5 = 9.30, L6 = 7.10, L7 = 6.3, L8 = 6.2, L9 = 3, L10 = 4.75, Lg = 10.25.
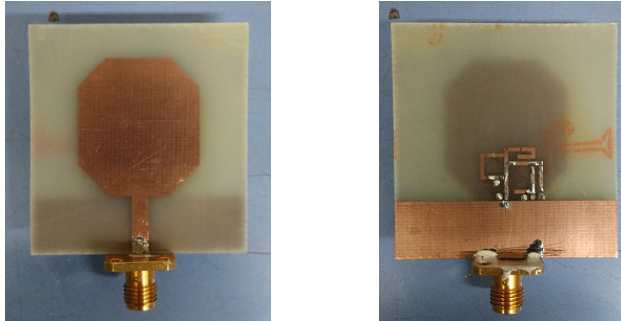
Figure 8. Photograph of the modified geometry in Figure 7.
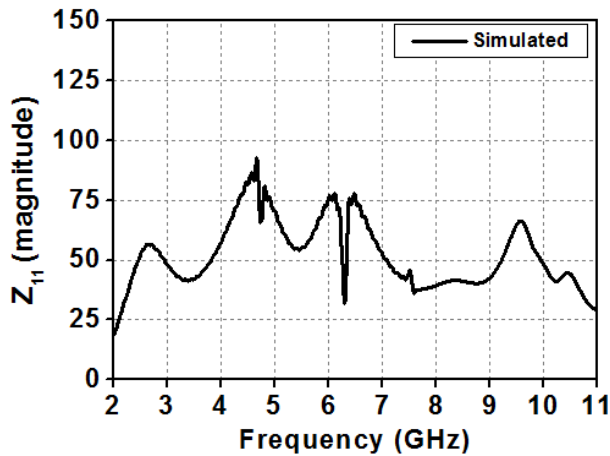
Figure 9. Input impedance of the proposed antenna when all diodes are OFF.

(a)

(b)

(c)

(d)

(e)
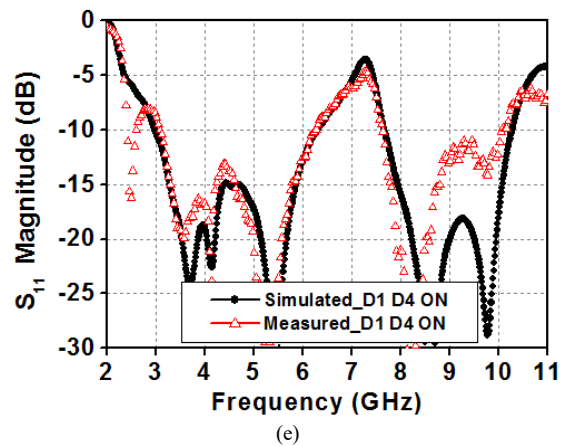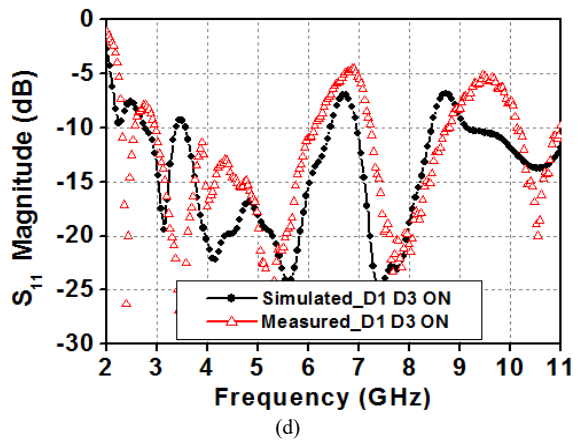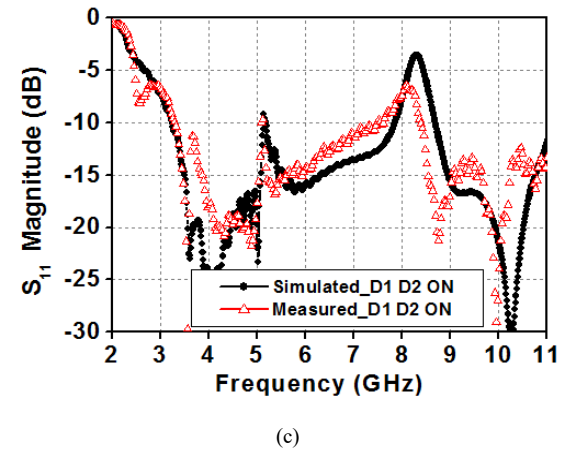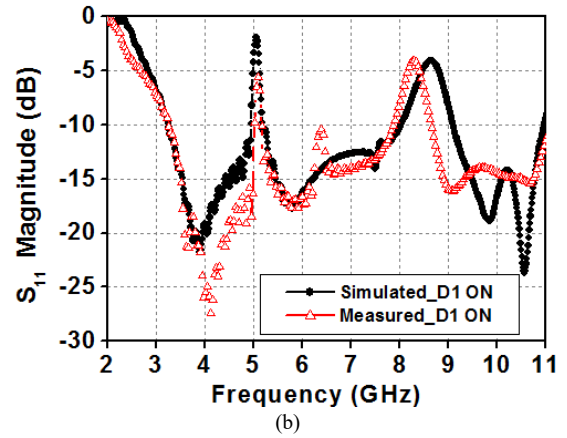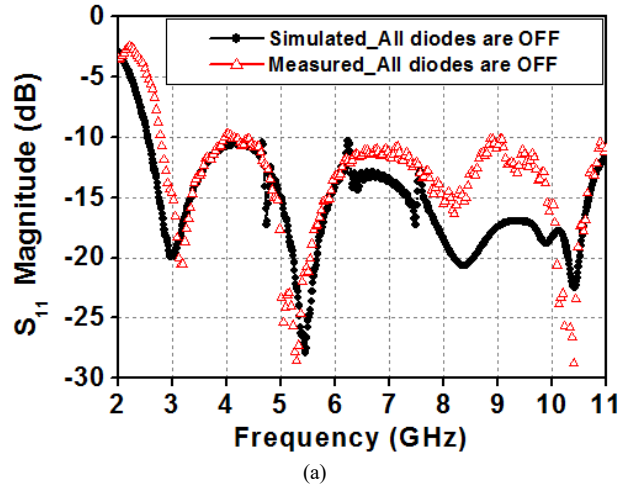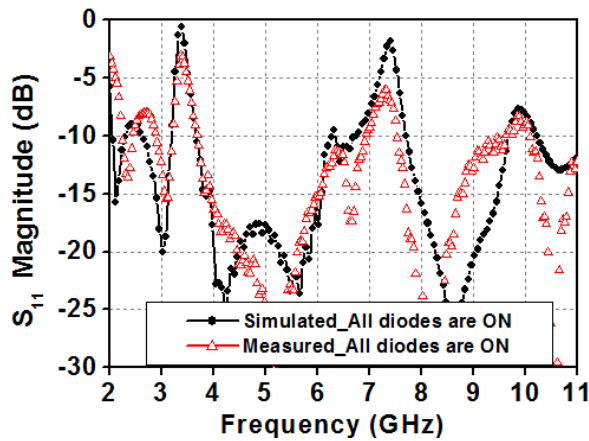
(f)

Figure 10. Reflection coefficients of the modified antenna as shown in Figure 7.

4. Switchable and tunable antenna design

The switchable and tunable reconfigurable UWB modified octagonal monopole patch antenna is proposed by modifying the structure as shown in Figure 11. Front side of the structure is same as used in the previous one (Section 2).
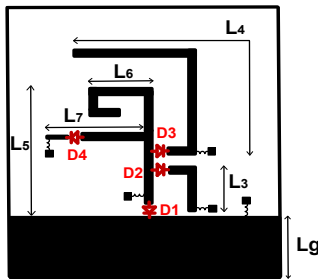
Figure 11. Switchable and tunable notch-band antenna: L3 = 5.1, L4 = 14.8, L5 = 9.45, L6 = 5.5, L7 = 11.2, Lg = 10.3.
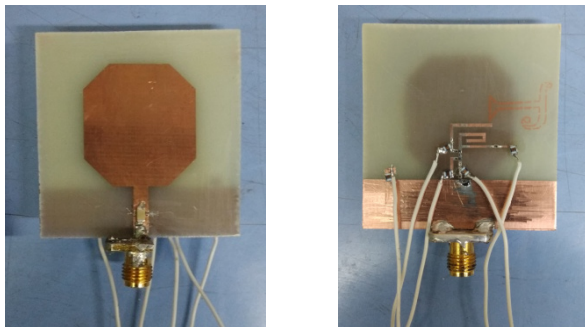
Figure 12. Photograph of the switchable and tunable notch-band antenna as given in Figure 11.

One varactor and three PIN diodes are used to make antenna switchable and tuneable. Photograph of the printed circuit is shown in Figure 12. An input impedance plot, with all diodes OFF, is shown in Figure 13. In the first state, when all diodes are OFF, the reflection coefficient is below -10 dB from 2.94 to 11.85 GHz as shown in Figure 14(a). In state 2, diode D1 is ON, two notched-bands are created at 4.8 and 8.25 GHz as shown in Figure 14(b). In state 3, diodes D1 and D2 are ON, the notched-band created at 8.25 GHz is suppressed by connecting L3 stub as shown in Figure 14(c). In state 4, diodes D1 and D3 ON, notched-bands are shifted
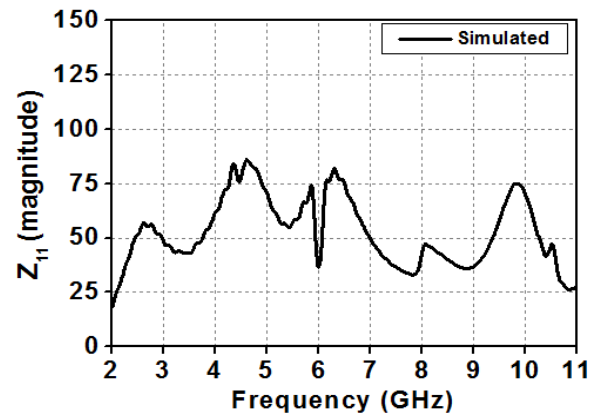
Figure 13. Input impedance of the proposed antenna when all diodes are OFF.

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

(i)
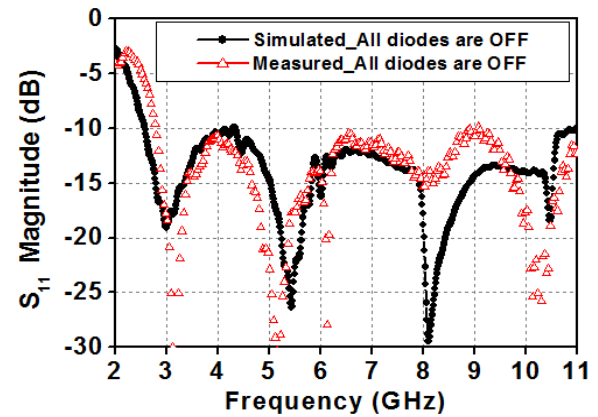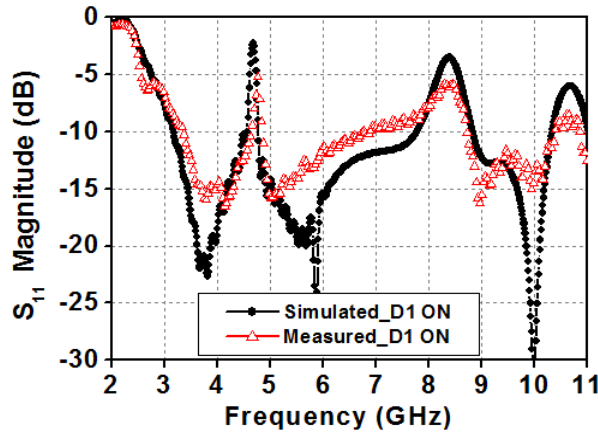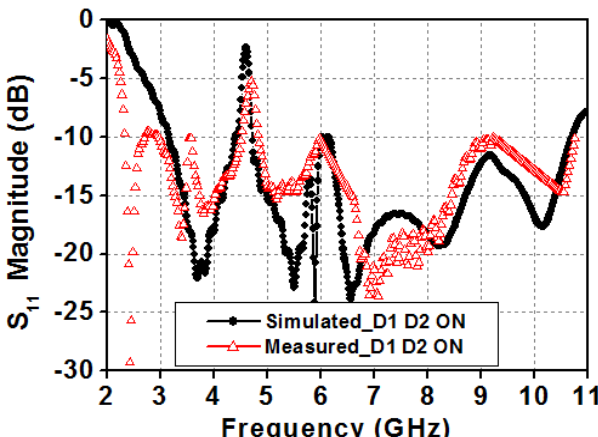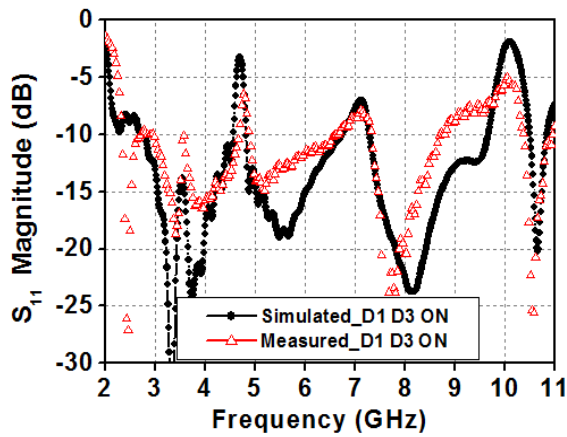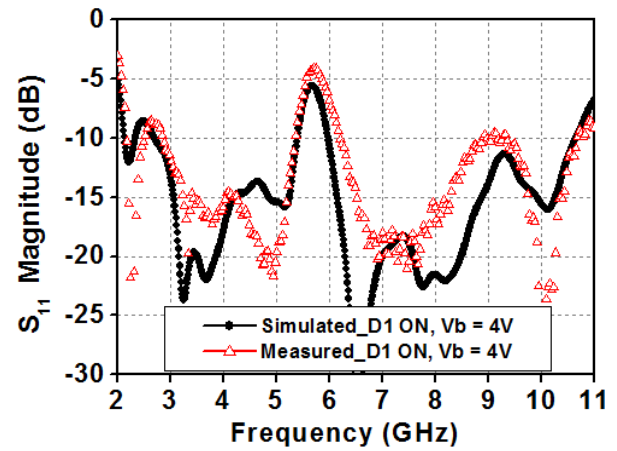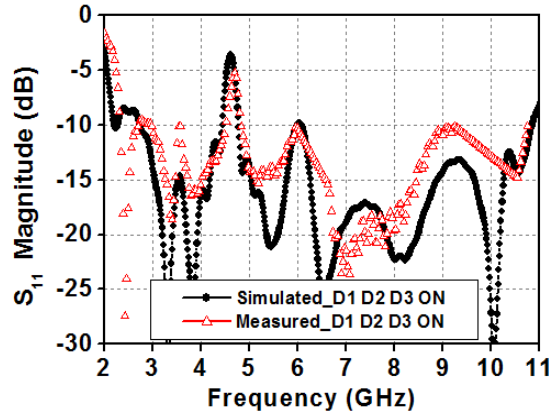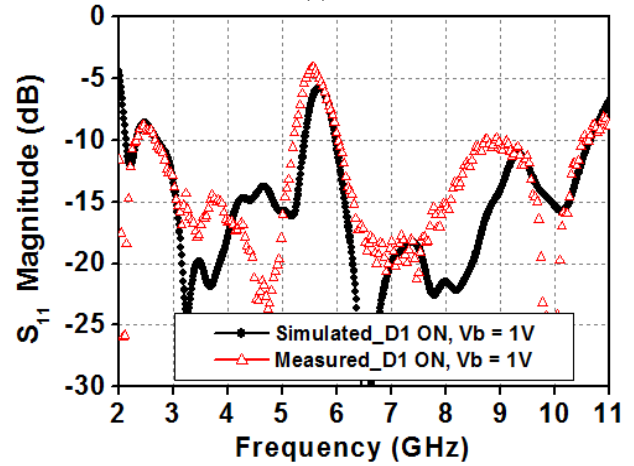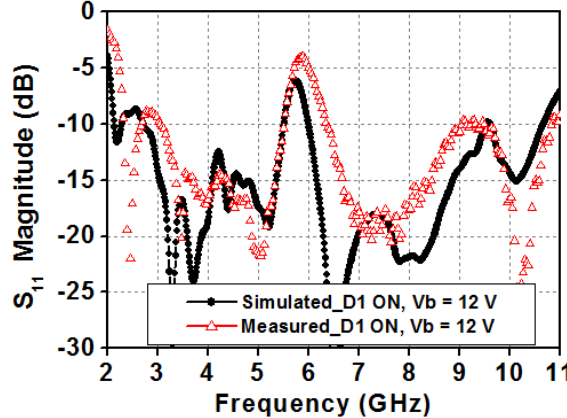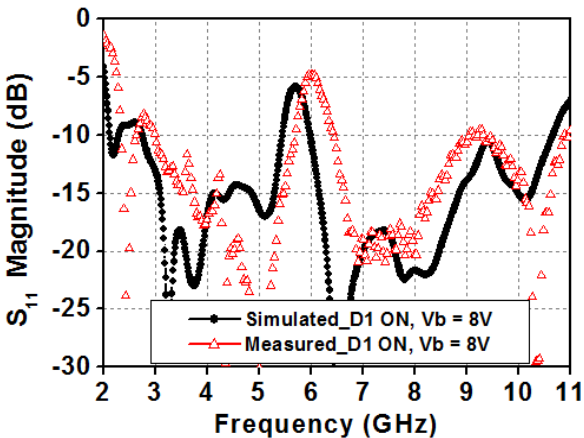
Figure 14. Reflection coefficients of the modified switchable and tunable antenna.

by changing the overall reactance as shown in Figure 14(d). In state 5, diodes D1, D2 and D3 are ON, one notch is obtained at 4.75 GHz as shown in Figure 14(e).

Continuous tuning is also obtained from the structure, diode D4 (varactor diode, model MA46H120) is integrated as shown in Figure 11. Capacitance of the diode varies from 0.1 pF to 1 pF with bias voltage varies from 12 V to 1V, respectively. The results are plotted with different bias voltages as shown in Figure 14(f) to (i). Notch-bands are shifted from higher to lower side with centre frequencies of 6 and 5.5 GHz, respectively.
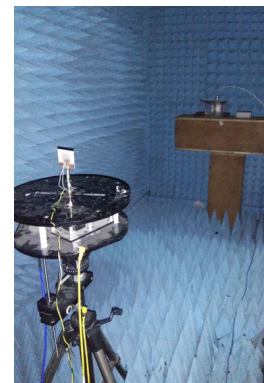
Figure 15. Photograph of the measurement setup.

The radiation patterns of the switchable and tuneable octagonal monopole patch antenna are measured in the anechoic chamber. Photograph of the measurement setup is shown in Figure 15. A double-ridged horn antenna is used to illuminate the test antenna. The proposed (test) antenna is placed on a rotating table which rotates horizontally. Figure 16 shows the normalized radiation patterns. The cross-polarization level is better than 12 dB at all measured frequencies.
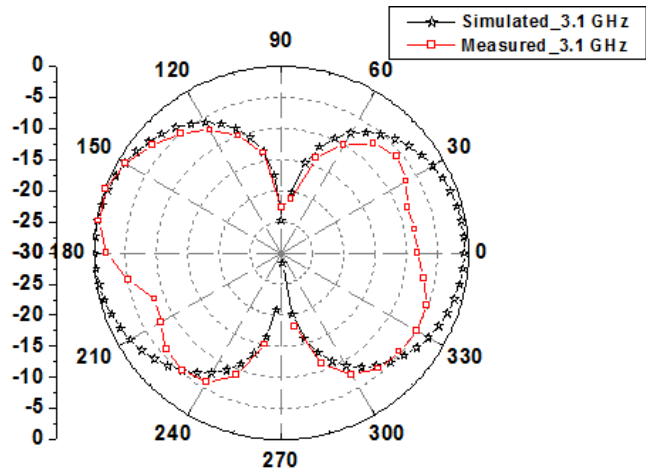
(a) yz - plane
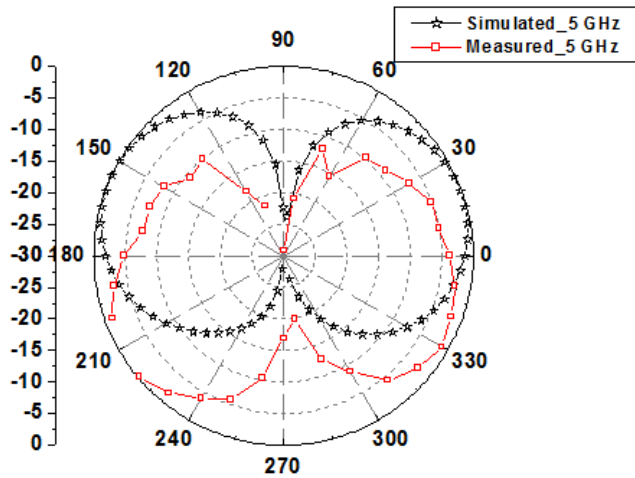
(b) yz - plane
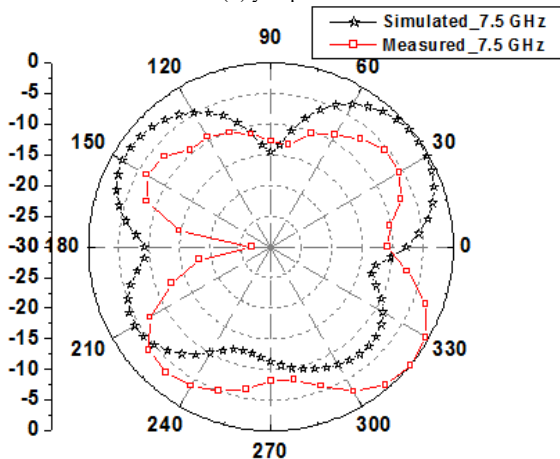
(c) yz - plane
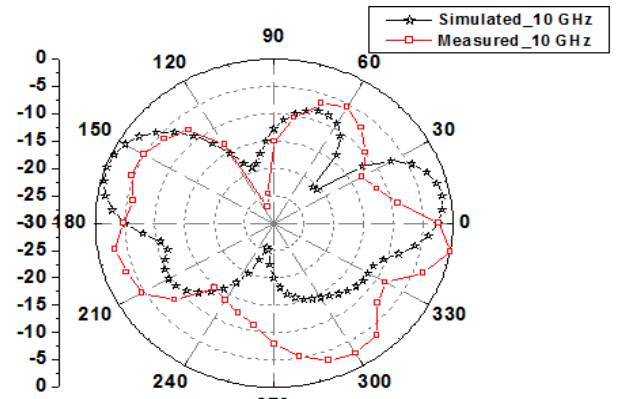
(d) yz - plane
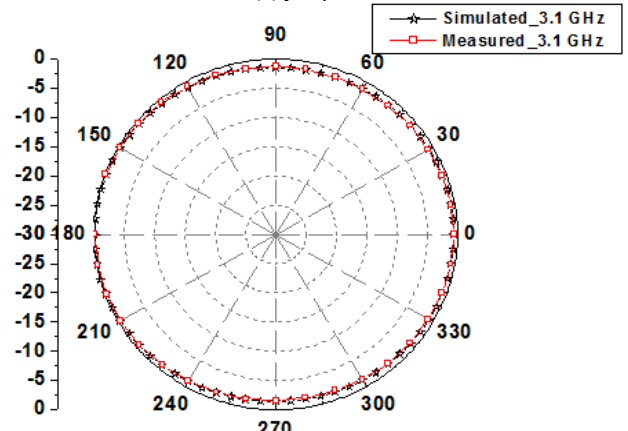
(e) xz - plane
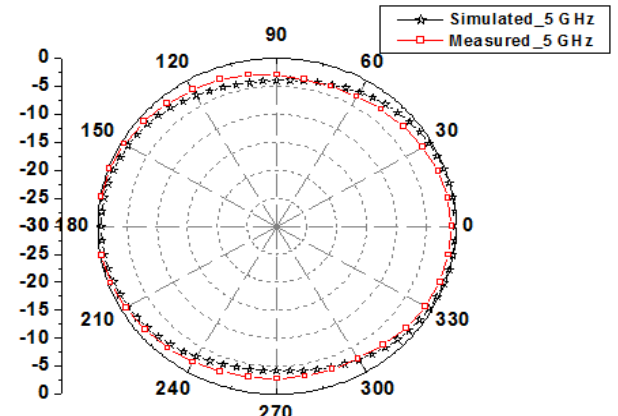
(f) xz - plane
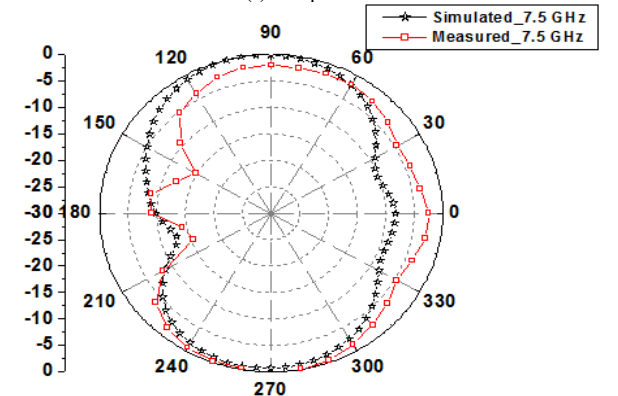
(g) xz - plane

(h) xz - plane
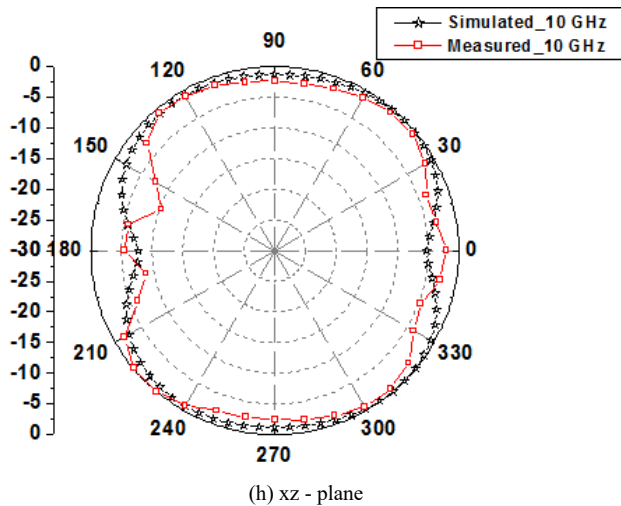
Figure 16. Normalized radiation patterns of the proposed switchable and tunable antenna.

The measured gains are better than 3.12 dB in all states except at the notched-bands. The gains in the notched-bands are below -7.4 dB. The discrepancies in the results are due to fabrication imperfections and use of insulated wires. Insulated wires affect the radiation pattern. Further, the sharp notches can be obtained from this structure by using lumped elements. The filters made from microstrip line have some bandwidth which is more than that of achieved with the lumped elements.

Table 2: Comparisons of the proposed work with earlier reported ones

	Frequency Range (GHz)	No. of notched-bands	Total size L×W (mm²)	Type of switching
[9]	2 – 12.5	Two notched-bands @ 5.3, 7.4 GHz	38.5 × 46	Discrete
[17]	3.1 – 14	One tuning band @ 5.1-5.9	22 × 22	Continuous
[18]	3.1 – 12*	One tuning band @ 3.6-5.2	18 × 21	Continuous
[19]	2.9 – 12	Four notched-bands @ 3.5, 4.6, 5.2, 5.8 GHz	30 × 31	Discrete
[20]	3.1 – 10.7	One notched-band @ 5 GHz	30 × 27.4	Discrete
This work (with varactor diode)	2.8 – 11.5	Five discrete notched-bands @ 4.5, 4.75, 4.85, 8.25, 10 GHz and one tuning band @ 5.5 – 6 GHz	42 × 40	Discrete & Continuous

* Estimated from the data given in the paper

Comparison with earlier reported papers is given in Table 2. In [9], one quasi-complementary split-ring resonator is etched to achieve dual notched frequency bands while in [20], one notched-

band is created by embedding a π-shaped slot in the structure, but the advantage is the compact structure. In [19], four notched-bands are obtained by inserting four U-shaped and one L-shaped slot. Continuous switching was achieved in [17] and [18]. In all these designs except [17], notched-bands were created by perturbing the radiating structure. In this paper, the notched-bands are created by using stubs as band stop filters (BSFs) on the back side of a radiator instead of changing the front side which has a radiator and feed line. In such cases, the radiation patterns are less affected while changing the different states of the antenna [17]. In electromagnetic scattering environment, antenna must have freedom to suppress the selected band. The proposed antenna can suppress the bands with centre frequencies of 4.5, 4.75, 4.85, 8.25, 10 and 5.5-6 GHz.

5. Conclusion

An experimental investigation of reconfigurable UWB modified octagonal microstrip monopole patch antenna with switchable and tunable band-notched characteristic is proposed in this paper. Initially, the design is started with creation of switchable notched-bands, after that some additional notches are obtained by using a modified structure. Finally, by employing a varactor diode, the tuning of notched-bands is achieved. The proposed designs are simple in geometry. The proposed antenna has freedom to suppress many interfering bands. The tunable band-notched reconfigurable UWB antenna is suitable for modern wireless communication systems.

References

[1] T. Kaiser, F. Zheng, and E. Dimitrov, "An overview of ultra-wide-band systems with MIMO," Proc. IEEE, 97(2), 285–312, Feb. 2009.
[2] J. Ren, W. Hu, Y. Z. Yin, and R. Fan, "Compact printed MIMO antenna for UWB applications," IEEE Antennas Wireless Propag. Lett., 13, 1517–1520, 2014.
[3] Wang-Sang Lee, Dong-Zo Kim, Ki-Jin Kim and Jong-Won Yu, "Wideband planar monopole antennas with dual band-notched characteristics," IEEE Trans. Microw. Theory Tech., 54(6), 2800-2806, June 2006.
[4] P. Gao et al., "Compact printed UWB diversity slot antenna with 5.5-GHz band-notched characteristics," IEEE Antennas Wireless Propag. Lett., 13, 376–379, 2014.
[5] J. -M. Lee, K. -B. Kim, H. -K. Ryu, and J. -M. Woo, "A compact ultra wideband MIMO antenna with WLAN band-rejected operation for mobile devices," IEEE Antennas Wireless Propag. Lett., 11, 990–993, 2012.
[6] L. Kang, H. Li, X. Wang and X. Shi, "Compact Offset Microstrip-Fed MIMO Antenna for Band-Notched UWB Applications," IEEE Antennas Wireless Propag. Lett., 14, 1754-1757, 2015.
[7] T. K. Roshna, U. Deepak, V. R. Sajitha, K. Vasudevan and P. Mohanan, "A Compact UWB MIMO Antenna With Reflector to Enhance Isolation," IEEE Trans. Antennas Propag., 63(4), 1873-1877, April 2015.
[8] J.-F. Li, Q.-X. Chu, Z.-H. Li, and X.-X. Xia, "Compact dual band notched UWB MIMO antenna with high isolation," IEEE Trans. Antennas Propag., 61(9), 4759–4766, Sep. 2013.
[9] W. T. Li, Y. Q. Hei, W. Feng and X. W. Shi, "Planar Antenna for 3G/Bluetooth/WiMAX and UWB Applications With Dual Band-Notched Characteristics," IEEE Antennas Wireless Propag. Lett., 11, 61-64, 2012.
[10] S. Wei, Y. Yin, Y. Huang, S. Fan, L. Kang and X. Li, "Novel smiling face-shaped antenna for dual band-notched ultra-wideband applications," 2010 IEEE International Conference on Ultra-Wideband, Nanjing, 2010, 1-3.
[11] He Huang, Ying Liu and Shuxi Gong, "A novel uniplanar differentially-fed UWB polarization diversity antenna with dual notch bands," 2015 9th European Conference on Antennas and Propagation (EuCAP), Lisbon, 2015, 1-4.
[12] D. Yu, G. Xie, Z. Liao and W. Zhai, "Novel dual band-notched omni-directional UWB antenna on double substrates crossing," 2011 4th IEEE International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications, Beijing, 2011, 95-97.

[13] L. Kurra, M. P. Abegaonkar, A. Basu and S. K. Koul, "A band-notched UWB antenna using uni-planar EBG structure," 2013 7th European Conference on Antennas and Propagation (EuCAP), Gothenburg, 2013, 2466-2469.

[14] X. Wang, Z. F. Yao, Z. Cui, L. Luo and S. X. Zhang, "CPW-fed band-notched monopole antenna for UWB applications," 2008 8th International Symposium on Antennas, Propagation and EM Theory, Kunming, 2008, 204-206.

[15] J. Dong, G. Fu, J. Zhao and X. Wang, "Dual band-notched UWB antenna with folded SIRs," 2012 International Conference on Microwave and Millimeter Wave Technology (ICMMT), Shenzhen, 2012, 1-3.

[16] A. K. Tagantsev, V. O. Sherman, K. F. Astafiev, J. Venkatesh, and N. Setter, "Ferroelectric Materials for Microwave Tunable Applications," Journal of Electroceramics, 11, 5-66, 2003.

[17] R. Zaker, C. Ghobadi and J. Nourinia, "Novel Modified UWB Planar Monopole Antenna With Variable Frequency Band-Notch Function," IEEE Antennas Wireless Propag. Lett., 7, 112-114, 2008.

[18] Mahdi Nouri and Sajjad Abazari Aghdam, "reconfigurable uwb antenna with electrically control for triple on-demand rejection bandwidth," Microwave Optical Technology Lett., 57(8), August 2015.

[19] Z. H. Wu, F. Wei, X. W. Shi and W. T. Li, "A compact quad band-notched UWB monopole antenna loaded one lateral L-shaped slot," Progress in Electromagnetics Research, 139, 303–315, 2013.

[20] Y. -L. Zhao, Y. -C. Jiao, G. Zhao, L. Zhang, Y. Song and Z. -B. Wong, "Compact planar monopole UWB antenna with band-notched characteristic," Microwave Optical Technology Lett., 50(10), 2656–2658, Oct. 2008.

[21] Ruchi Paliwal, Rajesh K. Singh and Shiban K. Koul, "Reconfigurable UWB Monopole Antenna with Switchable Frequency Notched Bands," IEEE Applied Electromagnetics Conference, Dec. 2017, 1-4.

[22] CST Studio Suite, [Online]. Available: http://www.cst.com/

[23] Data Sheet of MA4SPS402 PIN diodes, MA-Com, Application Note.

Analysis of Drain Current Transient Response of Gate Pulse Voltage in AlGaN / GaN High Electron Mobility Transistors

Hirohisa Taguchi[*], Kazuto Akahori, Takuma Shimazu, Honoka Tanabe

Department of Electrical and Electronic Engineering, School of Engineering, Chukyo University, Nagoya, 4668666, Japan

A B S T R A C T

Recently, several studies focus on a GaN material system that exhibits a significant probability of use in power devices including wide-gap semiconductors. However, the GaN-HEMT is also a structure that easily leads to crystal defects in AlGaN and i-GaN heterojunction. The aim of the study involved investigating the cause of the current collapse in GaN-HEMT after device construction. A GaN-HEMT with a field plate structure was subject to an environmental temperature change from 300 K to 400 K. A pulse voltage was applied to the gate electrode, and the transient response characteristic of the drain current was analyzed. Given the application of the pulse voltage on the gate electrode, charging and discharging of 2 DEG carriers was repeated with respect to crystal defects near the gate electrode. The charge / discharge reduction was observed via a sampling oscilloscope as a transient response. The transient response exhibited an evident dependence on temperature change. The dependence indicated a time constant change, and thus it was possible to calculate the activation energy of crystal defects trapping carriers. The results suggested that the crystal defect evaluation of GaN-HEMT was possible via transient response analysis of 2DEG carrier by using the proposed method.

1. Introduction

The study is an extension of a previous study originally presented in TENCON 2017 [1]. Recently, several studies focus on GaN material system that exhibits a significant probability of use in power devices with respect to wide-gap semiconductors [2]. The GaN exhibits high breakdown voltage characteristics derived from the wide gap structure, and it is possible to exhibit an output exceeding 1 A in the HEMT structure [3]. Extant studies investigated the application of GaN-HEMTs to green ICT systems as high efficiency devices [4]. The device structure of GaN-HEMT is constructed with a superlattice structure of AlGaN and i-GaN. The superlattice structure exhibits a piezo effect and spontaneous polarization effect and generates two-dimensional electron gas at the interface between AlGaN and i-GaN [5]. Thus, a transistor layer that exhibits a low ON resistance is obtained by constructing a channel layer in GaN that should originally correspond to a high resistance material. Therefore, the GaN-HEMT is a power device exhibiting high withstand voltage, high power operation, and low ON resistance. It is expected that it can be potentially utilized in a wide range of applications including power amplification

mechanism of mobile phone base station, power units of artificial satellites, and sensor applications in server power supply [6].
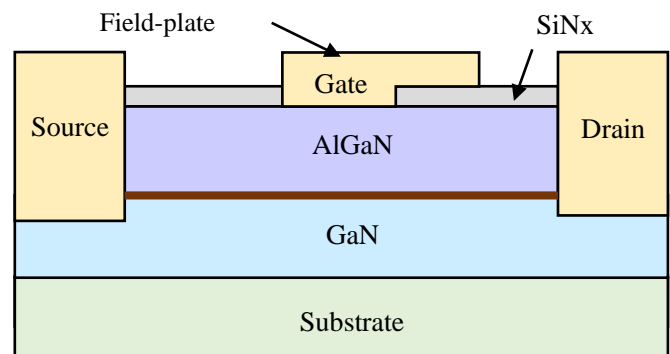
Figure 1 Cross section of the GaN HEMT used in this study

However, the GaN-HEMT is also a structure that easily causes crystal defects in AlGaN and i-GaN heterojunction [7]. A current collapse phenomenon occurs wherein 2DEG carriers are trapped in the crystal defect and the transfer current value sharply decreases with respect to DC response characteristics [8]. Measures including the addition of a surface passivation film

[*]Hirohisa Taguchi, ,   Email: htaguchi68@gmail.com

(SiNx), overhanging the direction of the drain electrode of the gate electrode structure (gate plate), and addition of a recess structure between the gate and drain were adopted with respect to the GaN-HEMT structure to suppress the current collapse phenomenon [9, 10]. However, crystal defects are known to increase due to the energization of the device or extreme environmental temperature change, and this deteriorates device characteristics. Additionally, after device construction, it is difficult to identify crystal defects generated in GaN-HEMT [11].

Therefore, the aim of the study involves investigating the cause of current collapse of GaN-HEMT after device construction. A GaN-HEMT with a field plate structure is subject to an environmental temperature change from 300 K to 400 K. A pulse voltage is applied to the gate electrode, and the transient response characteristic of the drain current was analyzed. Given the pulse voltage applied to the gate electrode, charging and discharging of 2 DEG carriers are repeated with respect to crystal defects near the gate electrode. The charge / discharge reduction is observed with a sampling oscilloscope as a transient response. The transient response exhibits an evident dependence on the temperature change. The dependence indicates a time constant change, and thus it is possible to calculate the activation energy of crystal defects trapping carriers. The results suggest that the crystal defect evaluation of GaN-HEMT is possible via the transient response analysis of a 2DEG carrier by using the proposed method.

2. Experimental Method

2.1. Target Device

Figure 1 shows the cross section of the GaN HEMT used in the study. The GaN HEMT includes an i-GaN layer built on a SiC substrate as a channel layer. As a characteristic structure of the HEMT, it includes a field plate that extends the drain side of the T-shaped gate. This suppresses the current collapse phenomenon during normal temperature and high temperature operation.

2.2. Experimental Set Up

Figure 2 shows the experimental diagram in the experiment. In the experiment, the gate voltage was pulsed, and the change in the drain current is observed via a sampling oscilloscope. The base voltage of the gate pulse is set to 0 V, and the rectangular wave pulse is set at the peak voltage of 2 V. The period is 50 ms, and the duty ratio is 0.5. The drain voltage is variable at 0 to 7 V, and the response characteristic observation is set at 3 V. Pulse response characteristics are observed via a sampling oscilloscope (HP83480A, Max: 50 GHz). In order to observe environmental temperature changes, AlGaN / GaN HEMT (Devise under the Test: DUT) is introduced into a nitrogen furnace. The environmental temperature change ranges from 296 K to 373 K. All measurement systems are high frequency compatible (50 GHz).

3. Experimental Results

Figure 3 shows the I-V characteristics of the GaN-HEMT used in the experiment. The blue line in the figure shows the result when only DC voltage is applied. The orange line is obtained by adding high frequency to the gate electrode portion. In the case of only DC voltage, the linear region characteristic of the FET is

confirmed immediately after the drain voltage is applied. However, when the drain voltage is 0.2 V or more, the current collapse phenomenon appears, and the current value plummets. When a high frequency component is applied to the gate electrode portion, the carrier trapping effect possessed by crystal defects near the gate electrode is suppressed by the field plate effect, and an almost normal I-V characteristic is obtained. The phenomenon was interpreted as the self-heating of GaN-HEMT in previous studies [12].

Figure 4 shows the time axis response of the drain current when a pulse voltage is applied to the gate electrode portion. The time axis response characteristic in nine points in 10K increments from and Figure. 4 (b) shows the enlarged view of the peak portion where temperature dependence is confirmed. As shown in figure 4
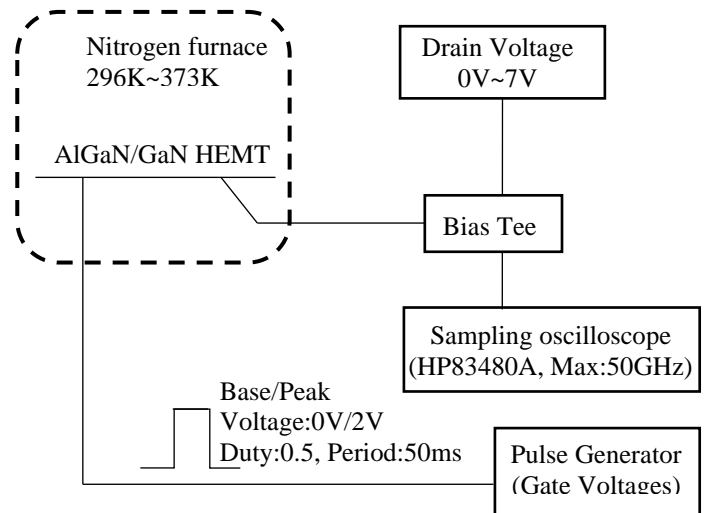
Figure 2 Experimental diagram in this experiment.
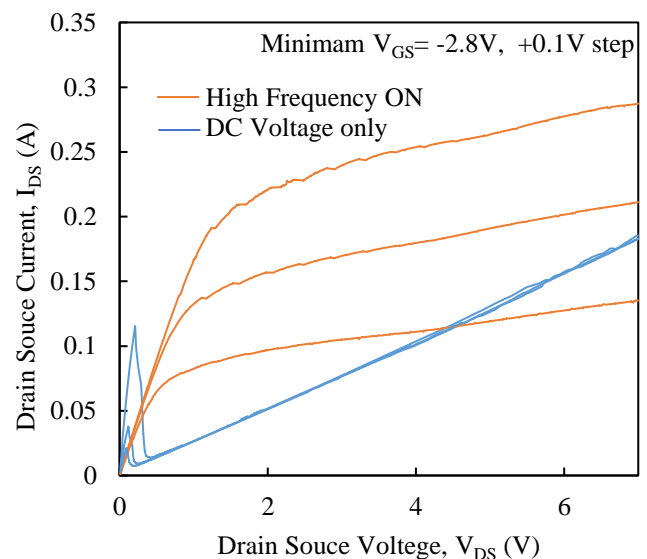
Figure 3: I-V characteristics of the GaN-HEMT.

(a), three types of peak responses are confirmed. Peak 1 in figure 4 (a) in the temperature dependence is not observed. While a gate pulse voltage is applied, a stable 2DEG current exists under the gate, although the 2 DEG current value plummets at the moment
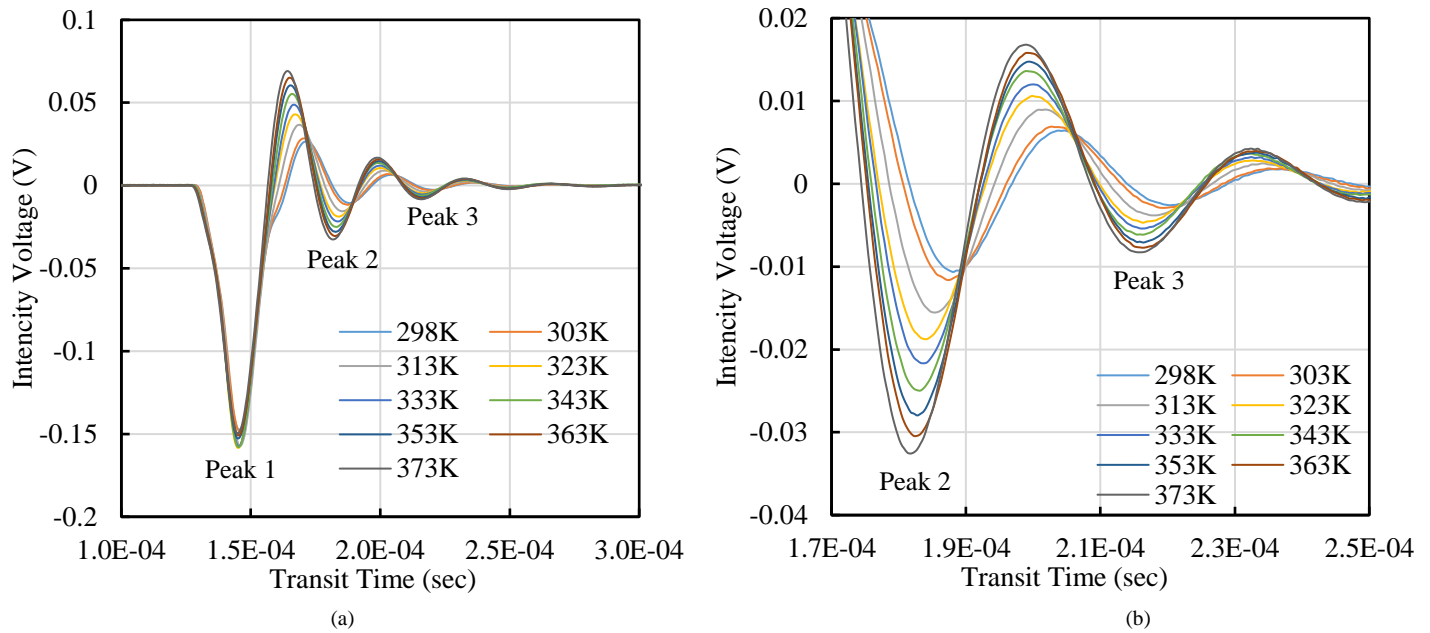
Figure 4 Time-axis response to the drain pulse voltage of the AlGaN/GaN HEMT.
(a) Overall view (b) Enlarged view of peak 2 and peak 3

when the pulse voltage switches to the OFF-State. The peak 1 does not exhibit temperature dependency, and it is considered as a response characteristic when the 2DEG carrier is directly swept from under the gate.296 K is observed by applying a temperature change of up to 373 K. Figure. 4 (a) shows the overall view of the time axis response,

However, peak 2 and peak 3 exhibit an evident temperature dependence as shown in Figure. 4 (b). The temperature dependence in this case exists in terms of both the peak height and peak shape. The peak height indicates the temperature dependence of the number of carriers based on the peak. The peak shape indicates the carrier sweeping process. When the horizontal axis is considered on the time axis, it is interpreted as a transient response. Thus, the shape change indicates a time constant change that limits the transient response. This is discussed in the next section.

4. Discussions

Peak 2 exhibited an evident temperature dependence. The peak height of Peak 2 decreased with increases in the temperature. The peak height was the response voltage and is dependent on the number of carriers. It was considered that the increases in temperature strengthen the piezo effect and increase the thermal carrier supplied from the semiconductor hetero interface. Both the 2 DEG carrier supplied by the voltage and the carrier supplied by the heat were trapped in crystal defects in the AlGaN. The trapping probability was based on the number of carriers wherein thermal carriers increase with increases in the temperature and the number of carriers accumulated in crystal defects also increases [13]. It was considered that the captured carrier is emitted from crystal defects by turning the gate pulse voltage as OFF-State. The emission phenomenon was interpreted as a peak response shown in Figure 4, and transient response characteristics were analyzed.

The response time constant was calculated from the transient response characteristic. As shown in Figure 5, the response time constant decreased with increases in the temperature. Based on the temperature dependence, an Arrhenius plot was executed, and the activation energy ΔE of transient response characteristics was calculated from the slope of the plot. Additionally, ΔE was calculated as approximately 0.575 eV. The value was extremely close to the activation energy of crystal defects in reference [13]. Thus, the transient response indicated by the peak 2 was interpreted as a carrier discharge phenomenon trapped in the crystal defect of the AlGaN layer under the gate.

A pulse voltage was applied to the gate electrode, and thus it was possible to repeat the process of generating / extinguishing the 2 DEG directly under the gate electrode. When only 2DEG carrier generating / extinguished (as shown in Figure 4), peak 2 and peak 3 based on the temperature were not confirmed. When the pulse voltage corresponded to the ON-State at the gate electrode, the 2 DEG carrier was formed under the gate. At the same time, the 2DEG carrier was trapped by crystal defects present in the AlGaN layer near the gate electrode. When the gate pulse voltage turned OFF-State, the 2 DEG carriers were first swept, and the peak 1 was confirmed. There was no reason to observe the temperature dependence in this case. Subsequently, the carriers trapped in crystal defects were gradually released, and peak 2 and peak 3 appeared as transient responses.

We attempted the same analysis for peak 3 although the number of trap carriers was not sufficiently acquired, and the peak intensity was low. An evident temperature dependence was confirmed as shown in the shape in Figure 4. However, an effective analysis result was not obtained. Evidently, the response time of peak 3 was slower than that of peak 2. Additionally, the peak height was lower than that of peak 2. Given that the carrier trap to the crystal defect formed peak 3, a period of time elapsed

in releasing the carrier, and it was considered as the characteristic of the crystal defect in the deep part of the crystal. The possibility was considered wherein the presence of crystal defects in the i-GaN crystal depth builds a channel layer [14].
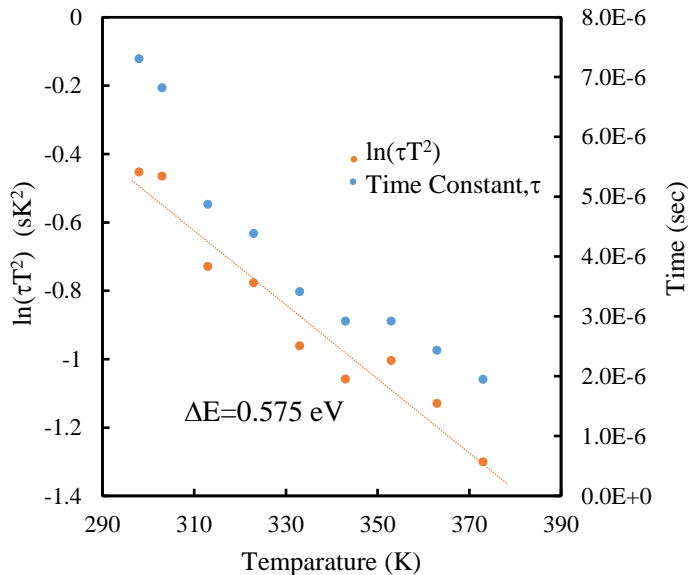
Figure 5 Arrhenius plot using the time constant.

In the study, a pulse voltage was applied to the gate electrode of the completed GaN-HEMT, and the transient response change of the drain current was precisely observed and analyzed. This was a simple method. However, it enabled the evaluation of crystal defects in the completed GaN-HEMT. Although the ICTS method and the DLTS method were used for the same analysis, the method was an extremely simple method when compared with the aforementioned methods and exhibited excellent versatility.

5. Conclusions

The aim of the study involved investigating the cause of current collapse of GaN-HEMT after device construction. A change in the environmental temperature from 296 K to 373 K was applied to GaN-HEMT with a field plate structure. A pulse voltage was applied to the gate electrode, and the transient response characteristic of the drain current was analyzed. Given the application of the pulse voltage to the gate electrode, charging and discharging of 2 DEG carriers was repeated with respect to crystal defects near the gate electrode. The charge / discharge reduction was observed with a sampling oscilloscope as a transient response. The transient response exhibited an evident dependence on temperature change. The response time constant is calculated from the transient response characteristic. As shown in Figure 5, the response time constant decreased with increases in the temperature. Based on the temperature dependence, an Arrhenius plot was executed, and the activation energy ΔE of transient response characteristics was calculated from the slope of the plot. Additionally, ΔE was calculated as approximately 0.575 eV. The value was extremely close to the activation energy of crystal defects in extant studies. The dependence indicated a time constant change, thereby facilitating the calculation of the activation energy of crystal defects trapping carriers. It was suggested that the crystal defect evaluation of GaN-HEMT is possible via transient response analysis of 2DEG carrier by the method.

References

[1] A. Otake, S. Nakano, K. Yamanaka, H. Taguchi, "Temperature dependence of drain current transient response by gate pulse voltage in AlGaN / GaN high electron mobility transistors" TENCON 2017 IEEE Region 10 Conference, Penang, Malaysia, 2017. https://doi.org/10.1109/TENCON.2017.8228128

[2] S. Nakamura, T. Mukai, M. Senoh, "Candela-class high-brightness InGaN/AlGaN double-heterostructure blue light-emitting diodes" Appl. Phys. Lett. 64(13), 1687–1689, 1994. https://doi.org/10.1063/1.111832

[3] G. Nishio, K. Nakatani, T. Ishizaki, "Design of harmonic processing circuit for microwave GaN-HEMT power amplifier", IEEE International Meeting for Future of Electron Devices, Kansai (IMFEDK), Kyoto, Japan, 2015. https://doi.org/10.1109/IMFEDK.2015.7158583

[4] R. Quay, D. Schwantuschke, E. Ture, F. van Raay, C. Friesicke, S. Krause, S. Müller, S. Breuer, B. Godejohann, P. Brückner, "High-power microwave GaN/AlGaN HEMTs and MMICs on SiC and silicon substrates for modern radio communication" Phys. Status Solidi A, 215(9), 1700655(1–7), 2018. https://doi.org/10.1002/pssa.201700655

[5] K. Shinohara, C. King, A. Carter, E. Regan, A. Arias, J. Bergman, M. Urteaga, B. Brar "GaN-based field-effect transistors with laterally gated two-dimensional electron gas" IEEE Electron Device Letters, 39(3), 417–420, 2018. https://doi.org/10.1109/LED.2018.2797940.

[6] H. Watanabe, T. Fukami, H. Saito, A. Tomiki, O. Ceylan, H. Nunomura, O. Shigeta, T. Shinke, K. Kojima, "High speed downlink system for small satellite and high-efficiency x-band GaN SSPA" IEEE MTT-S International Microwave Symposium (IMS2014), Tampa, FL, USA, 2014. https://doi.org/10.1109/ MWSYM.2014.6848670

[7] H. Sasaki, T. Hisaka, K. Kadoiwa, T. Oku, S. Onoda, T. Ohshima, E. Taguchi, H. Yasuda, "Ultra-high voltage electron microscopy investigation of irradiation induced displacement defects on AlGaN/GaN HEMTs" Microelectronics Reliability, 81, 312–319, 2018. https://doi.org/10.1016/j.microrel.2017.10.005.

[8] H. Huang, Y. Liang, G. Samudra, T. Chang, C. Huang, "Effects of gate field plates on the surface state related current collapse in AlGaN/GaN HEMTs" IEEE Transactions on Power Electronics, 29(5), 2164–2173, 2013. https://doi.org/ 10.1109/TPEL.2013.2288644.

[9] W. Waller, M. Gajda, S. Pandey, J. Donkers, D. Calton, J. Croon, S. Karboyan, J. Šonský, M. Uren, M. Kuball, "Impact of silicon nitride stoichiometry on the effectiveness of AlGaN/GaN HEMT field plates" IEEE Transactions on Electron Devices, 64(3), 2017. https://doi.org/10.1109/TED. 2017.2654800

[10] D. Čučak, M. Vasić, O. García, J. Oliver, P. Alou, J. Cobos, A. Wang, S. Martín-Horcajo, M. Romero, F. Calle, "Physics-based analytical model for input, output, and reverse capacitance of a GaN HEMT with the field-plate structure" IEEE Transactions on Power Electronics, 32(3), 2017. https://doi.org/10.1109/TPEL.2016.2569404

[11] S. Huang, Q. Jiang, S. Yang, C. Zhou, K. Chen, "Effective passivation of AlGaN/GaN HEMTs by ALD-grown AlN thin film", IEEE Electron Device Letters, 33(4), 516–518, 2012. https://doi.org/10.1109/LED.2012.2185921.

[12] C. Wang, Y. Xu, X. Yu, C. Ren, Z. Wang, H. Lu, T. Chen, B. Zhang, R. Xu, "An electrothermal model for empirical large- signal modeling of AlGaN/GaN HEMTs including self-heating and ambient temperature effects" IEEE Transactions on Microwave Theory and Techniques, 62(12), 2014. https://doi.org/10.1109/TMTT.2014.2364821.

[13] M. Mielenz, J. Brox, S. Kahra, G. Leschhorn, M. Albert, T. Schaetz, H. Landa, B. Reznik, "Trapping of topological-structural defects in Coulomb crystals" Phys. Rev. Lett. 110, 2013. https://doi.org/10.1103/PhysRevLett.110.133004.

[14] J. Joh, J. Alamo, "A current-transient methodology for trap analysis for GaN high electron mobility transistors", IEEE Trans. On Electron Devices, 58(1), 132–140, 2011. https://doi.org/10.1109/TED.2010.2087339.

Development and Testing of Intelligent Wheelchair Controller for Quadriplegic Patients

Mohammed Faeik Ruzaij Al-Okby [*,1], Sebastian Neubert [2], Norbert Stoll [2], Kerstin Thurow [3]

[1]*Technical Institute of Babylon, Al-Furat Al-Awsat Technical University (ATU), Babylon, Iraq*

[2]*Institute of Automation, University of Rostock, Rostock 18119, Germany*

[3]*Center for Life Science Automation (celisca), University of Rostock, Rostock 18119, Germany*

A R T I C L E   I N F O

A B S T R A C T

In this research paper, the development and evaluating of a smart controller for electrical powered wheelchairs are presented. The controller aimed to assist quadriplegic, paralyzed, and handicap patients who cannot use their hands to drive an electrical wheelchair by using a joystick controller. Two sub control units have been combined in one hybrid system to create the current version which are voice, and head tilt controllers. They are activated and operated in parallel at the same time to allow the wheelchair user to choose the preferred control method. The voice controller is activated by the user's voice command to control the wheelchair instead of a joystick. The head tilt controller uses the user's head motions to create control commands instead of the joystick controller. The head tilt controller design is based on using two embedded MEMS orientation detection modules as input measurement units. The system uses a modern low power consumption microcontroller to analyze the received information and data from inputs and creating the prompt control commands and send it to the wheelchair motors driver as an output unit.

1. Introduction

The use of assistive and rehabilitation applications has increased rapidly in recent years because of the huge increase in the numbers of handicapped and paralyzed people. In the traditional wheelchair, joysticks controller has been used to generate the control commands, which requires good control of the user's hand. However, quadriplegic and paralyzed patients cannot use a conventional joystick controller. Thus, several new technologies, methods, and inventions have to be developed to serve this class of patients and to reduce their sufferance. Nowadays, several input signals from the user's body can be used such as body gesture and motions [1-3], brain electrical signals [4-6], user voice commands [7-9], and electrical activity in the body muscles (EMG) [10-11]. Several approaches have been used for each signal. Each approach has its advantages and disadvantages, more details and information are available in [12-13]

In this paper, the design, and evaluation of the hybrid controller for electrical wheelchairs are explained. The controller design takes into consideration the target users' requirements. The target users lost the ability to use their upper and lower limbs and

cannot handle a conventional joystick controller. The novelty of the proposed system is based on using more than one controller flexibly to allow the user choosing the preferred controller. The novel implementation of the voice controller allows the system to avoid false positive (FP) errors by using two different voice recognition modules combined by a false positive errors cancelation algorithm. The head tilt controller has a unique design to perform several tasks based on using many functions to improve the performance of the system in different situations such as applied emergency stop or driving the system in ramps and non-straight roads. The system has been tested with the implementation of several tasks, some important factors such as response time and stop distance have been measured to check the new system controller's performance in comparison to an original joystick controller. The combination of several sensors and detection units in the voice and head tilt controllers produce a durable and simple to use system that helps the target users to enhance the quality of daily life. The proposed work is an improvement and a development of previous work that can be found in [14-15].

2. Methodology

There are three main layers of the system which are the input layer, the main processor, and the execution platforms. The system main layers have been explained in Fig. 1. The input layer

[*]Mohammed Faeik Ruzaij Al-Okby, Technical Institute of Babylon, Al-Furat Al-Awsat Technical University (ATU), Babylon, Iraq. Phone: +9647722365739. Email: Mohammed.al-Okby@celisca.de, abnalfratain@yahoo.com.

includes all the sensors and modules that are responsible for picking up the control signals from the user side and sending it to the main processing unit. The main processor analyses and processes the data and information of the sensors and modules in the input layer. It is responsible for the generation and sending of the control commands to the output layer which is also called the execution platforms. The following section explains each layer in detail:
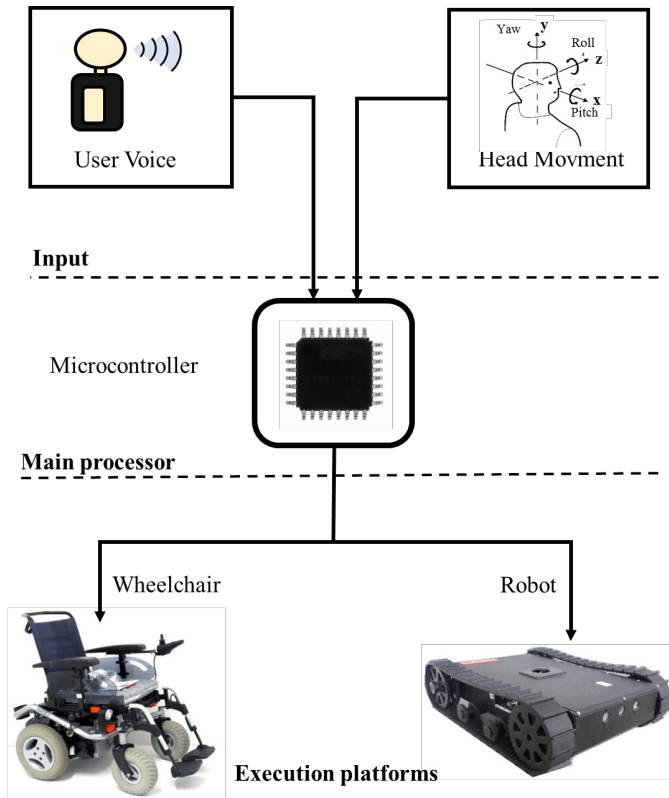
Figure 1: System structure.

2.1. Input Layer

The input layer deals with the signals and parameters picked up from the user body to generate specific control commands. It has two main signals from the wheelchair users which are the user voice command and the user's head motions. There are two units for signal acquisition which are the voice recognition (VR) unit and the orientation detection (OD) unit. These two units are the main core of the voice and head tilt controllers.

The VR unit is a combination of two sophisticated modules using different voice recognition processors and algorithms. This makes the voice controller more reliable and allows the system to detect and eliminate the FP errors. The two voice recognition modules work together in parallel and send the VR results to the microcontroller as command index value using a universal asynchronous receiver/transmitter (UART) bus or by sending a binary value for a specific input/output port pin.

The first voice recognition module used in the voice controller is the Easy VR module which uses the strong voice processor

RSC-4128 from Sensory, USA. The Easy VR module has the ability to operate in Speaker Dependent SD as well as Speaker Independent SI modes. The SD means that the system detects the voice command depending on the individual characteristic of the voice command. This makes the system respond only to the person who was trained before to the system. SI mode means, the system responses to any user giving the selected voice commands. For SD mode, the module uses the well-known Dynamic Time Warping (DTW) algorithm. The module uses the hidden Markov model (HMM) for the SI mode. The module can be trained with up to 32 voice commands.

The second module is the SpeakUP Click^{TM} module (MikroElektronika, Serbia). It only uses the DTW algorithm and is operating in SD mode only. This module can work standalone or with the host processor. It can be trained with up to 200 voice commands. More details regarding the used VR module can be found in [16-17].

The second acquisition unit in the input layer consists of two modern orientation detection sensors from Bosch sensor tech. Inc., Germany called BNO055. Both modules are identical and operate at the same bus speed and settings. They use I^{2}C bus to communicate with the host processor; each module uses its own I^{2}C address. The two modules are fixed in different locations on the system depending on the required orientation measurement. The first module is used to generate the motion control command. A traditional PC headset has been used to fix the first module in the middle of the headset. The second BNO055 module is used to detect the road slope or wheelchair reference orientation. It is fixed on the wheelchair chassis. The internal structure of BNO055 module includes 4 MEMS sensors combined with an ARM 0 microcontroller. The MEMS sensors are a gyroscope, magnetometer, temperature sensor, and accelerometer. A data fusion algorithm has been used to analyze the data from the four sensors and feeds the host processor by a ready to use orientation information in different measurement styles such as quaternion vector, linear acceleration, rotation vector, and Euler angles. In current work, the angle degree measurements of the Euler angles Pitch, Roll, and Yaw have been selected as input to the processing unit to represent the user's head and reference orientation of the wheelchair [18].

2.2. Main processing Layer

The main processing layer interface between the input and output layers. The received information and data from the input unit are transported to the main processing unit using several communication busses and ports. The received data represent the user commands and is processed by the main microcontroller using a specific algorithm depending on the type of sensor sending the information. The results from the microcontroller will be converted to a specific hex-decimal message to control the wheelchair components. The Silicon labs EFM32GG990F1024 microcontroller has been used as the main controller of the system. Several sufficient input/output ports with wide types of

communication buses have been embedded such as UART, I²C, USB, and general-purpose input-output pins GPIO [19].

One of the important tasks of this layer is to prevent the conflict between the modules and sensors in the input layer. The head tilt controller is activated as main controller when the system starts. The voice controller is activated for complementary functions such as controlling speed level and turning on and off lights and signals. The user can select the voice controller as main motion controller by giving the voice command "voice" and he can return to the head tilt as main controller using the voice command "orientation". Fig. 2 shows the system flowchart and the transition between the sub-controllers.

Figure 3: Sabertooth 2x32 Amp motor driver

3. System Description

In the presented work, the two sub-controllers are activated together to cover all the required wheelchair tasks. The head tilts controller cannot cover all the control commands of the wheelchair. It is mainly used for motion and speed adjustment control commands and not covers other commands like turning on and off lights or sounding alarms and signals. The voice controller was tested at different noise levels for several voice commands in [9], [14]. In this current work, it is used as an assistive controller with a few commands for controlling additional parameter.

The head tilt controller has two main orientation measurements, the first is the orientation of the user's head which is responsible for generating the motion control commands. The generation of the motion commands depends on the value and the direction (the axis) of the head tilt angle. The second orientation measurement is for the reference wheelchair orientation. This measurement is used to calibrate the system in case of passing non-straight roads. The output of the orientation modules has been set in the form of Euler angles. Three Euler angles have been used to represent the body orientation around the main axes x,y,z using the tilt angles pitch (P), roll (R), and yaw (Y) respectively. The modules send the orientation data via I²C bus to the main microcontroller, which receives and processes the head and reference orientation information. The motion control commands are generated by specific motions of the user's head. The speed and direction of the control command are determined based on the tilt axis for direction and the angle degree value for the speed. There are five main motion commands for controlling the presented system. Each command has specific regions of tilt angles to control the wheelchair motion. The forward command starts and ends by tilting the user head between (0°<P<25°) and (5°<R<45° and -45°<R<-5°). The left and right commands start from 0 to maximum speed by tilting the user head between (-10°>P>0°) and (5°<R<45° and -45°<R<-5°). The stop region is located between (-10°>P>0°) and (-5°>R>5°). The backward command tilt angles are located between (-10°>P>-45°) and (5°<R<45° and -45°<R<-5°). All other tilt angles are not used for generating control commands. This region is used for activating error messages in the system display or for activating emergency stop functions to protect the user and the system in specific cases such as the headset fall down or unexpected collision. Fig. 4

Figure 2: System flowchart

2.3. Output Layer

The output layer is responsible for the execution of the control commands. It is implementing motion commands for controlling the wheelchair electrical motors by changing the motors rotation direction and speed. In the presented system, the electrical powered wheelchair Meyra Smart 9.906 (MEYRA GmbH, Germany) is used as output or execution layer. The robotic motor driver Sabertooth 2x32 Amp (Dimension Engineering LLC., USA, see Fig. no. 3) has been used to control the wheelchair motors instead of original VR2 motor driver unit. The same motor driver unit has been used in jaguar lite robot and H20 robot which enables to use the system in robotic applications, too.

shows the programmed head tilt angles for the roll and pitch Euler angles.

a) Pitch (P)

b) Roll (R)
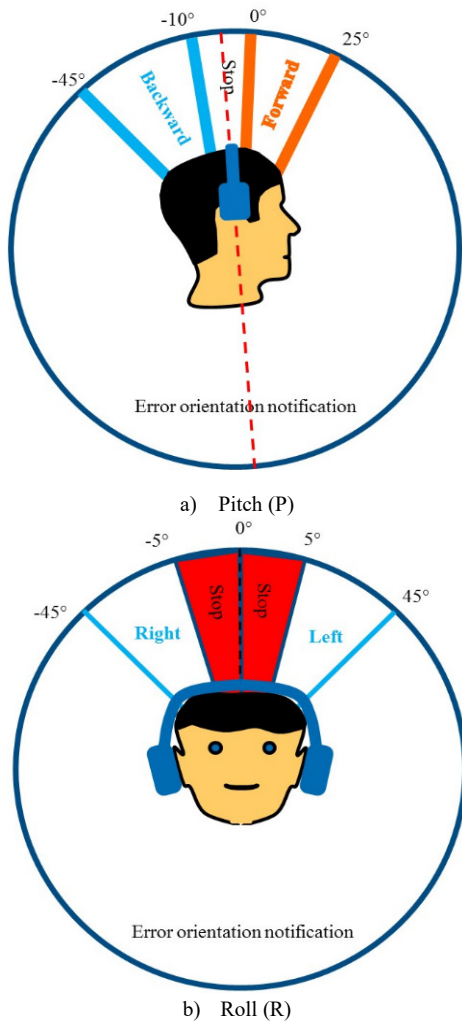
Figure 4: Control commands regions

4. Experimental Results

4.1. System Indoor Test

Several tasks have been tested to check the performance of the hybrid system. The tasks are divided into an individual task for each controller and for both controllers together. A group of complementary voice command has been tested such as "faster = increase the tilt to speed ratio", "slower = set the tilt to speed to the basic ratio", "light off = turn lights off", "light on = turn lights on", and "stop". The head tilt controller was tested with the motion commands. 14 users tested the system with several tasks. All the 14 users are normal user without disabilities. The users used the system as quadriplegics patient without using their upper and lower limbs. Each task was repeated 10 times by all users. The tested tasks included the most important functions of the system. The tasks have been performed in a known path including a narrow corridor and a furnished laboratory. The test had two parts. In the first part, the wheelchair should be driven in a 1.5 m wide narrow corridor. In this part, the user should give the voice command "faster" to change the head tilt controller sensitivity

from the primary speed 0°-25° P tilt angle → 0-2 km/h to the medium speed 0°-25° P tilt angle → 4 km/h (first task T1). In the second task, the user gives the voice command "light on" to activate the system lights T2. When the user reaches the laboratory door, he reduces the wheelchair speed sensitivity to the basic level using the voice command "slower" T3. Inside the laboratory, the user executes another group of tasks. The user needs to pass the laboratory entrance and avoid the collision with some obstacles (see Fig no. 5) T4. After driving the system and passing narrow corridor, the user needs to activate the emergency stop function of the head tilt controller by tilting his head beyond the programmed ranges for pitch or roll Euler angles T5. Then, the user needs to drive the wheelchair to the second laboratory door for exit. Before exiting, the user deactivates the system lights using the voice command "light off" T6. Finally, the user should exit the laboratory and stop the wheelchair in the outer corridor using "stop" voice command T7. The navigation map of the system tests is shown in Fig. 5.
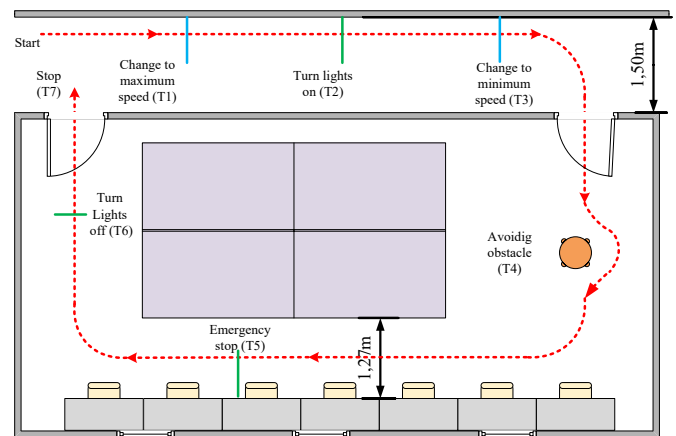
Figure. 5: System tests map

Table 1 includes the system indoor test results in the form of correct command execution (accuracy), the number of false positive error FP, and the number of false negative error FN for every task. The results clearly show that the head tilt controller performance is better than voice controller with 0 % errors and 100 % correct commands accuracy.

Table 1: Indoor test results in the form of correct command execution

Task	Accuracy	FP	FN
T1	95.71 %	0	6
T2	96.42 %	0	5
T3	95.71 %	0	6
T4*	100 %	0	0
T5*	100 %	0	0
T6	97.85 %	0	3
T7	96.42%	0	5

* Head tilt commands

4.2. Reaction time test

The reaction time in this test refers to the time period between giving the command by the user and the complete execution of

the command by the system in real time. Two parameters have been measured in this test which are the time and the distance. The three controllers, head, voice, and joystick were tested in the same environment with identical test parameters and conditions such as user weight, wheelchair speed, and the surrounding noise in the test environment. A fixed ≈ 1.716 km/h wheelchair speed was selected for the test for all three controllers. The wheel rotation per minute (RPM) was calculated and adjusted to 27.6 RPM individually for each controller using the digital tachometer DT-2234C. Only one user with a weight of app. 85 kg performed the test for the three controllers. All tests have been taken at the same laboratory with a surrounding noise of ≈ 54 dB.

The reaction time and the distance to stop the system are calculated using video frames from a high-speed camera. The camera can record the test video in a range from 30 to 1000 frames per second. The measured time period starts when the user gives the stop control command and finishes when the wheelchair stopped completely. The test has been repeated 10 times for each controller and the average reaction times and distances have been calculated. The same steps repeated identically for the head tilt, joystick, and voice controllers. The "stop" command was given in different ways depending on the used controller. The joystick controller has a specific joystick range for stop command which can be easily recorded by the camera. The voice controller will stop the system when the user gives the voice command "stop" and to make it easy to be recorded by the camera the user was asked to lift his hand up when he gives the "stop" command. Finally, for the head tilt controller, the camera can easily record the stop command which is activated by returning the user's head to the range (-10°>P>0°, -5°<R<5°). Fig. 6 illustrates the capturing of the stop command frame for all controllers.

a) Head tilt controller

b) Joystick controller
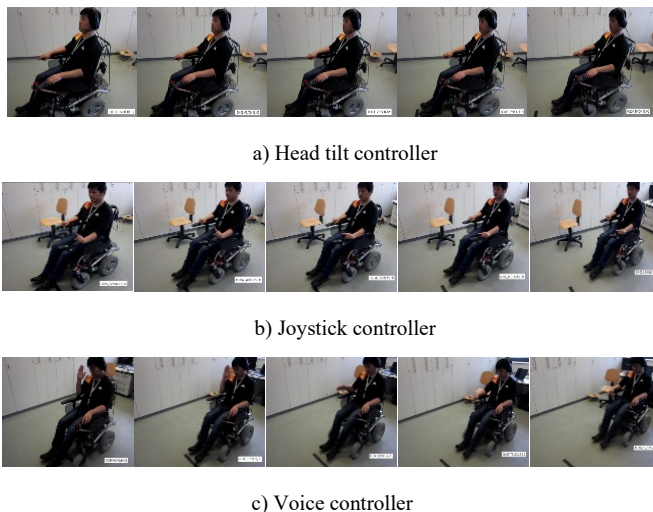
c) Voice controller

Figure 6: Stop command frames of the different controllers

Fig. 7 explains the statistical chart of the calculated reaction time and stop distance of the three controllers. The charts show the mean values for the measured parameter with the standard deviation (STDEV) for 10 samples for each controller. The results showed that the head tilts controller performance is better

compared to the other controllers for the two calculated parameters.
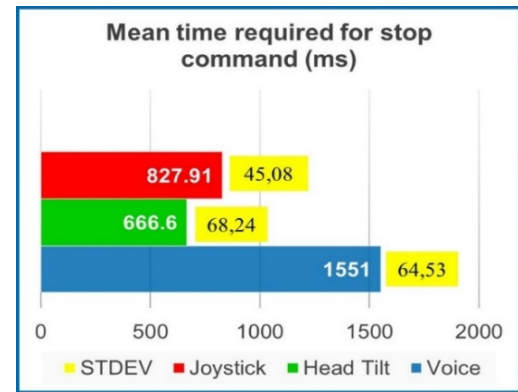
a)   Reaction time
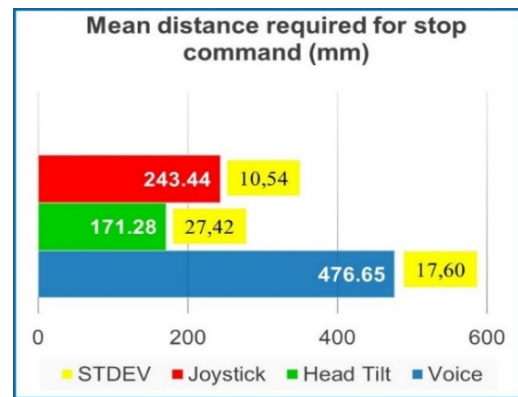
b)   Stop distance

Figure. 7: Reaction time and distance for the stop command

4.3. Questionnaire

For the evaluation of the system usability, the fourteen users in the previous tests were asked seven questions about their feeling, adaptation, and opinions about the system. Each question should get a numerical evaluation from 0 minima up to 10 maximums. Table 1 summarizes the evaluation questions. Fig. 8 and Table 2 explain the questionnaire results.
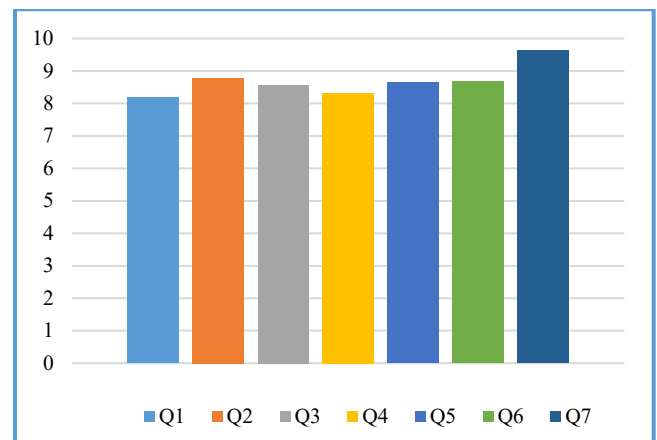
Figure 8: Questionnaire result for the system

Table 2: Questionnaire evaluation questions

Question	Question
Q1	Is it easy to use head tilt controller?
Q2	How do you find the comfortability of wearing a headset?
Q3	How do you find the reaction time for head tilt controller?
Q4	How do you find the reaction time for voice controller?
Q5	How do you find the voice command response?
Q6	Is it easy to pass narrow area?
Q7	Is it easy to control lights and signals?

Table 3: Questionnaire results

User	Q1	Q2	Q3	Q4	Q5	Q6	Q7
U1	7,5	8	7,5	7,5	9	10	10
U2	7,5	8	7	9	7,5	10	10
U3	9	10	10	8	10	7,5	10
U4	8,5	9	8	7,5	9	9	10
U5	7	8	7,5	7	7,5	8	9
U6	7,5	9	10	9	10	8	10
U7	9	8	10	10	8	10	10
U8	9	8	9	10	8	8	10
U9	8	9	10	9	10	8	10
U10	9	10	10	8	10	8	10
U11	8,5	9	8	7,5	9	9	10
U12	8	10	8	8	8	9	9
U13	9	9	8	8,5	7	10	8
U14	7,5	8	7	8	8	7,5	9
Mean	8,214	8,785	8,571	8,3573	8,642	8,714	9,642
SDV	0,726	0,801	1,206	0,928	1,063	0,974	0,633

The abbreviation STDEV refers to the standard deviation. The results of the questionnaire show differences in the system evaluation between the users. The highest mean value with the minimum standard deviation was observed for question Q7 ("Is it easy to control lights and signals?"). The highest difference between the users' answers with a standard deviation of 1.206 was observed for Q3 ("How do you find the reaction time for head tilts controller?"). This may result from the fact, that the users were previously informed about the command confirmation function of the head tilts controller that makes a 100 ms delay in the starting of each motion command. The abbreviation STDEV refers to the standard deviation.

5. Conclusions

In this paper, the development and testing of a hybrid wheelchair controller have been described. The system consists of two sub-controllers which are the voice and head tilts controller. The voice controller uses two voice recognition modules combined to enhance the recognition accuracy and reduce the errors. The head tilts controller tests revealed a good performance in indoor as well as outdoor tests. The tests show that the head tilts controller is easier and more accurate in wheelchair controlling than the voice controller and it is thus selected to be the main controller. The combination of voice and head tilt controllers make the system more flexible and covers all wheelchair functions such as turning lights, sound alarms, and signals on and off. It also enables the selection of the speed range.

Acknowledgment

References

[1] Henrik Vie Christensen A and Juan Carlos Garcia B,"Infrared Non-Contact Head Sensor, for Control of Wheelchair Movements," Book Title Book Editors IOS Press, 2003.

[2] Farid Abedan Kondori, Shahrouz Yousefi, Li Liu, Haibo Li, "Head Operated Electric Wheelchair", in Proc. 2014 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI), San Diego, CA, USA, 6-8 April 2014, pp. 53-56.

[3] Mohammed Faeik Ruzaij, Sebastian Neubert, Norbert Stoll, Kerstin Thurow, "Auto Calibrated Head Orientation Controller for Robotic-Wheelchair Using MEMS Sensors and Embedded Technologies", in Proc. 2016 IEEE Sensors Applications Symposium (SAS 2016), Catania, Italy, 20-22 April, 2016, pp. 433-438.

[4] Kazuo Tanaka, Kazuyuki Matsunaga, and Hua O. Wang, "Electroencephalogram-Based Control of an Electric Wheelchair", IEEE Transactions on Robotics, Vol. 21, No. 4, August 2005.

[5] K. Tanaka, K. Matsunaga, and H. O. Wang, "Electroencephalogram-Based Control of an Electric Wheelchair," IEEE Transactions on Robotics, Vol. 21, No. 4, pp. 762–766, 2005.

[6] I. Iturrate, J. Antelis, and J. Minguez, "Synchronous EEG brain-actuated wheelchair with automated navigation," in Proc. of IEEE International Conference on Robotics and Automation (ICRA '09), Kobe, Japan, 2009, pp. 2318–2325.

[7] Masato Nishimori, Takeshi Saitoh, and Ryosuke Konishi, "Voice Controlled Intelligent Wheelchair", in Proc. SICE Annual Conference 2007, Sept. 17-20, 2007, Kagawa University, Japan, pp. 336-340.

[8] Aruna. C, Dhivya Parameswari. A, Malini. M and Gopu. G, "Voice Recognition and Touch Screen Control Based Wheelchair for Paraplegic Persons", in Proc. Green Computing, Communication, and Electrical Engineering, Coimbatore, 2014, pp. 1-5.

[9] Mohammed Faeik Ruzaij, Sebastian Neubert, Norbert Stoll, Kerstin Thurow, "Hybrid Voice Controller for Intelligent Wheelchair and Rehabilitation Robot Using Voice Recognition and Embedded Technologies", Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol. 20 No. 4-2016, Fuji Technology Press Ltd, Tokyo, Japan, pp. 615-622.

[10] Satoshi Ohishi and Toshiyuki Kondo, "A Proposal of EMG-based Wheelchair for Preventing Disuse of Lower Motor Function", in Proc. Annual Conference of Society of Instrument and Control Engineers (SICE), August 20-23, 2012, Akita University, Akita, Japan, pp. 236-239.

[11] Z. Yi, D. Lingling, L. Yuan, and H. Huosheng, "Design of a surface EMG based human-machine interface for an intelligent wheelchair," in Proc. of 2011 10th International Conference on Electronic Measurement Instruments (ICEMI), Chengdu, China, 2011, pp. 132–136.

[12] Mohammed Faeik Ruzaij, Sebastian Neubert, Norbert Stoll, Kerstin Thurow,"Design and Testing of Low Cost Three-Modes of Operation Voice Controller for Wheelchairs and Rehabilitation Robotics", in Proc. 9th IEEE International Symposium on Intelligent Signal Processing WISP2015, Siena, Italy, May 15 -17, 2015, pp. 114-119.

[13] Mohammed Faeik Ruzaij Al-Okby, Sebastian Neubert, Norbert Stoll, Kerstin Thurow, "Low-Cost Hybrid Wheelchair Controller for Quadriplegias and Paralysis Patient", Advances in Science, Technology and Engineering Systems Journal (ASTESJ), Vol. 2, No. 3, pp. 687-694, 2017.

[14] Mohammed Faeik Ruzaij, Sebastian Neubert, Norbert Stoll, Kerstin Thurow, "Multi-Sensor Robotic-Wheelchair Controller for Handicap and Quadriplegia Patients Using Embedded Technologies", ", in Proc. 2016-9th International Conference on Human System Interactions (HSI), Portsmouth, United Kingdom, 6-8 July 2016. pp.103-109.

[15] Mohammed Faeik Ruzaij, Sebastian Neubert, Norbert Stoll, Kerstin Thurow, "Design and implementation of low-cost intelligent wheelchair controller for quadriplegias and paralysis patient", in Proc. 2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMI), Herl'any, Slovakia, 26-28 January 2017, pp. 399-404

[16] Speak Up Click User Manual Ver.101, MikroElektronika, Belgrade, Serbia, 2014.

[17] EasyVR 2.0 User Manual R.3.6.6., TIGAL KG, Vienna, Austria, 2014.

[18] BNO055 data sheet, [online]. Available: https://ae-bst.resource.bosch.com/media/_tech/media/datasheets/BST_BNO055_DS000_14.pdf. [Accessed: 15 Jun 2016].

[19] EFM32GG990 DATASHEET, [Online]. Available: http://www.silabs.com. [Accessed: 20 December 2015].

Contactless Power and Bidirectional Data Transmission via Magnetic Field

Jia-Jing Kao, Chun-Liang Lin[*], Chih-Cheng Huang, Hau-Shian Jian

Department of Electrical Engineering, National Chung Hsing University, Taichung, Taiwan

A R T I C L E   I N F O

A B S T R A C T

At present, the wireless charging and data transmission systems developed by major car manufacturers rely heavily on additional transmission equipment. This paper presents a newly developed wireless power and bidirectional data transmission scheme without adding any radio frequency (RF) devices. The battery charge status and vehicle related information can be transmitted bidirectionally in the wireless manner by the compensation capacitor between two isolated units. Based on the function, this system can quickly inform the user when an emergency event occurs during power transmission. An inverter is used on the primary side of the system, and the generated AC power is transmitted to the load at the secondary side through mutual inductance; the secondary side of the system adjusts the load current to transmit data back. The primary side uses the zero voltage switching (ZVS) method to receive data, and the trimming of the current curve to transmit commands; the secondary side of the system receives commands and uses carrier cycles for decoding. Experimental verification shows applicability of the proposed system.

1. Introduction

The principle of inductive power transfer (IPT) is the mutual magnetic coupling between the transmission windings. It transfers the main power source on the fixed side to a plurality of movable devices by means of the air gap [1]. The issue is becoming popular due to the increasingly global concern of energy conservation. However, most of the previous research tasks only focus on the issue of IPT without considering simultaneous power and data transmission. This research intends to deal with the issue with a novel idea. The paper presented here is an extension of our work originally presented in the *2017 12th IEEE International Conference on Industrial Electronics and Applications* [2] by including extended results.

Traditional transferring power is accomplished through the contact of the metal contacts (Plug-In). Although this kind of power transmission can meet most of the requirements, it is inconvenient to be used in many occasions, such as moisture, snow, oil exploitation, and underwater power supply. In order to avoid the shortcomings of contact power transmission, contactless power transfer (CPT) has been developed, which is one of the popular wireless power transfer (WPT) technologies. It has no metal contact and hence avoiding the weakness.

The wireless charging method has been receiving attention over a wide range of industrial and civilian applications in recent years which leads to a variety of utilizations such as contactless electric vehicle charging [3-5], bidirectional IPT system [6-8],

electric vehicle and power transmission between grid [9], wireless power and data transmission with fully implantable stimulator [10], portable-telephone charging [11], dynamic charging of electric vehicle [12-13], and wireless power and data transmission system [14].

In literature [7], the proposed IPT system has been applied to vehicle battery charging with high efficiency. Its control loop is composed of external and internal paths, used to handle power control between two sides with an additional wireless transmission equipment to transfer data from the secondary side to the primary side. This design requires better transmission security between two isolation units. The IPT system of [15] uses non-contact technology in the drilling machine, and two sets of windings are used for power supply and data transmission respectively. The design would lead to extra production cost.

The architecture of this research task is the use of a flyback converter, which is used for wireless power and bidirectional data transmission, and adopts ZVS technology to reduce switching loss and EMI noise. The secondary side of the system performs data modulation by instantaneously adjusting load current, and the primary side of the system receives data under ZVS. In addition, the system trims the current transmitting command on the primary side, and the secondary side of the system receives command and decodes data. The authors of [15] proposed a transient load detection method, which mainly uses energy injection and free resonance mode to detect the load. Although a similar approach was used here, wireless power and bidirectional data transmission

[*] Chun-Liang Lin, Email: chunlin@dragon.nchu.edu.tw

are the focus of this research. Our design proposes an effective power and bidirectional data transmission solution under the architecture of the IPT system. The technology can be combined with a vehicle positioning system to realize a wireless vehicle charging station as the final goal. As for a practical application in the micro grid, when a driver is charging his/her car, if the charge station can immediately know the voltage, current, power and related information of the charging car, the accident can be effectively managed. Beside of this, if the car is used as a mini power generator, the grid side may identify car ID information and count how much power it contributing to the grid.

2. Description of System

2.1. Circuit Design

This research task considers a wireless vehicle charging system with the function of synchronous power and bidirectional data transmission. The primary side (Grid side) uses a utility power supply (AC 110V / 60Hz). The secondary side (Vehicle side) stores the energy transmitted from the grid side in the vehicle battery. In addition to power transmission, both sides can send and receive messages simultaneously through the same conduction windings. At startup, the primary side of the system performs two steps in sequence, namely payload detection and soft start. First, it detects whether there is a payload on the secondary side of the system. If no payload is detected on the pad's inductance, the primary side stops power output. If a load is detected, the soft start current will give rise to a preset level, and this mechanism minimizes the inrush current flowing through the switch. The schematic diagram of the proposed system is illustrated in Fig. 1.

The flyback converter adopted here is widely used in AC-DC conversion or DC-DC conversion. Its operational principle is similar to the buck-boost converter. In addition to voltage conversion, there is electrical isolation between the input and output circuits.

The system consists of two circuits that are magnetically coupled to each other. The grid-powered battery charging station (GPBCS) converts the main power supply to direct current through a rectifier and a filter. The transmitter generates power at the primary windings and transmits it to the secondary windings, which is located inside of the vehicle, by mutual inductance to complete the wireless power transfer.

The power switch (MOSFET) N_2 at the secondary side has two functions. First, it bypasses the output current and then sends "T_X data" such as output voltage and current back to the primary side. The way in which the primary side receives data is to detect the period of the feedback voltage V_C. For data communication, when the T_X data is received and confirmed by the primary side, the switch (N_1) at the primary side sends a command to the secondary side. The current transducer (CT) at the secondary side is connected to the circuit to receive the command.

As can be seen from Fig. 1, the magnetic coupling coefficient and the load quality factor Q of the secondary side are closely related to the mutual inductance M. Its equation is as follows:

$$k = \frac{M}{\sqrt{L_p L_s}}$$

$$Q = \omega C_s R_L$$

where L_P and L_S are primary side inductance and secondary side inductance, respectively and C_S is secondary side capacitance, R_L is load on the secondary side. The operating frequency and rated power of the power supply are mainly limited to the switching components used, both must be balanced according to the copper losses. In this research, the appropriate Q and k were chosen to maintain efficiency of power transmission [5], and S/N ratio, thereby controlling quality of the transmitted data.
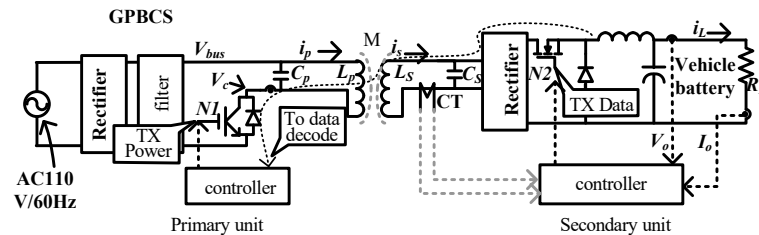
(a) Transmission path of data feedback to the primary (Grid) unit
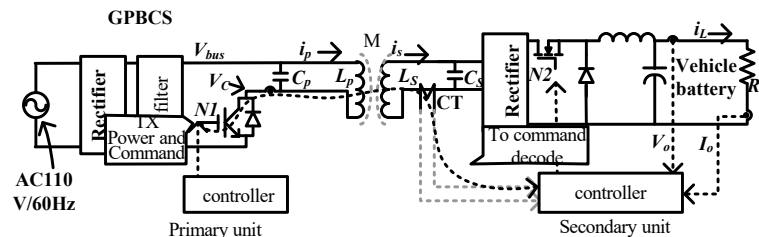
(b) Transmission path of command sending to the secondary (Vehicle) unit

Figure 1. Simultaneous wireless power and bidirectional data transmission scheme

The simultaneous wireless power and bidirectional data transfer system adopts parallel–parallel (PP) reactive power compensation topology. The primary side for parallel compensation is used to generate a large primary current; the secondary side suitable for battery charging current sources [16]. The PP type compensation topology is illustrated in Fig. 2 (a). The topology under mutual inductance coupling can be represented by the simplified circuit depicted in Fig. 2 (b).

As shown in Fig. 2, the impedance reflected from the secondary side back to the primary side is given by

$$Z_r = \frac{\omega^2 M^2}{Z_s} \tag{3}$$

The voltages of the primary windings and secondary windings are expressed as follows:

$$V_p = -j\omega M I_s + j\omega L_p I_p$$

and

$$V_s = -j\omega L_s I_s + j\omega M I_p$$

The current of the secondary winding can be expressed by the following equation:

$$I_s = \frac{j\omega M I_p}{Z_s}$$

with the resonant frequency of both sides being

$$\omega = \frac{1}{\sqrt{L_p C_p}} = \frac{1}{\sqrt{L_s C_s}}$$

The secondary side also transmits data to the primary side with the same mutual inductance. The load impedance of the primary side can be known for the network combination of the primary side and the secondary side, namely,

$$Z_t = \frac{1}{j\omega C_p + \dfrac{1}{j\omega L_p + Z_r}}$$

$$C_p = \frac{\left(L_p L_s - M^2\right) C_s L_s^3}{M^4 C_s R + L_s \left(L_p L_s - M^2\right)^2} \tag{9}$$

(b) Mutual inductance coupling circuit
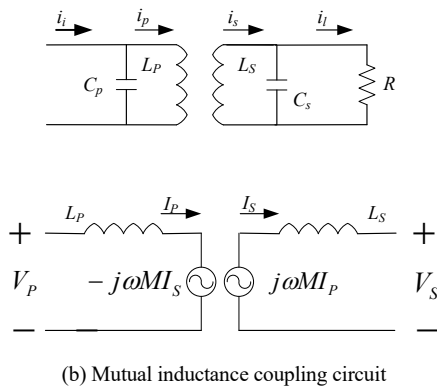
Figure 2. Compensation topology and the simplified circuit

To minimize the VA rating, the imaginary part of the load impedance is adjusted to zero when the secondary side is working at the resonant frequency. The selected primary side capacitance C_P is as follows

3. Wireless Data Transmission

3.1. Communictaion Protocol

The data packet can be divided into four parts, namely start segment, data segment, checksum segment, and stop segment. The entire packet architecture is referenced to the 1-Wire communication protocol and fine-tuned to construct a packet suitable for use with the system. In addition, the universal asynchronous receiver transmitter (UART) specific start segment and stop segment are also retained in the system to separate the packets. The major data types in the data segment are command, voltage value, current value, power, and stop power transmission. The checksum segment is used to detect the entire packet error.

The messages we are about to transmit are divided into two categories based on their attributes. The first category is normal messages including output voltage and current, which is handled with a slower rate. The second category is emergency messages such as overvoltage and overcurrent, which is processed with a faster rate for instantaneous reaction by the system. In addition, when T_X data of the secondary side is sent back to the primary side, the system will calculate the charging efficiency. The primary side is the command issuer, which will send the T_X cmd to the secondary side. Table I summarizes the specifications for data transmission.

Table 1: Protocol for data transmission

Category	Baud rate	Start segment	Data segment	Checksum segment	Stop segment
Normal messages	120 Bd.	1 bit	8 bits	8 bits	1 bit
Emergency messages	Operating freq. Bd.	1 bit	4 bits	4 bits	1bit

3.2. Data Transmission

When the primary side transmits power to the secondary side, description of data transmission between two sides is explained as in Fig. 3. In this figure, the period between zero-crossing will coincide with the AC line. The time base clock on the primary side is the "Digital AC line" signal generated by the zero-crossing detection circuit. In order to be synchronized with the primary side of the Digital AC line, the secondary side uses a current sensor to measure I_S, which is then compared to produce a "Digital I_S" signal.

Signal modulation technique is commonly adopted in communication and electronic systems to mix and transmit signals for data transmission in one or more carrier waves. General modulation modes include amplitude-shift keying (ASK), frequency-shift keying (FSK), and phase-shift keying (PSK). In which, FSK has been widely used in digital signal modulation for communication systems such as emergency broadcasts. It uses discrete frequency changes of the carrier to modulate data for transmission. In which, the carry frequency f_C is used to

discriminate digital "0" or "1". Referring to Fig. 3, in the preamble section, T_{OFF} of the duty cycle of N_2 is fixed. When the secondary side transmits data in the data section, the primary side receives the data in the shadow section; when the primary side transmits the command in the cmd section, the secondary side receives command in the shadow section.
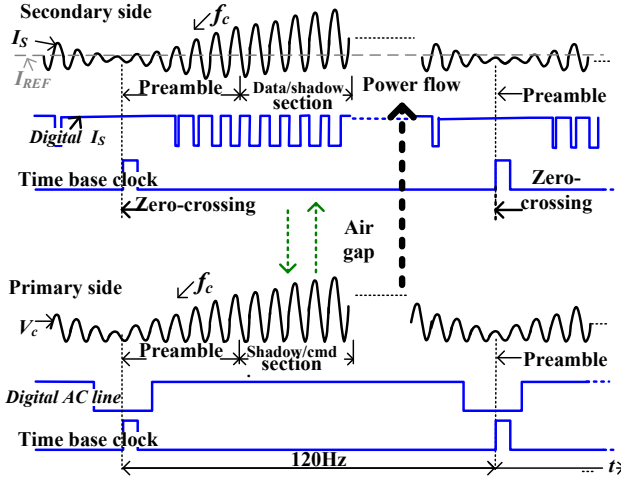
Figure 3. Description of bidirectional data transmission via the mutual magnetic coupling circuits

3.3. Data Decoding

When data is allocated in the data section on the secondary side and modulated to $T_Z(0)$, the controller reduces the duty cycle of N_2 and the output current (See Fig. 1). The reduced current causes a time delay on the secondary side, which we denote it as $\Delta\tau$. Further, as can be seen from Fig. 4, N_2 is turned off in the c-d section so that the current I_S flows through the load R_L is stored in C_S; when N_2 works in the d-e section, it is turned on, and C_S will discharge to R_L. The delayed discharge of C_S directly affects V_C on the primary side, causing V_C to generate a delay time at 0V. At this time, the d-e section corresponding to V_C will have a bump. Since N_1 operates based on ZVS, the turn on time of N_1 is also delayed. The additional delay time produced by the a-e section is also defined as $\Delta\tau$ after comparison with the e-g section. The time delay $\Delta\tau$ resulting from modulation of the digit "0" is as follows

$$T_Z(0) = T_C(1) + \Delta\tau \tag{10}$$

where $T_Z(0)$ under modulation denotes the period of the carrier wave for digit "0" and $T_C(1)$ for the digit "1". The term $\Delta\tau$ can be referred to the S/N ratio. An increase in $\Delta\tau$ will result in better data transmission quality, making it more robust against noise, and it is more advantageous for the controller on the primary side to decode the signal. However, it might deteriorate the power transmission quality.

It can be known from the above that the secondary side can send the numbers "0" and "1" by $T_Z(0)$ and $T_C(1)$ respectively in the data section; the primary side receives the modulated data, which being decoded as "0" or "1" is based on $T_Z'(0)$ or $T_C(1)$. The rule for the primary side distinguishes the digital "0" is determined by

$$T_Z'(0) > T_C(1) + \Delta ref \tag{11}$$

where $_{\Delta ref}$ is a threshold set to identify the digit "0" at the primary side. Illustration of the idea is depicted in Fig. 5 which explains the function for modulating a digit "0" at the center of the data string where curve 1 is I_S, curve 2 is the switching state of N_2, curves 3 and 4 are the gate voltages of V_C and N_1, respectively. After changing the duty cycle of N_2 (100% adjusted to 94%), it can be seen that the waveform of I_S becomes the defined $T_Z(0)$; the waveform of the synchronous V_C also becomes the defined $T_Z'(0)$. As can be seen from the results in the figure, when the periods of I_S and I_P are the same, the operating frequency will be affected when I_S changes.
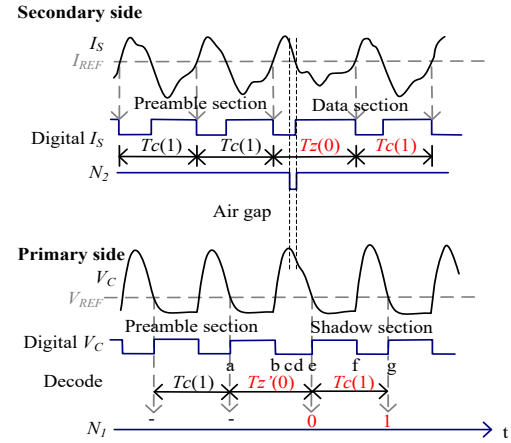
Figure 4. Data decoding method

Fig. 6 shows the varying duty cycle of N_2 vs. $\Delta\tau$. It can be seen from the figure that after adjusting the duty cycle, a near linear relationship can be obtained between 85% and 94%. The relationship is as follows:

$$\text{Modulated } T_{off} \square \ \Delta\tau + \varepsilon \tag{12}$$

where the modulated T_{off} is duty off of N_2 for the secondary side to send the digit "0", the uncertainty $|\varepsilon| \le 0.5\mu s$ is set in our experiments. If the modulated $T_{off} + \varepsilon > \Delta ref$, the primary side decodes data as the digit "0".
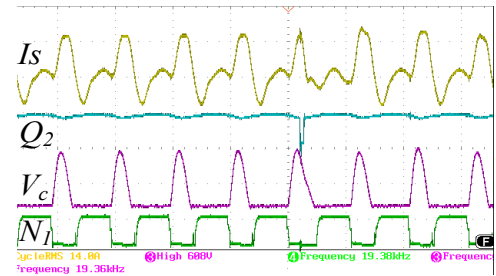
Figure 5. The digit "0" is modulated by reducing the duty ratio of N_2 from 100% to 94% when transmission power is at 700 W
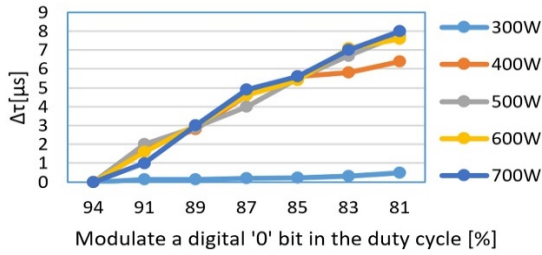
Figure 6. Relationship between duty cycle of N_2 and $\Delta\tau$

3.4. Command Transmission

In order to prevent conflicts during data transmission, the system-defined data arrangement string is shown in Fig. 7 where T_X cmd is the corresponding command message generated on the secondary side after sending the message on the primary side. The reset time is set to wait for the next string. As for T_X data, it is the feedback from the normal message of the secondary side, such as output voltage and current. Fig. 8 depicts how the system processes T_X cmd synchronously during the wireless power transfer.
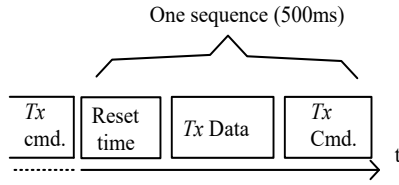
Figure 7. Data transmission scheme

When the system is operating normally and there is no emergency on the secondary side, the primary side places command in the Cmd. section with baud rate 120 Bd. To react, the secondary side detects command accordingly in the shadow section from "Digital I_S". When modulation is performed to the digital "0", the primary side changes the duty cycle of the switch N_1. As seen from Fig. 9, the time required for T_{ON} is shorter when modulating "0" and the time required for T_{ON} when modulating "1" is longer. The reduced current causes a loss of time on the primary side, which is what we call $\Delta\tau$. Referring to Fig. 8, the primary side will arrange data to be sent in the Cmd. section. The turn on time N_1 in the b-c section will be shorter than other parts. The current change of I_P will affect the secondary side, which is to be explained later. It causes a reduced time period $\Delta\tau$ at the shadow part in the a-d section than that in the d-e section. It can be found from the upper sub-figure of Fig. 8 that the term $\Delta\tau$ generating the digital messages "0" is defined as

$$T_Z^{P2S}(0) = T_C(1) - \Delta\tau \qquad (13)$$

where $T_Z^{P2S}(0)$ denotes the primary side modulating a digit "0" to the secondary side. The secondary side receives command and decodes it to be "0" or "1" depending on $T_C(1)$ and $\Delta\tau$. The secondary side distinguishes the digital "0" based on the following criterion:

$$T'^{P2S}_Z(0) < T_C(1) - \Delta ref \qquad (14)$$

where $T_Z'^{P2S}(0)$ denotes the secondary side receiving a digit "0" and Δref is a threshold for judging the digit "0" at the secondary side. If $\Delta\tau > \Delta ref$, the primary side trims the duty cycle of N_1 as

$$\text{Trimmed } T_{on} \approx \Delta\tau + \varepsilon \qquad (15)$$

where $\Delta\tau$ and ε are defined as above. If Trimmed $T_{on} > \Delta ref$, the secondary side decodes data as the digit "0". Figs. 10 and 11 display the function for modulating a digit "0" ($\Delta\tau = 4\mu s$) at the center of the data string. It is seen that $T_Z^{P2S}(0)$, by modulating the duty cycle of N_1, is relatively shorter than that of modulating a "1" by $4\mu s$. Fig. 9 shows simultaneous wireless 700 W power transfer while simultaneously transferring Tx cmd the under the air gap of 25 mm and 0 mm lateral offset.
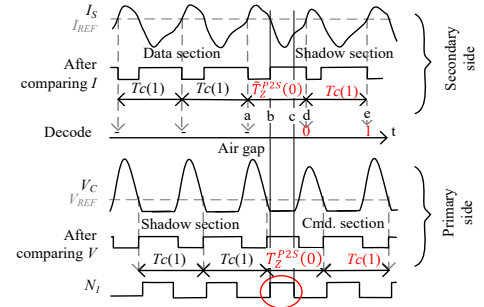
Figure 8. Working mechanism of command transmission
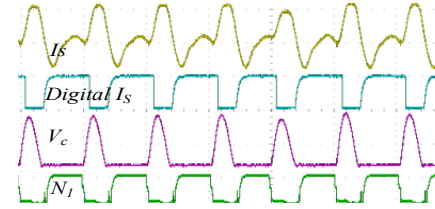
Figure 9. Simultaneous transmission of T_X cmd at 700 W power transmission (Under 25 mm air gap and 0 mm offset)

3.5. Emergency Message

If there is abnormal voltage or current detected during power transmission, the system will stop charging with the highest priority.

We reduce the duty cycle of N_2 from 100% to 94% to generate an additional delay time $\Delta\tau = 4\mu s$ and to confirm the result. The results are shown in Figs. 10 and 11 for the cases of 400 W or 700 W power transmission. It is seen that the primary side can still correctly decode data whether the IPT system works at low or high power.
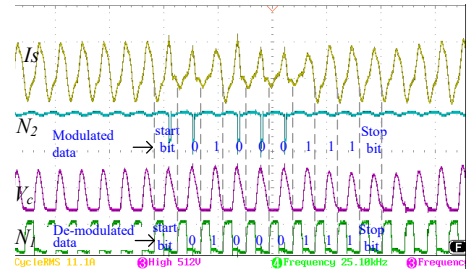
Figure 10. Synchronous transmission of emergency message when transmitting power at 400 W
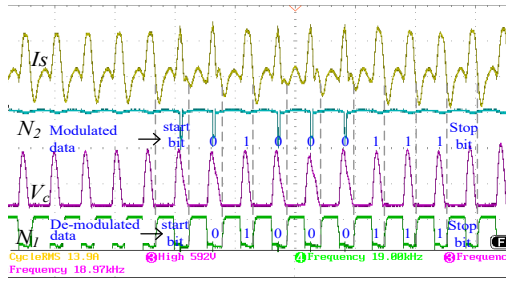
Figure 11. Synchronous transmission of emergency message when transmitting power at 700 W

3.6. Transmission Windings

IPT is a kind of electricity conversion technology. For operation, the primary side windings pass the alternating current I_P and the alternating magnetic field intensity H is generated which generates the magnetic flux density ψ via the air medium coupling. This affects the secondary side windings and induces the electromotive force ε (see Fig. 12).

The windings used in this research are the spiral type and the winding ends are not connected. In the real-world experiment, in order to reduce the internal impedance of the windings, the coil has a wire diameter of 2.0 mm to increase the efficiency of transmission. The diameter of the windings is 165 mm, the resonant frequency of the windings is 22 kHz when loaded, and the resonant frequency of the windings is 32 kHz without load.
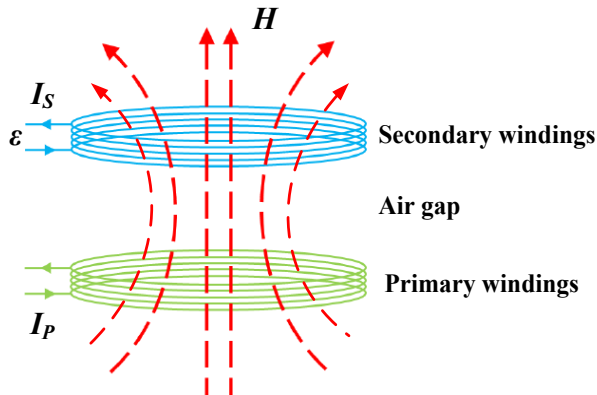
Figure 12. Electromagnetic induction diagram of two transmission windings
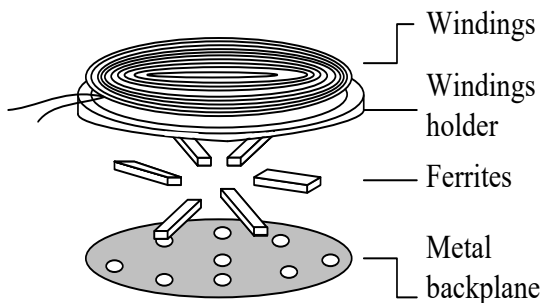
Figure 13. Transmission windings construction used in this research

3.7. Interface

The human machine interface (HMI) control circuit is shown in Fig. 14. It uses the Arduino MEGA 560 as the coordination hub of the entire HMI control circuit, and equipped with an HD44780

universal 16-inch liquid crystal display (LCD) to display the transmission information. The user control interface is implemented by a simple button circuit to facilitate the status display of bidirectional power and data transmission. It communicates with the MCU of the IPT system through a UART communication interface.

The Arduino MEGA 2560 receives the decoded data packet in real time. It also detects the loop of the button instantly. When the user presses the button, it sends a data packet to the IPT system for use by the encoding circuit. Fig. 15 shows the entity of the HMI.
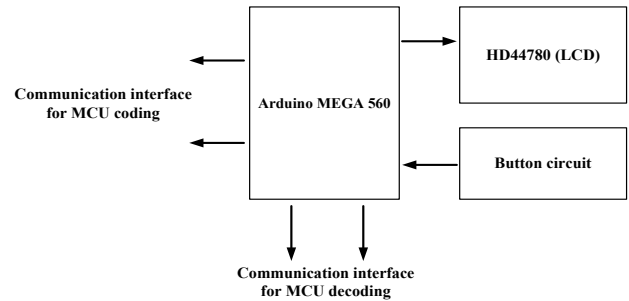
Figure 14. HMI diagram
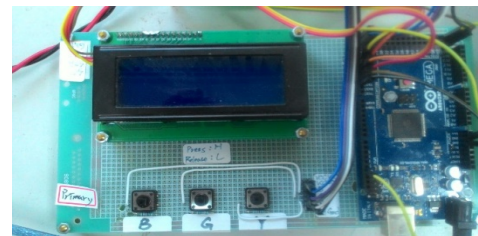
Figure 15. HMI circuit

3.8. Measurement

The primary side controls output power according to the demand of the secondary side. The later simultaneously receives power and measured voltage, current, and power via the way of magnetic field coupling.

The secondary side employs a voltage attenuator and a voltage amplifier to detect charging voltage and current. The crest factor of both signals was designed to equal to 3. These normalized analog signals were connected to a 12-bit ADC device shown as in Fig. 16. In the figure, in addition to measurement, the tasks of the CPU include data encoding, data decoding, and HMI process. To reduce power consumption, the ADC device utilizes the under-sampling method [17-18] to sample signals. It adopts low sampling rate and long measurement period to improve measurement accuracy. The measured raw voltage and current at the sampling time are given by

$$v\_raw_i = \sqrt{\frac{1}{N}\sum_{j=1}^{N} v\_Gain\left(v_j - v\_dc\_offset\right)^2} \tag{16}$$

$$i\_raw_i = \sqrt{\frac{1}{N}\sum_{j=1}^{N} i\_Gain\left(i_j - i\_dc\_offset\right)^2} \tag{17}$$
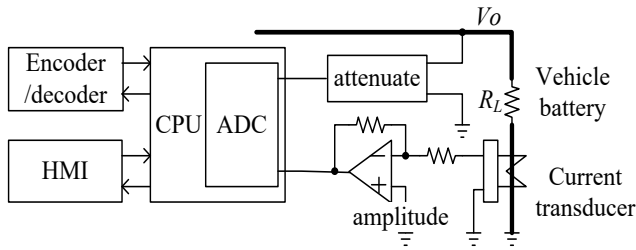
Fig. 16. The secondary control unit with measurement

where N is data update rate, v_Gain and i_Gain are voltage and current gain, respectively. These gains were generated by full-scale calibration. The terms v_j and i_j are the jth voltage and current sampling values, respectively, v_dc_offset and i_dc_offset are inherent dc-offsets.

Voltage and current are obtained by calculating

$$AVG(v_{raw_{i-K-1}} + v_{raw_{i-K-2}} + .. + v_{raw_{i-1}} + v_{raw_i}) \quad (18)$$

$$I_i = AVG(i_{i-K-1} + i_{raw_{i-K-2}} + .. + i_{raw_{i-1}} + i_{raw_i}) \quad (19)$$

where AVG denotes "average" and K is index of measurement period.

We use the under-sampling approach to measure charging voltage, current, and power. Comparison of the parameters of under-sampling and over-sampling methods is given in Table II. The experimental results are shown in Fig. 17. The measurement error is larger at low output power because of the nonlinear distortion of voltage attenuator and amplifier.
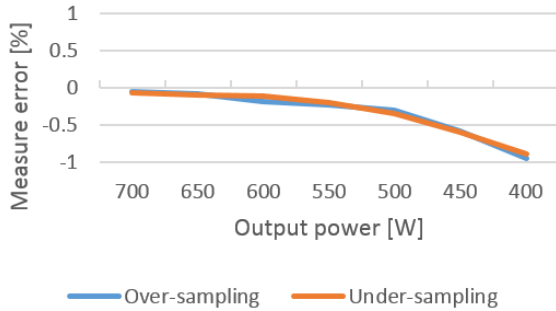
Figure 17. Comparison of the under-sampling and over-sampling methods

Table 2: Measured parameters of under sampling method vs. over-sampling

Parameter	Sampling type	
	Under sampling	Over sampling
ADC resolution [bits]	12	
Crest factor	3	
Data update rate [ms]	250	250
Sampling rate [kHz]	$\fallingdotseq 5$	200
Measurement period [ms]	2000	250

3.9. Efficiency

The primary side controls output power based on the ZVS method. The operational frequency was varied by various output

power or demand between 19~30 kHz. The sampling rate was 200 kHz. The channels of sampling voltage and current are V_{bus} and i_P, respectively. Their crest factor is 3. Both signals are connected to the input of ADC and controller for measurement.
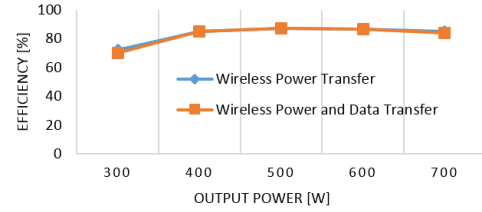
Figure 18. Comparison of efficiency change with output power

Drop of the power transfer efficiency is shown in Fig. 18 when power and data are transmitted simultaneously. As seen from the figure, this will only result in a slight decrease in the power transfer efficiency.

4. Experimental Results

Since the primary side operates at ZVS, the operating frequency of the system can be changed by the load R_L. In addition, Q_1 on the primary side can have less switching loss and higher efficiency under ZVS operation. If the air gap of two transmission windings changes, the operating frequency of the system will vary. To resolve solve the problem, both units measure the period of T_C (1) in the preamble section as a reference.

We conduct the following experiments to verify robustness of data transmission quality (TQ) while there is misalignment between two induction pads or there is load change. We consider Δref to be $\Delta \tau / 2$ and parameters listed in Table III. Fig. 19 shows the system hardware and test environment. Power transmission in both sides are controlled by ARM Cortex-M0. The clock speed of the microcontroller is 24 MHz. The data transmission quality (TQ) is defined by

$$CQ \square \frac{pass\ count}{total\ count}\% \quad (20)$$

When transmitting a normal message, the total variable count is 1,000. The pass count for the number of correct packets is defined by the following equation:

$$CS \square no.\ of\ Data\ 2's \quad (21)$$

Data received is treated as a correct message if the checksum (CS) is equal to $Data2's$.
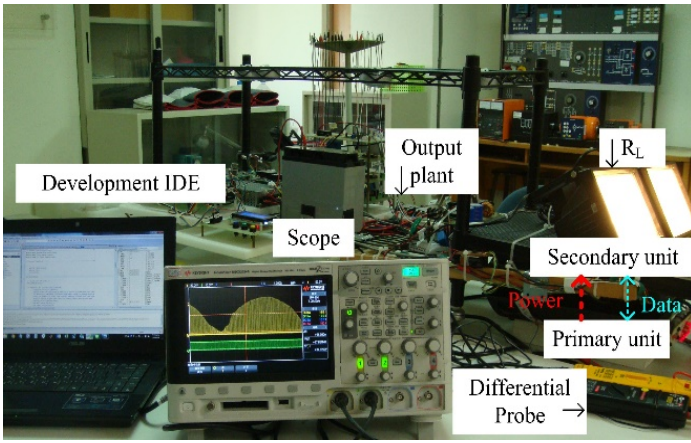
Figure 19. Hardware relaization of the IPT system

Fig. 20 presents the result of data TQ. It was recorded by adjusting the lateral shift of two pad's center from -20 mm to +20 mm. It is found from the experimental results that regardless of the lateral shift between the two pads, an increase in $\Delta\tau$ can significantly improve the data TQ. In addition, when the lateral shift of the two pads is maintained within -5 mm to +5 mm, and $\Delta\tau$ is set to between $3\mu s$ to $5\mu s$, the data TQ can be as high as 90% or more.

Fig. 21 shows the correlation between output power and data TQ. It can be seen from the figure that communication performance is acceptable (Variation less than 10%) when the system works within 400 W to 700 W and $\Delta\tau$ lying between $3\mu s$ and $5\mu s$.

Table 3: Parameters of the prototype IPT system [1]

Operating frequency		22 kHz
Windings type		Spiral
Power MOSFET	V_{DS}	500 V
	V_{GS}	±20 V
Rated output power		700 W
Rated load		40 Ω
Wire diameter		2.0 mm
Grid side inductance		70 uH
Grid side capacitance		0.35 uF
Grid side induction coil outer diameter		165 mm
Grid side winding number		20 turns
Mutual inductance		26.1 uH
Vehicle side inductance		69.4 uH
Vehicle side capacitance		0.35 uH
Vehicle side induction coil diameter		165 mm

Vehicle side winding number	20 turns
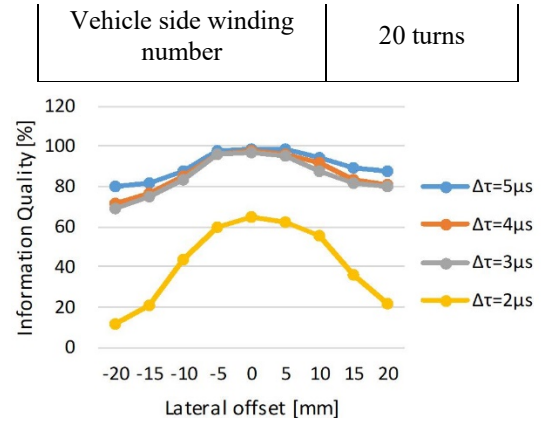
Figure 20. Data TQ for different lateral offsets between two induction pad centers of L_P and L_S with the nominal air gap 25 mm
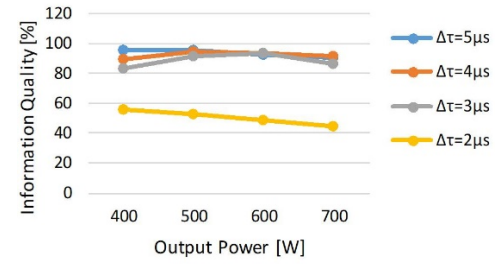
Figure 21. Data TQ vs. output power

Fig. 22 shows that the results of data TQ vs. air gap when the conduction windings L_S and L_P lies between 15 mm and 25 mm. It is quite clear that performance of data transmission drops quickly when the air gap is greater than 25 mm. However, the experimental results were produced by the prototype sysytem. In general, the optimal air gap of data transmission between two windings may change with charcteristics of the system.
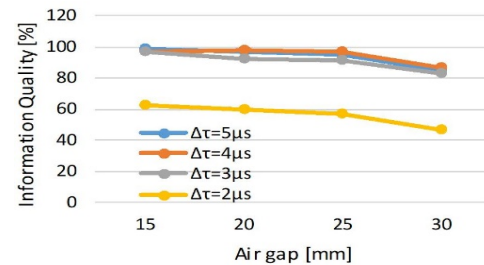
Figure 22. Data TQ vs. air gaps between two induction pads

5. Discussions

While experiments presented in this report demonstrate satisfactory results in either power transmission or data transmission, the power transfer efficiency might drop rapidly due to visible misalignment of two induction pads between vehicle chassis and power station as revealed in Fig. 20. Practically, it is impossible for the driver to move car perfectly so that there is no misalignment between two induction pads. There is a need to develop an adaptive bidirectional wireless power and data transmission control strategy for the optimal TQ under the situation of lateral misalignment between two induction pads. In addition, a clear relationship between data transmission efficiency and physical parameters of the LC tank remains to be investigated.

6. Concluding Remarks

This research report presents a newly developed power and bidirectional data transmission design. The major contribution is that the state of charge of the vehicle battery and vehicle ID information could be mutually communicated through two contactless units. The design exhibits several advantages. For example, the primary (Grid) side of the system can control the output current and achieve steady-state compensation of the secondary (Vehicle) side of the system by wireless transmission. In addition, it costs relatively less while compared with blue tooth or other RF communication technologies because no extra data communication device devices needed. Most importantly, it is a kind of transmission via magnetic field, therefore, there is no serious issue on EMC/EMI. That is, the system is robust against RF interference which makes it to be applicable in the indoor parking lot. To verify applicability, experiments have been conducted with different lateral offsets and vertical shifts between two induction pads which demonstrates performance robustness in power and data transmission.

Acknowledgment

References

[1] J. Dai, D. C. Ludois, "A survey of wireless power transfer and a critical comparison of inductive and capacitive coupling for small gap applications" IEEE Trans. Power Electron., 30(11), 6017-6029, 2015. https://doi.org/10.1109/TPEL.2015.2415253

[2] C. C. Huang, H. H. Jian, J. J. Kao, C. L. Lin, Y. T. Kuo, "New Design of Wireless Power and Bidirectional Data Transmission" in 2017 12th IEEE Conference on Industrial Electronics and Applications, Siem Reap, 2017. https://doi.org/10.1109/ICIEA.2017.8282873

[3] X. Qu, H. Han, S. C. Wong, C. K. Tse, W. Chen, "Hybrid IPT topologies with constant current or constant voltage output for battery charging applications" IEEE Trans. Power Electron., 30(11), pp. 6329-6337, 2015. https://doi.org/10.1109/TPEL.2015.2396471

[4] E. A. Mehdi, K. Chma, J. Stefani, "Rapid-charge electric-vehicle stations" IEEE Trans. Power Delivery, 25(3), pp. 1883-1887, 2010. https://doi.org/10.1109/TPWRD.2010.2047874

[5] C. H. Ou, H. Liang, W. Zhuang, "Investigating wireless charging and mobility of electric vehicles on electricity market" IEEE Trans. Ind. Electron., 62(5), pp. 3123-3133, 2015. https://doi.org/10.1109/TIE.2014.2376913

[6] T. Diekhans, R. W. D. Doncker, "A dual-side controlled inductive power transfer system optimized for large coupling factor variations and partial load" IEEE Trans. Power Electron., 30(11), pp. 6320-6328, 2015. https://doi.org/10.1109/TPEL.2015.2393912

[7] U. K. Madawala, M. Neath, D. J. Thrimawithana, "A power–frequency controller for bidirectional inductive power transfer systems" IEEE Trans. Ind. Electron., 60(1), pp. 310-317, 2013. https://doi.org/10.1109/TIE.2011.2174537

[8] D. J. Thrimawithana, U. K. Madawala, M. Neath, "A synchronization technique for bidirectional IPT systems" IEEE Trans. Ind. Electron., 60(1), pp. 301-309, 2013. https://doi.org/10.1109/TIE.2011.2174536

[9] Y. Ma, T. Houghton, A. Cruden, D. Infield, "Modeling the benefits of vehicle-to-grid technology to a power system" IEEE Trans. Power Syst. 27(2), pp. 1012-1020, 2012. https://doi.org/10.1109/TPWRS.2011.2178043

[10] Q. Xu, D. Hu, B. Duan, J. He, "A fully implantable stimulator with wireless power and data transmission for experimental investigation of epidural spinal cord stimulation" IEEE Trans. Neural and Rehabilitation Eng., 23(4) pp. 683-692, 2015. https://doi.org/10.1109/TNSRE.2015.2396574

[11] Y. Jang, M. M. Jovanovic, "A contactless electrical energy transmission system for portable-telephone battery chargers" IEEE Trans. Ind. Electron., 50(3), pp. 520-527, 2003. https://doi.org/10.1109/TIE.2003.812472

[12] S. Jeong, Y. J. Jang, D. Kum, "Economic analysis of the dynamic charging electric vehicle," IEEE Trans. Power Electron., 30(11), pp. 6368-6377, 2015. https://doi.org/10.1109/TPEL.2015.2424712

[13] G. R. Nagendra, L. Chen, G. A. Covic, J. T. Boys, "Detection of EVs on IPT Highways" IEEE Emerging and Selected Topics in Power Electron., 2(3), pp. 584-597, 2014. https://doi.org/10.1109/JESTPE.2014.2308307

[14] T. Bieler, M. Perrottet, V. Nguyen, Y. Perriard, "Contactless power and information transmission" IEEE Trans. Ind. Electron., 38(5), pp. 1266-1272, 2002. https://doi.org/10.1109/TIA.2002.803017

[15] Z. H. Wang, Y. P. Li, Y. Sun, C. S. Tang, X. Lv, "Load detection model of voltage-fed inductive power transfer system" IEEE Trans. Power Electron., 28(11), pp. 5233-5243, 2013. https://doi.org/10.1109/TPEL.2013.2243756

[16] C. S. Wang, O. H. Stielau, G. A. Covic, "Design considerations for a contactless electric vehicle battery charger" IEEE Trans. Ind. Electron., 52(5), pp. 1308-1314, 2005. https://doi.org/10.1109/TIE.2005.855672

[17] C. S. Wang, G. A. Covic, O. H. Stielau, "Power transfer capability and bifurcation phenomena of loosely coupled inductive power transfer system" IEEE Trans. Ind. Electron., 51(1), pp. 148-157, 2004. https://doi.org/10.1109/TIE.2003.822038

[18] Y. Liu, J. J. Yan, H. T. Dabag, P. M. Asbeck, "Novel technique for wideband digital predistortion of power amplifiers with an undersampling ADC" IEEE Trans. Microw. Theory Techn., 62(11), pp. 2604-2617, 2014. https://doi.org/10.1109/TMTT.2014.2360398

Towards an Efficient Federated Cloud Service Selection to Support Workflow Big Data Requirements

Mohamed Adel Serhani[1*], Hadeel Al Kassabi[2,1], Ikbal Taleb[2]

[1]*Information Systems and Security Department, College of Information Technology, 15551, UAE*

[2]*Concordia Institute for Information System Engineering, Concordia University, H3G 1M8, Canada*

A B S T R A C T

Cloud Computing is considered nowadays an attractive solution to serve the Big Data storage, processing, and analytics needs. Given the high complexity of Big Data workflows and their contingent requirements, a single cloud provider might not be able alone to satisfy these needs. A multitude of cloud providers that offer myriad of cloud services and resources can be selected. However, such selection is not straightforward since it has to deal with the scaling of Big Data requirements, and the dynamic cloud resources fluctuation. This work proposes a novel cloud service selection approach which evaluates Big Data requirements, matches them in real time to most suitable cloud services, after which suggests the best matching services satisfying various Big Data processing requests. Our proposed selection scheme is performed throughout three phases: 1) capture Big Data workflow requirements using a Big Data task profile and map these to a set of QoS attributes, and prioritize cloud service providers (CSPs) that best fulfil these requirements, 2) rely on the pool of selected providers by phase 1 to then choose the suitable cloud services from a single provider to satisfy the Big Data task requirements, and 3) implement multiple providers selection to better satisfy requirements of Big Data workflow composed of multiples tasks. To cope with the multi-criteria selection problem, we extended the Analytic Hierarchy Process (AHP) to better provide more accurate rankings. We develop a set of experimental scenarios to evaluate our 3-phase selection schemes while verifying key properties such as scalability and selection accuracy. We also compared our selection approach to well-known selection schemes in the literature. The obtained results demonstrate that our approach perform very well compared to the other approaches and efficiently select the most suitable cloud services that guarantee Big Data tasks and workflow QoS requirements.

1. Introduction

Cloud Computing is a promising potential venue for processing Big Data tasks as it provides on-demand resources for managing and delivering efficient computation, storage, and cost-effective services. However, managing and handling Big Data implicates many challenges across several levels, among which are the difficulty of handling the dynamicity of the environment resources, the dataflow control throughout the service compositions, and guaranteeing functional and performance quality. Therefore, abundant Cloud Service Providers (CSPs) offering comparable services and functionalities proliferate in the market to meet the growing challenging demands. Subsequently, the selection of the most appropriate cloud provider is recognized to be a challenging

task for users. Not only appropriate in terms of functionality provisioned, but also satisfying properties required by the user such as specific levels of quality of service and reputation, especially with the exaggerated cloud providers' marketing claims of guaranteed QoS levels.

Hence, providing an automatic and modest means for selecting a cloud provider which will enable Big Data tasks and guarantee a high level of Quality of Cloud Service (QoCS) is a necessity. Moreover, modeling and evaluation of trust among competing cloud providers enables wider, safer and more efficient use of Cloud Computing.

Therefore, it is necessary to propose a comprehensive, adaptive and dynamic trust model to assess the cloud provider Quality of Service prior to making selection decisions.

*Mohamed Adel Serhani, Email: serhanim@uaeu.ac.ae

A large number of CSPs are available today. Most pf CSPs offers a myriad of services, for instance, Amazon Web Service (AWS) offers 674 varying services which are classified according to locality, Quality Of Service, and cost [1]. Automating the service selection to not only rely of simple criterion such as cost, availability, and processing power, but to consider service quality agreement is crucial. Current CPS selection approaches support straightforward monitoring schemes and do not provide a comprehensive ranking and selection mechanism. For instance, CloudHarmony [2] supports up-to-date benchmark results that do not consider the price while Cloudorado [3] supports price measurement, however neglects other dynamic QoS properties.

Selecting the best CSP reveals twofold objectives, and adds value to both CSPs and Big Data users as well as applications. CSPs provision services that attract clients' interest and support their processing and storage needs. However, users must ensure that services they were offered meet their expectation in terms of quality and price.

Difficulties linked to CSP selection to handle Big Data tasks include for example the following: 1) The limited support for Big Data users in describing their various QoS needs of different Big Data tasks. 2) The difficulty to search in a high dimensional database or repository of CSPs. 3) The challenge to consider the continuous variations in the QoS needs and the Big Data related requirements. And 4) The limited support for mapping Big Data task quality requirements to the underlying cloud services and resources quality characteristics. By doing so, we can guarantee an end-to-end quality support from the top-down Big Data quality consideration to cloud services and resources quality enforcement.

Our main objective in this work is to build a full-fledged approach that supports Big Data value chain with the best cloud services and resources that are trustworthy, automatically scale, and support complex and varying Big Data quality requirements. This is possible with the development of a comprehensive cloud services selection model that fulfills the needs of a Big Data job with the efficient supporting cloud services. Our solution will impose QoS of Big Data processes through dynamic provisioning of cloud services by one or multiple CSPs that will ensure high quality cloud services and fulfill crucial Big Data needs. we propose in this paper a selection approach which includes three phases as follows: our first selection scheme, eliminates CSPs that cannot support the QoS requirements of a Big Data job, which decreases the next selection stage search scope. Consecutively, our second selection stage extends the Analytic Hierarchy Process (AHP) approach to provide selection based on ranking cloud services using various attributes such as Big Data job characteristics, Big Data task profile (BDTP), Quality of Service and considering the continuous changes in cloud services and resources.

The third phase consists of selecting cloud services among different cloud providers, this happens mainly if none of the cloud providers can support the BDTP solely. In addition, if the Big Data job is possibly split into smaller jobs, during the three selection phases, our approach maps the upper quality requirements of the Big Data job to lower level matching quality characteristics of cloud services.

2. Related Work

Cloud service selection attracted the attention of researchers because of its crucial role in satisfying both the users' and providers' objectives having high quality service while optimizing resource allocation and costs. They proposed various approaches to handle and manage the cloud service selection problem. In this section we outline and classify these approaches and emphasize on their strengths and weaknesses.

A broker-based system is described in [4] where the authors proposed a multi-attribute negotiation to select services for the cloud consumer. The quality data is collected during predefined intervals and analyzed to detect any quality degradation, thus allowing the service provider to allocate additional resources if needed to satisfy the SLA requirements. Another broker-based framework was proposed to monitor SLAs of federated clouds [5] with monitored quality attributes measured periodically and checked against defined thresholds. Additionally, in [6], the authors proposed a centralized broker with a single portal for cloud services, CSP, and cloud service users. The authors in [7] proposed a distributed service composition framework for mobile applications. The framework is adaptive, context-aware and considers user's QoS preferences. However, this framework is not suitable for for cloud service selection due to heterogeneity and dynamicity nature of the cloud environments.

The authors in [8] proposed a broker–based cloud service selection framework which uses an ontology for web service semantic descriptions named OWL-S [9]. In this framework, services are ranked based on a defined scoring methodology. First, the services are described using logic-based rules expressing complex constraints to be matched to a group of broker services. Another service selection system was proposed in [10] where the authors proposed a declarative ontology-based recommendation system called 'CloudRecommender' that maps the user requirements and service configuration. The objective of the system is to automate the service selection process, and a prototype was tested with real-world cloud providers Amazon, Azure, and GoGrid, which demonstrated the feasibility of the system.

In [11], a declarative web service composition system using tools to build state charts, data conversion rules, and provider selection policies was proposed. The system also facilitates translation of specifications to XML files to allow de-centralized service composition using peer-to-peer inter-connected software components. In addition, the authors in [12] proposed a storage service selection system based on an XML schema to describe the capabilities, such as features and performance.

Optimizing the performance is a significant issue in Cloud Computing environments. In other words, better resource consumption and enhanced application performance will be achieved when embracing the appropriate optimization techniques [13]. For example, minimizing the cost or maximizing one or more performance quality attributes. In [14], a formal model was proposed for cloud service selection where the objective is to not only the cost but also the risks (e.g., cost of coordination and cost of maintenance). In this evaluation, the model studies different cost factors, such as coordination, IT service, maintenance, and risk taking. Furthermore, the risks are denoted in terms of integrity, confidentiality, and availability.

The authors in [15] proposed a QoS-aware cloud service selection to provide SaaS developers with the optimized set of

composed services to attend multiple users having different QoS level requirements. They used cost, response time, availability, and throughput as different QoS attributes. The ranking of services is evaluated using integer programming, skyline, and a greedy algorithm providing a near-optimal solution.

Different optimization techniques were adopted for cloud service selection in the literature. One of which were proposed in [16], which used a probabilistic and Bayesian network model. The authors modeled the discovery of cloud service as a directed acyclic graph DAG to represent the various entities in the system. In [18], the authors model cloud service selection as a multi-objective p-median problem according to pre-defined optimization objectives. Their objectives are to optimize the QoS, the number of provisioned services, the service costs, and network transmission costs simultaneously in the given continuous periods. The model also supports the dynamic changing users' requirements over time. Similarly in [17], the authors suggested a service selection model based on combining fuzzy-set multiple attribute decision making and VIKOR. Nevertheless, the discrepancies among user requirements and the providers were not addressed.

The authors in [19] incorporated the IaaS, PaaS, and SaaS service subjective quality attributes based on user preference and applied fuzzy rules based on training samples for evaluation of cloud services quality. A resource management framework is proposed in [20] using a feedback fuzzy logic controller for QoS-based resource management to dynamically adapt to workload needs and abide by SLA constraints. Also, fuzzy logic was adopted in [21] to allow for a qualitative specification of elasticity rules in cloud-based software for autonomic resource provisioning during application execution. A CSP ranking model was proposed in [22] based on user experience, and service quality using an intuitionistic fuzzy group decision making for both quantifiable and non-quantifiable quality attributes to help users select the best CSP conferring to their requirements.

Another cloud service recommendation system was presented in [23] with a selection based on similarity and clustering according to user QoS requirements for SaaS, including cost, response time, availability, and throughput. The users are clustered according to their QoS requirements and are ranked based on multiple aggregation QoS utility functions. Their approach is composed of different phases, starting with clustering the customers and identifying the QoS features, then mapping them onto the QoS space of services, clustering the services, ranking them, and finally finding the solution of service composition using Mixed Integer Programming technology.

Additionally, Multiple Criteria Decision Making (MCDM) models and fuzzy synthetic decision were commonly used in combination for service selection. In [24], fuzzy synthetic decision was applied for selecting cloud providers taking into consideration user requirements. Furthermore, the authors in [25] adopted fuzzy-set theory to evaluate cloud providers trust based on quality attributes related to IaaS. Also in [26], the authors proposed a framework for QoS attributes-based cloud service ranking by applying AHP techniques. A case study was presented to evaluate their framework. Yet, this work was limited to using the measurable QoS attributes of CSMIC rather than including the non-measurable QoS criteria as well [17]. Other works used AHP approach for coud service selection, such as in [1], where the authors adopted MCDM method using AHP to select CPs based on real-time IaaS quality of service. Similarly, The authors in [27]

distributed cloud resource management based on SLA and QoS attributes. They adopted AHP to cope with the cloud environment changes during the resource selection process. However, both works exhibit the limitation of only considering the QoS of the cloud services as their selection basis.

Web services frequently undergo dynamic changes in the environment such as overloaded resources. Hence, the authors in [28] proposed a multi-dimensional model, named AgFlow, for component services selection according to QoS requirements of price, availability, reliability, and reputation. The model optimizes the composite service QoS required by the user and revises the execution plan conforming with resource performance dynamic changes. The authors in [29] proposed an SLA renegotiation mechanism to support and maintain QoS requirements in cloud-based systems. They use historical monitoring information including service statuses such as availability, performance, and scalability to predict SLA violations.

Few existing cloud federation projects are based on brokering technologies for multi-cloud composed services. Hence, more research needs to be done towards a standardized methodology for handling interoperability and standard interfaces of interconnected clouds [30]. Trustworthiness evaluation models among different cloud providers were proposed and focus on a fully distributed reputation-based trust framework for federated Cloud Computing entities in cloud federation. In this model, trust values are distributed at each cloud allowing them to make service selection independently [31]. Trust modeling was also tackled in federated and interconnected cloud environments where both consumers and different cloud providers need to trust each other to cooperate [32].

The literature is missing a comprehensive selection model that incorporates all cloud service layers, dimensions, and components in a multi-dimensional model that satisfies service selection for such constrained Big Data applications. Additionally, among the several methods used to determine the user's QoS preference, none exhibits the flexibility to make it responsive to the user's point-of-view as well as comprehends the specific characteristics related to Big Data applications. Accordingly, service selection models are to take into consideration the subsequent requirements: 1) Transparency for stakeholders (such as, customers, CPs, and service brokers), 2) Simple interface that is user friendly, easy to, configure, control and integrate 3) Maintainable and self-adapting to service layers, such as, SaaS, IaaS, and PaaS, and 4) Require low communication overhead by using low number and lightweight messages between stakeholders.

We aim in this work to build a complete, flexible, and QoS driven solution to assess different CSPs' services' capabilities of handling various Big Data tasks. Hence, we develop a three-phase cloud service selection scheme that considers the task complexity and the dynamicity of cloud resource and services. The first step in the selection process consists of apprehending required Big Data quality of service, define and endorse these requirements using the proposed Big Data Task Profile (BDTP). It adopts three selection phases to assess in real-time the CPs QoS and their corresponding services and choose only those that match these requirements.

3. Big Data Task Profile

We explain in this section the main elements of our Big Data specification model as depicted in Figure 1. For every different Big Data task, we model the related profile categories. Additionally, we model a set of attributes and characteristics classifications for
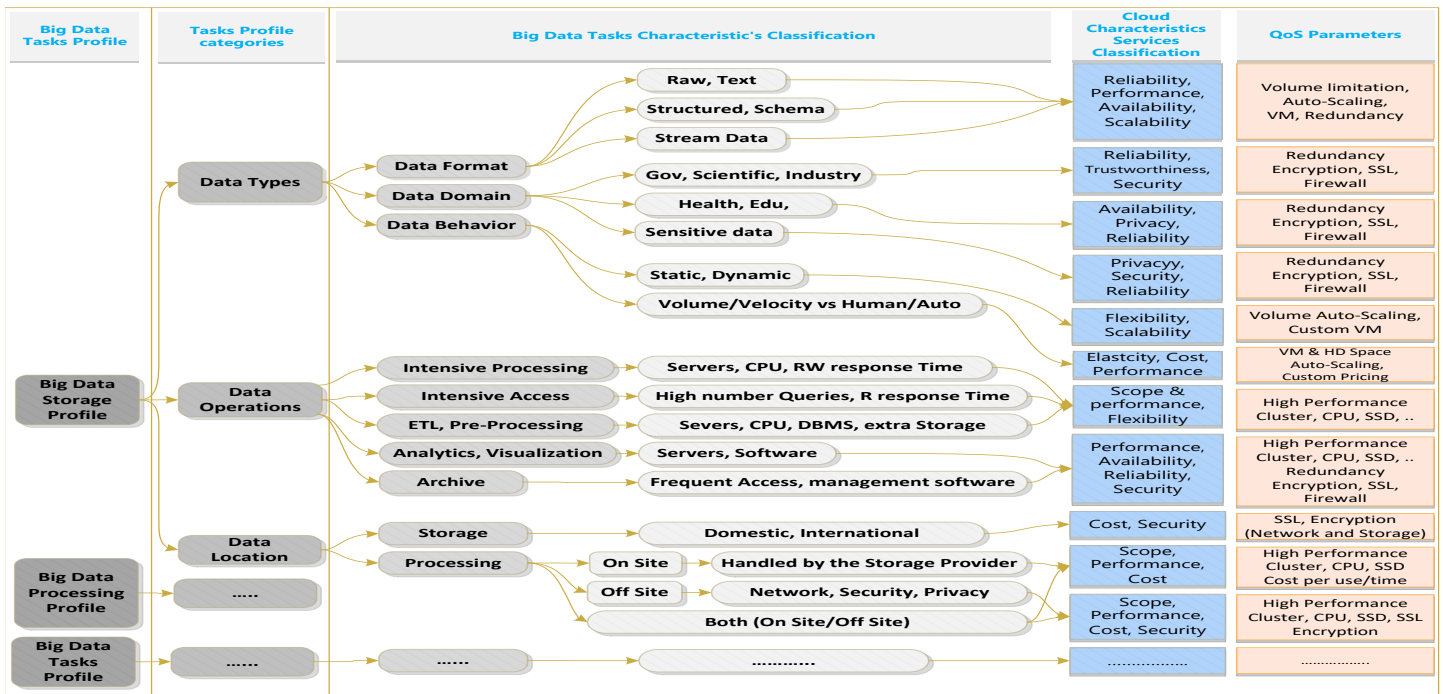
Figure 1.   Profile-based Specification Model for BD Tasks

each category. Furthermore, we map the Big Data characteristics to its corresponding cloud attribute and services.

3.1. Big Data Task Profile (BDTP) Specification

The BDTP specifies the main Big Data task requirements that need to be satisfied, and it is modeled as a set of triples: $R= \{DT, DO, DL\}$; where, DT refers to Data Type, and DO refers to Data Operation, and DL refers to Data Location. A Big Data request profiled based on BDTP, which defines the requirements and the most appropriate quality specifications that meet a certain Big Data task (such as, Big Data storage). For instance, Storage Profile specifies the following requirements:

Data Types Specifications:

a)   Format: structured, unstructured, semi-structured, or stream data.
b)   Domain: government, health, smart cities, etc.
c)   Behavior: volume static vs dynamic scale, and velocity.

Stored Data Possible Operations:

a)   Intensive processing
b)   Intensive access
c)   Extract Transform Load
d)   Analytics and visualization.
e)   Archive and backups only

Data Storage Location:

a)   Storage Preference
   • Local cloud service provider
   • Geographically disperse site: this involves considering the following properties: network bandwidth, and security of data.
b)   Data processing location:
   • On site: security and cost requirements (high or low).

• Off site: network, security, cost, and servers requirements

Table 1: Sample Profiles for Big Data Tasks

TaBig Data Tasks	Related Cloud Services	Needed Resources	Cloud Services Classification	QoS Parameters
Generation and Collection	PaaS, DaaS	A, C	d, c, e	1, 2
Preprocessing	LaaS	A, B, C, D	a, c, e, f	1, 2, 3, 4
Processing	PaaS, SaaS	A, B, C, D	a, c, e, f	1, 2, 3, 4
Analytics	SaaS	C, D	i, f	1, 3, 4
Visualization	SaaS	C, D	i, f	1, 3, 4
Storage	DaaS	A, C	a, b, c, d, e	1, 2
Transport	LaaS	A	a, b, c, d, e	2
A. Networks B. Servers C. Storage D. Applications E. Infrastructure	1. Storage RW Speed (SSD/HDD) 2. Network Speed Mb/s & Latency 3. CPU Speed, Core, Count 4. RAM size		a. Performance b. Security c. Reliability d. Availability e. Scalability f. Transformation g. Hetriogeneity h. Privacy i. Governance	

Figure 2 illustrates the events issuing succession that deal with a Big Data request. Once a request is received, the best suitable BDTP is selected from the stored profile, in addition, the requirement is normalized to generate a profile R. Then the profile is linked with the user's quality of service requirement to produce an updated profile R' which will assist in the 3-phase selection. In the first selection stage we generate a list of CSPs **CPList** that is used for the second selection phase to generate another list of cloud services **CSList**.

The tipples $R= \{DT, DO, DL\}$ represent the BDTP-based user requirements after mapping the appropriate BDTP profile to be
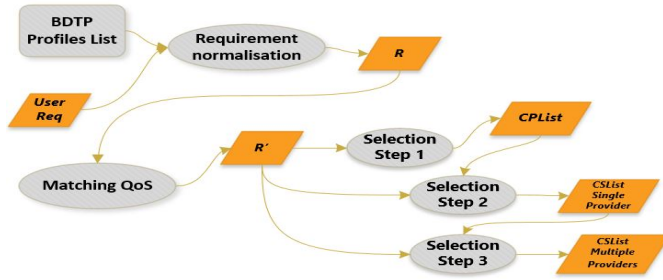
Figure 2.   Two-step CSP selection based on BDTP

Figure 3.   Workflow Example

used. Afterwards, the profiled requirement *R* are translated into **R' = {QoS₀, QoS₁,…,QoSᵢ}**, for usage in the first selection stage. The later produces a list of CSPs fulfilling **R'** and noted as: **CPList = {CP₁, CP₂,….,CPᵢ}**. Moreover, the next selection stage based on cloud services is initiated on the same **R'** to produce a set of cloud services list noted as: **CSList = {CS₁, CS₂,…., CSᵢ}**. The third selection step, retrieves the *CSList* from the second selection phase and look for other cloud services from different providers that can satisfy the request.

3.2. Big Data Workflow Profile (BDTP) Specification

In this section, we describe a simple workflow applied in a case where a patient needs to be continuously monitored to predict epileptic seizures before they actually occur. The monitoring process involves placing multi-channel wireless sensors on the patient's scalp to record EEG signals and continuously stream the sensory data to a smartphone. This process does not restrict the patient's movements. The continuous recorded sensor data, such as 1 GB of data per hour of monitoring is considered a Big Data. However, smartphones lack the capabilities to handle this Big Data, whereas Cloud Computing technologies can efficiently enable acquiring, processing, analyzing, and visualization data generated form monitoring. Figure 3 describes the epilepsy monitoring workflow, where task t_1 is the data acquisition task that is responsible for collecting the EEG data is from the scalp by sensor electrodes then transfers the signals to be preprocessed to computing environment or to temporary storage t_2, which is storing the raw EEG signals. Task t_3 performs data cleansing and filtering processes to eliminate undesirable and noisy signals. Task t_4, is the data analysis task where the EEG data is analyzed to mine meaningful information to provision diagnosis and help decision-making. Finally, t_5 is the task responsible for storing the results.

In this workflow, a task is modeled as a tuple $t\langle tn, in, out \rangle$, where, tn is the task name and in and out are the input and the output data set respectively. Task dependency is modeled in $E = \{(t_i, t_j)| t_i, t_j \in T\}$, where t_j is dependent on t_i when t_j is invoked after the t_i is completed. The data flow is modeled by tracking the task input and output states. For each task t_i, we keep information about the data parameters, type and format.

3.3. Matching the BDTP to Cloud Service QoS

As we define **R= {DT, DO, DL}** to be a triple including Data Types, Data Operations and Data Location, we map each request's parameters from high level task specification to a low-level cloud
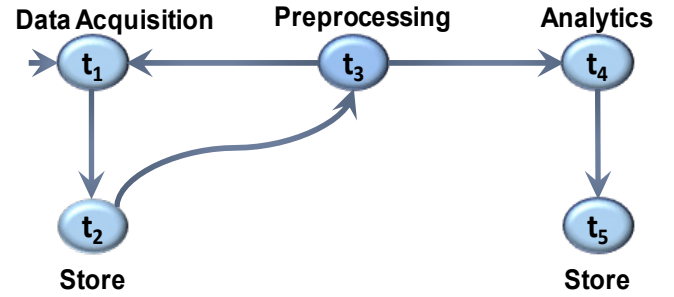
service's QoS attributes having values and ranges that satisfy each requirement of the BDTP. For each selection phase, the matching process engenders a predefined profile. The QoS Profile is continuously revised to incorporate customer's request needs even after mapping and adjustments of quality attributes. Table 1 illustrates the matching scheme of Big Data tasks to cloud services QoS attributes.

3.4. Web-based Application for Collecting of Big Data Workflow QoS Requirements

In this section, we describe a web-based application we developed for collecting Big Data workflow QoS preferences from the user and generating a quality specification profile, which will be used as basis for task and workflow quality-based trust assessment as shown in Figure 4. This GUI application, collects the quality specification that illustrates the main requirements of a Big Data workflow and its composed tasks. Some of the workflow quality requirements are application domain, data type, operations and location. Furthermore, the application collects the required quality information for every composed task in the workflow, such as quality dimension, quality attributes and the weight values required for the overall trust score calculation. In addition, output data quality is specified for each task along with the weights preferred by the user. Finally, a complete workflow quality profile is generated that enumerates the most suitable requirements and specifications, which fits each Big Data task, such as Big Data preprocessing.

4.   Cloud Service Selection Problem Formulation

One of the multi-criteria decision making methods is the Analytic Hierarchy Process (AHP) which is often used for such problems. It adopts a pairwise comparison approach that generates a preferences set mapped to different alternatives [33]. The advantage of AHP methodology is that it allows converting the subjective properties into objective measurements so they can be included in the decision-making, and hence permits the aggregation of numerical measurements and non-numerical evaluation. Additionally, it integrates the user's preference through getting the relative importance of the attributes (criteria) according to the user perception [1]. Accordingly, the quality attributes are represented as a hierarchal relationship, that matches the decision makers form of thinking [34]. Our recommended cloud service selection hierarchy is shown in Figure 5. This hierarchy clearly fits the mapping structure of Big Data to cloud services.

The AHP is intended to pairwise compare all different alternatives which are the quality attributes in our case. Therefore, the more quality attributes are considered, the larger the comparison matrix becomes and the higher number of comparison

Figure 4.   User interface for the collection of Big Data QoCS requirements

will be performed. Hence, we suggest to modify the original AHP approach as in [13].

The idea is to simplify the techniques to avoid the pairwise comparison by normalizing the quality attributes comparison matrix using geometric means which will decrease the required processing to reach a selection decision. Nevertheless, this modification will result in a converged weight matrix as a reason for adopting the geometric mean normalization and hence having a close attribute weight values. Eventually, the attribute priorities will diminish and will not satisfy the objective of this method. To solve this problem, we propose using the simple mean instead of geometric mean for normalization and calculating the attribute weights that matches the user priorities. We followed three steps in our selection approach given as:

4.1. Step1: Hierarchy Model Construction

We adopt the following definitions in our selection model [35]:

Definition 1: *The goal of decision problem which is the main objective and motivation. Here, the goal is the cloud service selection that best matches Big Data task profile conferring to the customer preference.*

Definition 2: *The alternatives which are represented with a set of various options open to the users to be considered in the decision. In our case, they are the group of available cloud services supplied by different cloud providers and matches the recommended BDTP.*

Definition 3: *The criteria of a decision problem. In this case, they are the quality attributes evaluated and upon which the*

comparison between alternatives is based on to eventually reach a decision. Specifically, they are the QoS attributes provided by the BDTP. The cloud services (alternatives) will be evaluated in comparison to the quality attributes (criteria) for measuring the matching level of the goal of the problem.

The QoS attributes (criteria) for our decision-making problem are depicted in Figure 1 where they are quantified and qualified using the BDTP by assigning acceptance threshold values or ranges of values [35].

$$CP = \{cp_i | i = 1,2,3, \dots n\} \qquad (1)$$

$$S = \{s_i | i = 1,2,3, \dots n\} \qquad (2)$$

where $\forall\ s_i \in S$ is offered by one $cp_i \in CP$

$$A = \{a_j | j = 1,2,3, \dots m\} \qquad (3)$$

$$P = \begin{bmatrix} p_{11} & \cdots & p_{n1} \\ \vdots & \ddots & \vdots \\ p_{1m} & \cdots & p_{nm} \end{bmatrix} \qquad (4)$$

where s_1, s_2 … s_n are the existing n alternative cloud services provided to the user. These services may be offered by various providers. a_1, a_2,..., a_m are the QoS attributes (criteria) from the BDTP mapped to the Big Data task required, for example: storage size, processing power, speed, availability, and reliability. p_{ij} is the performance of the ith alternative s with respect to the jth attribute.
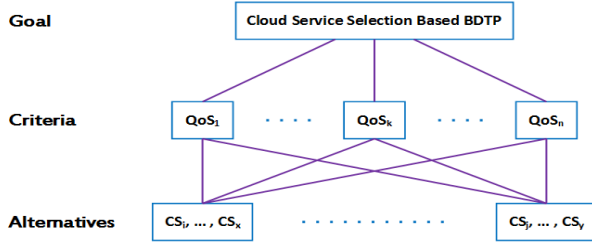
Figure 5.   Cloud service selection hierarchy model

4.2. Step2: Attributes Weights and Ranking

AHP scheme consists of mapping each property to a rank or a priority level compared to other criteria applied in different evaluations. Then, an importance level is given by a user for each property opposed to all others [35]. This is performed after building a pairwise comparison matrix using a weighbridge of level of importance. An attribute can be compared to itself and the related importance is set to 1. Therefore, the matrix diagonals are all set to 1 [34]. The importance level is within the range between 1 to 9, where 1 refers to the lowest importance attribute and 9 refers to the most important attribute having the highest value.

For *m* attributes, our pairwise comparison of attribute *i* with attribute *j* we get a square matrix $A_{M \times M}$ where r_{ij} designates the comparative importance of attribute *i* with respect to attribute *j*. This matrix has diagonal values assigned to 1. s.t. $r_{ij} = 1$ when $i = j$. Moreover, it contains reciprocal values across the diagonal, the ratio is inverted s.t. $r_{ji} = 1/r_{ij}$.

$$R = \begin{bmatrix} 1 & r_{12} & ...r_{1n} \\ 1/r_{12} & 1 & ...r_{2n} \\ \vdots & \vdots & 1 \end{bmatrix} \qquad (5)$$

Then, we define a normalized weight w_i for each attribute based on the geometric mean of the ith row. We choose the geometric

Decomposed BDTP	Match Score (%)	Storage BDTP1	Analytics BDTP2	Pre-process BDTP3
CP1	40%	√	X	X
CP2	70%	X	√	√
CP3	30%	X	√	X
CP4	30%	√	X	√

Decomposed BDTP	Match Score (%)	Storage BDTP1	Analytics BDTP2	Pre-process BDTP3
CP1	40%	0.7	0	0
CP2	70%	0	0.75	0.82
CP3	30%	0	0.8	0
CP4	30%	0.6	0	0.5

BDTP

Not 100% satisfied by one CP

Decomposable into multiple basic BDTPs

Decomposed BDTP	Selected Cloud Provider
BDTP1	CP1, 0.7
BDTP2	CP3, 0.8
BDTP3	CP2, 0.82

Multi-Cloud Provider Service Composition
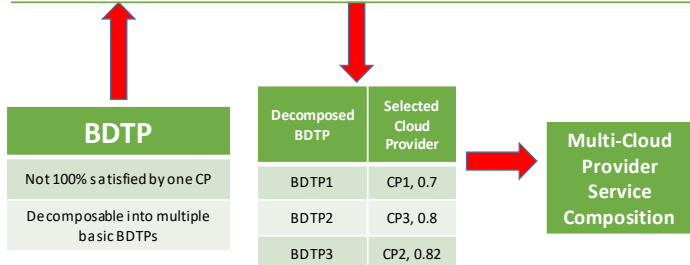
Figure 6.   Big Data workflow quality enforcement based on cloud service selection

mean methodology as an extended version of AHP for its simplicity, easiness of calculating the maximum Eigen value, and for decreasing the inconsistencies of judgment using $GM_i =$

Table 2: Multi-provider cloud service selection m-pcss

$\left(\prod_{j=1}^{m} r_{ij}\right)^{1/m}$ [34]. After that, the geometric means are normalized for all rows in the matrix using $w_i = GM_i / \sum_{i=1}^{m} GM_i$. Nevertheless, we get equal weights which disallow differentiation between attributes importance. Thus, we suggest to apply the normalized mean values for each row as follows:

$$M_i = \sum_{j=1}^{m} r_{ij} \qquad (6)$$

$$w_i = M_i / \sum_{i=1}^{m} M_i \qquad (7)$$

4.1. Step 3: Calculate the Ranking Score of All Alternatives

To generate the rating scores for each cloud service (alternative), we use Simple Additive Weighting method by multiplying weights obtained from eq. 7 w_j of each attribute *j* with its corresponding performance value in Matrix *P* from eq. 4. Then summing all resulted values as in:

$$Score_i = \sum_{j=1}^{m} w_j \times (m_{ij})_{normal} \qquad (8)$$

Where $(m_{ij})_{normal}$ is the normalized value of m_{ij} and $Score_i$ is the overall rating score of the alternative cloud service S_i. Finally, we select the cloud service (alternative) that has the highest score value:

$$S_{bestscore} = \max_{1 \le i \le n} Score_i \qquad (9)$$

5.   Model for Cloud Service Selection

We here describe our cloud service selection model to fulfill the quality of Big Data workflow over federated clouds. Figure 6 overviews how various Big Data processes, including storage, processing, and analytics can be provisioned with the cloud services and resources efficiently and with high quality. It details the main components involved in cloud service discovery and provisioning for Big Data value chain. Such components used for selection include service catalog, service broker, and service selector. However, components involved in cloud service provisioning in response to cloud service selection requests include resource selection, deployment, control, and monitoring.
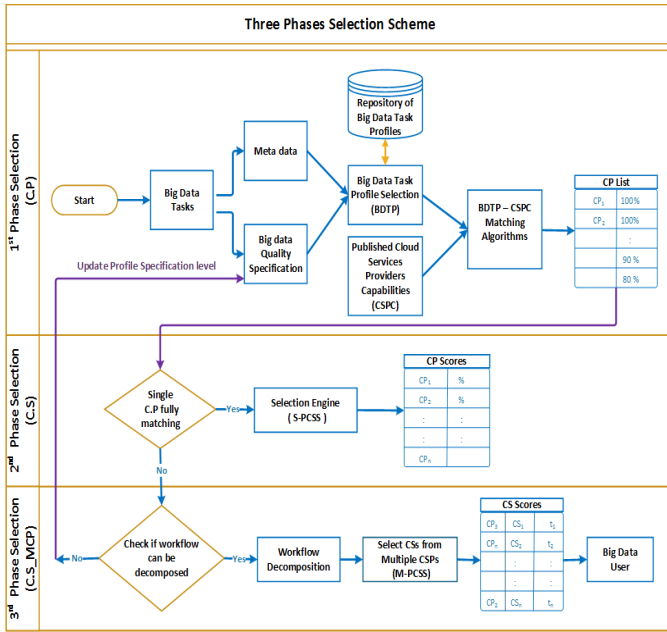
Figure 7.   Three phases cloud service selection model

5.1. Cloud Service Selection

As soon as a service request is issued to support Big Data processing and storage while guaranteeing certain QoS, cloud resources are reserved to deploy and process Big Data workflow over the cloud infrastructure. Then, the workflow execution is monitored to detect if any performance degradation occurred and respond with the appropriate adaptation actions to maintain high quality service provisioning.

Figure 7, describes the selection scheme which is implemented in three phases: the first phase involves choosing the most suitable CSPs that conform to the Big Data workflow requirements, however the second phase involves choosing among CSPs the services that fulfill the Big Data Task profile (BDTP). The third phase selection consists of conducting further selection strategy to choose services from different CSPs that satisfy different tasks of a single workflow and maximize the overall quality of the workflow. In the following, we describe in detail each of the three selection phases:

CSP selection phase: Big Data workflows described as an aggregation of tasks present a set of quality requirements, such as, trust, in addition, to extra information known as metadata, such as, type of data, and its characteristics. The Big Data task profile selection component takes as input the metadata and the Big Data quality specification to find and retrieve the closest suitable profile from the Big Data profile repository that responds to the task(s) quality requirements. Both selected profile and published cloud provider's competencies are used to trigger the execution of the CP-Profile matching algorithm which matches the BDTP profile to the CSP published competencies. A list containing scored CSPs is generated by this algorithm. A score granted to each provider refers the ratio of which the CSP is capable to accomplish the Big Data task(s) given the set of quality requirements.

CS selection phase with single provider: the second selection phase is initiated to choose the corresponding cloud services from the list of phase 1 selected CSPs according to two stages:

Stage 1: A single provider cloud service selection algorithm (S_PCSS) is performed if a specific cloud provider completely matches the QoS of the Big Data task. The output of this algorithm is a list of CSPs with their measured scores. Here, we provide an extension of the AHP Method to use a simple mean instead of geometric mean to measure the attribute weight. This leads to variation in the generated weight values for each attribute that matches the pairwise importance levels given by the user.

Stage 2: A process of decomposing Big Data workflow into tasks is triggered if no single CSP is able to fulfil the QoS of the BDTP. Tasks of the workflow should be independent and can be processed impartially. If a workflow cannot be decomposed into undependably executable tasks, a loopback to previous phase will allow reviewing the profile specification to meet the selection measures.

CS selection phase with multiple providers: the third selection phase. Once a workflow can be decomposed into a set of tasks, the multi-provider cloud service selection algorithm is implemented to cope with multiple service selection from various cloud providers to maintain the quality of aggregated workflow tasks. Table II depicts an example of BDTP decomposition into three independent profiles for storage, pre-processing, and analytics. A score is calculated for each CSP with regards to each profile and cloud providers that have the highest score are selected to handle each profile independently.

5.2. Selection Algorithms

According to the scheme described in Figure 7, we have developed three consecutive algorithms to support the three phases selection as follows:

The BDTP-CSPC algorithm: maps the BDTP with each CSP Capabilities (CSPC), for example, availability and cost. The selection is performed according to the providers' capabilities satisfaction without considering customer favoured priorities. Figure 8 describes the algorithm which requires the list of CSPs, the list of required quality attributes (profile) and the list of published quality attributes for each cloud provider. Then performs one-to-one matching of each pair of attributes (profile-published) and outputs a list of scored CSPs which completely match the BDTP. Each CSP is linked to a set of provided quality characteristics. The algorithm performs an evaluation of each CSP matching score based on the percentage of fulfilled quality attributes required by the BDTP. The BDTP-CSPC matching

---

Algorithm 1 BDTP-CSPC matching algorithm

Input: $CPList$ //List of CPs,
$profileAttrList$ //List of required quality attributes (profile),
$attrListPerCP$ //List of published quality attributes of each cloud provider

Output: $cpScoresList$ //Matching Score for each CP

1: **procedure** CPMATCHSCORE($CPList, attrListPerCP, profileAttrList$)
2:     **for all** $cp \in CPList$ **do**
3:         **for all** $attr \in profileAttrList$ **do**
4:             $cpAttrValue \leftarrow find(attr, attrListPerCP[cp])$
5:             **if** (matchRequirement ($cpAttrValue, attr$))
6:                 $count \leftarrow count + 1$
7:             $endif$
8:             $score[cpScoresList] \leftarrow count/size(profileAttrList)$
9:         **end for**
10:     **end for**
11:     **return** $cpScoresList[cp]$
12: **end procedure**=0

Figure 8.   BDTP-CSPC matching algorithm

Algorithm 2 Single Provider Cloud Service Selection (S_PCSS) algorithm

Input: *serviceList*, //List of CSs
profileAttrList, //List of required quality attributes (BDTP)
srvcAttrValue, //List of published quality attributes of each CS

Output: *srvcScoresList* //Matching Score for each CS

1: **procedure** EVALUATESERVICEAHP_S_CP(*serviceList*, *profileAttrList*, *srvcAttrValue*)
2:    $R \leftarrow buildComparativeImportanceMatrix(BDTP, profileAttrList)$
3:    $M \leftarrow R.size()$
4:    $sumMN \leftarrow 0$
5:    **for** $s_i \leftarrow 1, M$ **do** // each service (row)
6:       $MN_i \leftarrow 1$
7:       **for** $attr_j \leftarrow 1, M$ **do** // each attribute (col)
8:          $MN[s_i] \leftarrow MN[s_i] + R[s_i][attr_j]$
9:       **end for**
10:      $sumMN = sumMN + MN[s_i]$
11:   **end for**
12:   **for** $i \leftarrow 1, M$ **do** // each service (row)
13:      $weights[i] = MN[i]/sumMN$ // weight for each quality attribute
14:   **end for**
15:   **for** $i \leftarrow 1, serviceList.size()$ **do**
16:      $score_i \leftarrow 0$
17:      **for** $j \leftarrow 1, srvcAttrValue.size()$ **do**
18:         $score_i \leftarrow score_i + srvcAttrValue[i][j] * weights[j]$
19:      **end for**
20:      $srvcScoresList[i] \leftarrow score_i$
21:   **end for**
22: **end procedure**=0

Figure 9. Cloud Service Slection Algorithm - Single Provider

algorithm is scalable with the proliferation of cloud providers and quality attributes considered.

The S_PCSS algorithm: handles the second stage selection mechanism that considers thorough information about the attributes described in the BDTP to provide ranking values of the cloud services offered by the selected CSPs by the BDTP-CSPC algorithm. We adopted AHP and MADM to implement our selection strategy of cloud services. Figure 9 explains the single selection algorithm that uses a list of cloud services, the list of required quality attributes (BDTP), and the list of published quality attributes for each cloud service. Then, it generates a comparative matrix identifying the priority level of each published quality attribute in comparison to other quality attributes existing in the

Algorithm 3 Multi-Provider Cloud Service Selection (M_PCSS) algorithm

Input: *cpList*, //List of CPs
serviceList, //List of CSs
profileAttrList, //List of required quality attributes (BDTP)
srvcAttrValue, //List of published quality attributes of each CS
cpScoresList, //List of CPs scores

Output: *srvcScoresList* //Matching Score for each CS

1: **procedure** EVALUATESERVICEAHP_M_CP(*cpList*, *serviceList*, *profileAttrList*, *srvcAttrValue*, *cpScoresList*)
2:   **for** $cp_i \leftarrow 1, N$ **do** // each cloud service provider
3:      $cpServiceScore[cp_i] \leftarrow EvaluateServiceAHP\_S\_CP(serviceList, profileAttrList, srvcAttrValue[cp_i]))$
4:      **for** $attr_j \leftarrow 1, M$ **do**
5:         $cpServiceScore[cp_i] \leftarrow cpServiceScore[cp_i] \times cpScoresList[cp_i]$
6:      **end for**
7:   **end for**
8:   **for** $attr_j \leftarrow 1, M$ **do** // each attribute (col)
9:      $maxAttr \leftarrow 1$
10:      **for** $cp_i \leftarrow 2, N$ **do** // each cp (row)
11:         **if** $(cpServiceScore[cp_i][attr_j] \geq cpServiceScore[maxAttr][attr_j])$
12:            $maxAttr \leftarrow cp_i$
13:      **end for**
14:      $srvcScoresList[attr_j] \leftarrow tuple(maxAttr, cpServiceScore[maxAttr][attr_j])$
15:   **end for**
16:   **return** *srvcScoresList*
17: **end procedure**=0

Figure 10. Cloud Service Slection Algorithm - Multi- Provider

BDTP. Afterwards, this matrix is used to calculate and return a list of ranked cloud services with the highest scores and satisfy the Big Data task profile.

The M_PCSS algorithm: this algorithm handles the third stage selection where none of the CSPs fully supporting the Big Data workflow. In this situation, the workflow is decomposed into single independent tasks which will be processed by different cloud providers. Figure 10 describes the M_PCSS algorithm, the later takes as input the list of cloud providers, their offered cloud services and their calculated scores as well as the list of required quality attributes (BDTP), and the list of published quality attributes for each cloud service. It first applies the **S_PCSS** algorithm to receive the cloud service scores within each cloud provider. Then it finds the best matching services having the highest score among all cloud providers. Additionally, this algorithm favors the cloud provider that provides more services to minimize the communication and cost overhead due to data transfer and processing distribution. This is achieved by multiplying the cloud provider score to the service score to reach a final cloud service score.
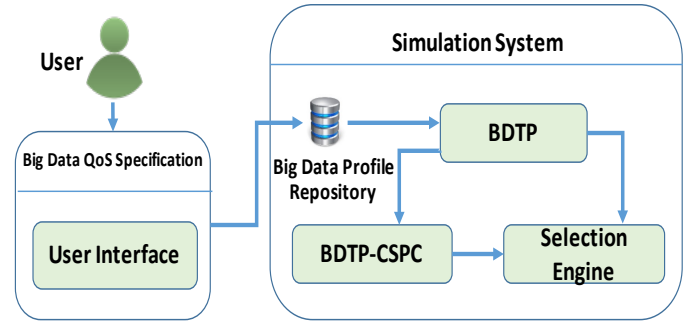
Figure 11. Simulator Components

6. Evaluation of Cloud Service Selection

This section details the experiments we conducted to assess the three-phase selection approach using various experimental scenarios.

6.1. Environment Setting

The setting and the simulation parameters we have used to conduct the experiments are described hereafter:

Setting and simulation parameters
Desktop: CPU Intel Core TM i7-3770K @ 3.40 GHz and Turbo Boost, DDR3 RAM 32GB, HD 1TB, and OS 64-bit.
Number of CSPs: 1 - 100.
Number of services provided by each CSP: 1 - 100.
QoS attributes: data size, distance, cost, response time, availability, and scalability.

6.2. Simulator

Figure 11 depicts the main modules of the JAVA simulator we have developed to implement the selection algorithms we have developed to support the three selection phases of cloud service providers and their related cloud services based on the BDTP and
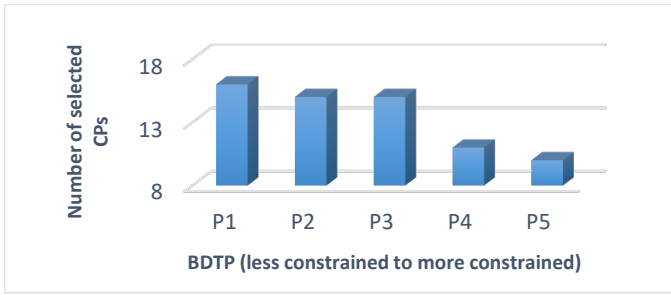
Figure 12. CSP profile-based matching with different constrains levels

Figure 13. CSP selected with different levels of BDTP strictness (Cost)

Figure 14. CSP selected with different levels of BDTP strictness (Response Time)

the AHP method. The simulator comprises five main components as follows:

BDTP component: this module classifies the Big Data task requests into three categories: data type, data operation and data location. It also sets the acceptance level (minimum, maximum, threshold), for each quality property and eventually normalizes the performance scores.

BDTP-CSPC component: integrates the full implementation of BDTP-CSPC selection algorithm we described above. This module measures a score for each cloud provider that matches the BDTP. CSPs scoring 100% are nominated to the second phase selection Engine.

Selection Engine: integrates the implementation of the S_PCSS algorithm. The later uses the BDTP and the selected CSPs nominated in the first phase, then implements AHP to rank and retrieve the set of cloud services from the list of CSPs that fulfil Big Data task. Moreover, the selection engine implements the M_PCSS selection algorithm to incorporate the implementation of selecting cloud services from different CSPs while calculating cloud services scores for each cloud provider. Afterwards, it selects the best matching cloud service with the highest score among all cloud providers.

Big Data QoS specification: it supports and guides users through an interface to specify the Big Data task quality attributes as depicted in Figure 4 above.

Big Data profile repository: serves as repository of Big Data task profiles. It is accessed to retrieve the appropriate profile when a Big Data task request is issued and a selection of suitable CSP and services need to take place to respond to the initiated request.

In addition, to the above implemented entities, the simulator generates multiple CSPs offering multiple cloud services having various QoS attributes performance levels to produce a CSP list that serves the selection algorithms. Other implemented modules include, communication interfaces, scoring schemes implementation, invocation interfaces, and storage management interfaces.

6.3. Experimental Scenarios

In this sub-section, we detail the various scenarios we have chosen to assess our 3-phase selection model and the related implemented algorithms. Scenarios were selected to validate three main properties: CSP selection accuracy, model scalability, and communication overhead.

In the following, we explain the developed scenarios to help evaluating our 3-Phase selection model.
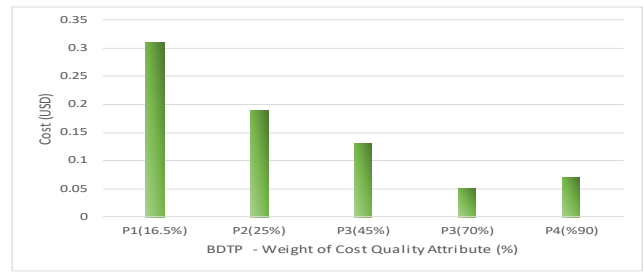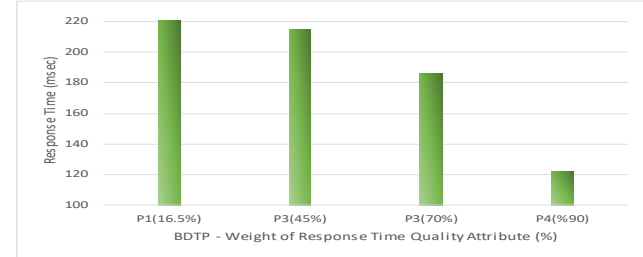
Scenario 1: evaluates the accuracy of the the first phase selection in terms of retrieving different Big Data task profiles while fixing the number of cloud providers to 20 CSPs. Figure 12 demonstrates that the less the number of selected CSPs the more the BDTP becomes constrained (e.g. includes extensive quality constraint to consider and evaluate).

Scenario 2: evaluates the accuracy of the the second phase selection based AHP while varying profiles and fixing the number of cloud providers. This will also retroactively validate the first selection results. Figure 13, demonstrates that the more constrained the BDTP is, which will add more weight on the cost quality attribute, the more the recommended CS provides a better cost. In the same manner, Figure 14, stresses the same results but
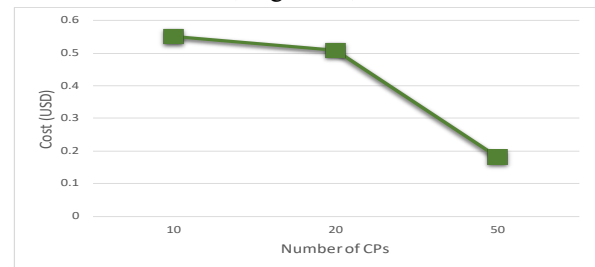
Figure 15. Cost of selected CSP with an increasing number of available CSPs

now with the response time quality attribute.
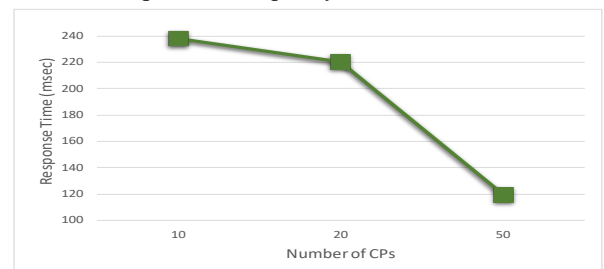
Figure 16. Response Time of selected CSP with an increasing number of available CSPs
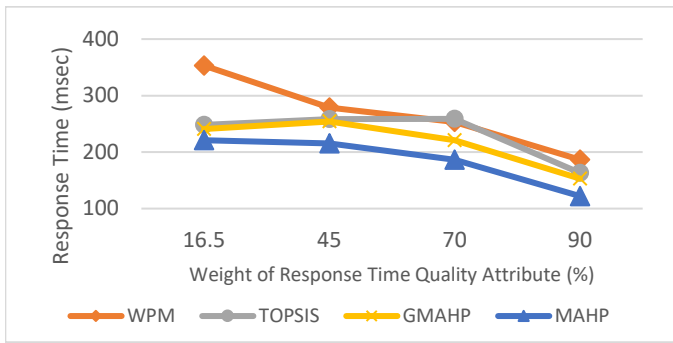
Figure 17. MAHP model behavior with different QoS attribute weight values benchmarked with other models.

Figure 18. Cost of selected task per algorithm.

Figure 19. Response time of selected task per algorithm.

Figure 20. Average response time of selected services using (MAHP) and (MAHP_M) benchmarked with other models.
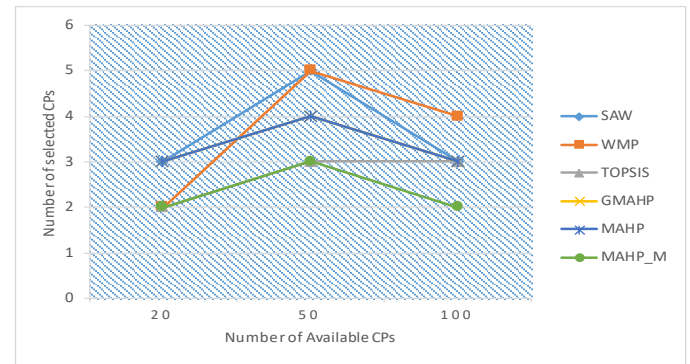
Figure 21. Number of differnet CSPs using (MAHP) and (MAHP_M) benchmarked with other models .

Solution (TOPSIS), and the modified Geometric Means AHP (GMAHP) to make cloud service selection. Comparison was conducted based on the response time quality attribute. Figure 17 demonstrates that MAHP gives better results compared to all other models, it provisions lower response times for all levels of selected quality attribute weights.

Scenario 5: we compare our 3-phase selection algorithm to other MADM selection methods by showing the cost and response time for each task composed in the workflow. As depicted in Figure 18 and Figure 19, the (MAHP) provisions lower task cost and response time respectively, and gives similar results as (GMAHP) and (TOPSIS). However, our *modified* AHP (MAHP_M) method provisions higher cost and response time per task than the (MAHP) since it gives higher preferences to selection of services from an existing cloud provider to minimize the communication and data transfer overhead.

Scenario 6: we compare the average response time of the workflow composed services using different selection schemes. The measured response time includes the communication and transfer overhead due to using services from different CSPs. As depicted in Figure 20, our (MAHP_M) method has the lowest response time because it minimizes the number of CSP among which services are selected, hence, minimizes the time wasted in communication overhead. In addition, Figure 21, shows the number of different CSPs providing the selected services and it shows that our (MAHP_M) method has the lowest number of CSPs and hence has the lowest overhead.

Scenario 3: evaluates the scalability of the selection model while increasing the number of CSPs and measuring the QoS cost and response time of the selected CSs after executing the 3-Phase selections respectively. Figure 15 and Figure 16, demonstrate that our 3-phase selection scheme scales perfectly as elucidated through a decrease in the cost and the response time respectively as the number of cloud providers increase. This is because more options are available to select among them which leads to better QoS fulfilment.

Scenario 4: we compare our 3-phase selection scheme with other MADM selection schemes. Our model used the *simplified* AHP using mean values of pairwise comparison matrix (MAHP). However, the other models used are Weighted Product Method (WPM), Technique for Order Preference by Similarity to Ideal

Figure 22. Workflow overhead and processing time of the MAHP and MAHP_M models benchmarked with other models.

Scenario 7: we compare the communication and data transfer overhead due to using different cloud providers. In this scenario, we used 100 CSPs and measured the total workflow execution time and the overhead time when using different selection methods. As shown in Figure 22, our (MAHP_M) method has the least overhead and accordingly total time amongst the rest of the methods. This is because our (MAHP_M) favors services that belong to already selected CSPs to minimize the overhead.

7. Conclusion

Big Data has emerged as a new paradigm for handling gigantic data and get valuable insights out of it. The special characteristics of Big Data reveals new requirements in terms of guaranteeing high performance and high quality of various Big Data processes (e.g. processing, storage, and analytics). The cloud infrastructure and resources are considered a perfect source of resources and services to support Big Data specific quality requirements. Selecting among myriad of cloud service providers the appropriate services and resources that meet these requirements is challenging given the diversity and the complexity of Big Data workflows.

In this paper, we proposed an efficient federated cloud service selection to support workflow Big Data requirements. It is a 3-phase selection scheme which is implemented through three phases. In the first selection phase, it captured the Big Data QoS requirements through the BDTP. However, in the second selection phase, a scored list of cloud services that satisfies the BDTP is generated. Finally, the third selection phase goes further and scored cloud services from different CSPs to better match the workflow quality requirements.

The main contributions of our selection scheme is the integration of a BDTP that ensures the QoS of Big Data tasks and is considered as a reference model for the three successive selection phases. In addition, revising the profile is advisable to have an efficient selection decision. We proposed a further contribution by extending the AHP method by adopting the mean values of pairwise comparison matrix alternative than using the geometric mean. The later shown weakness in producing a weight matrix with equal values of weights for all attributes. The last contribution is supporting workflow key requirements through the selection of multiple cloud services form multiple CSPs which maximized the Big Data complex workflow requirement fulfilment.

We conducted extensive experimentation to evaluate different properties of our 3-phase selection scheme. The results we have obtained proved that our selection model: integrated well the BDTP and guaranteed Big Data QoS requirements, scaled with the growing number of CSPs, performed better than the other MADM schemes such as TOPSIS, WPM, and the SAW, and enforced QoS requirement of Big Data workflows through varying cloud services from multiple CSPs.

For future work, we plan to have an extension for our selection scheme with more scenarios and complex Big Data workflows where other properties such as data security and privacy can also be considered. Furthermore, we are considering to assess our selection scheme against various selection techniques where we use an existing cloud environment.

References

[1] M. Zhang, R. Ranjan, M. Menzel, S. Nepal, P. Strazdins and L. Wang, "A Cloud Infrastructure Service Recommendation System for Optimizing Real-time QoS provisioning Constraints," IEEE Systems Journal, pp. 1-11, 2015. arXiv preprint arXiv:1504.01828

[2] "Cloudharmony," [Online]. Available: http://cloudharmony.com/. [Accessed 5 2017].

[3] "Cloud Computing Comparison Engine," [Online]. Available: http://www.cloudorado.com/. [Accessed 5 2017].

[4] E. Badidi, "A Cloud Service Broker for SLA-based SaaS provisioning," in International Conference on Information Society (i-Society), IEEE, 2013. https://doi.org/10.5121/ijccsa.2016.6201

[5] A. Al Falasi, M. A. Serhani and R. Dssouli, "A Model for Multi-levels SLA Monitoring in Federated," in IEEE 10th International Conference on Autonomic and Trusted Computing (UIC/ATC) Ubiquitous Intelligence and Computing, 2013. https://doi.org/10.1109/uic-atc.2013.14

[6] J. Lee, J. Kim, D.-J. Kang, N. Kim and S. Jung, "Cloud service broker portal: Main entry point for multi-cloud service providers and consumers," in IEEE 16th International Conference on Advanced Communication Technology, 2014. https://doi.org/10.1109/icact.2014.6779131

[7] Q. Z. Sheng, B. Benatallah, Z. Maamar, M. Dumas and A. H. Ngu, "Enabling personalized composition and adaptive provisioning of web services," in International Conference on Advanced Information Systems Engineering, 2004. https://doi.org/10.1007/978-3-540-25975-6_24

[8] L. D. Ngan and R. Kanagasabai, "Owl-s based semantic cloud service broker," in IEEE 19th International Conference on Web Services (ICWS), IEEE, 2012. https://doi.org/10.1109/icws.2012.103

[9] B. C. Grau, I. Horrocks, B. Motik, B. Parsia, P. Patel-Schneider and U. Sattler, "OWL 2: The next step for OWL," Web Semantics: Science, Services and Agents on the World Wide We, vol. 6, no. 4, pp. 309-322, 2008. https://doi.org/10.2139/ssrn.3199412

[10] M. Zhang, R. Ranjan, S. Nepal, M. Menzel and A. Haller, "A declarative recommender system for cloud infrastructure services selection," in International Conference on Grid Economics and Business Models, Springer, 2012. https://doi.org/10.1007/978-3-642-35194-5_8

[11] B. Benatallah, M. Dumas, Q. Z. Sheng and A. H. Ngu, "Declarative composition and peer-to-peer provisioning of dynamic web services," in 18th International Conference on Data Engineering. Proceedings., 2002. https://doi.org/10.1109/icde.2002.994738

[12] A. Ruiz-Alvarez and M. Humphrey, "An automated approach to cloud storage service selection," in the 2nd international workshop on Scientific cloud computing, ACM, 2011. https://doi.org/10.1145/1996109.1996117

[13] F. Pop and V. Cristea, "The Art of Scheduling for Big Data Science," in Big Data: Algorithms, Analytics, and Applications, Chapman and Hall, 2015, p. 105–120.

[14] B. Martens and F. Teuteberg, "Decision-making in cloud computing environments: A cost and risk based approach," Information Systems Frontiers, vol. 14, no. 4, pp. 871-893, 2012. https://doi.org/10.1007/s10796-011-9317-x

[15] Q. He, J. Han, Y. Yang, J. Grundy and H. Jin, "QoS-driven service selection for multi-tenant SaaS," in IEEE 5th international conference on Cloud Computing (Cloud), 2012. https://doi.org/10.1109/cloud.2012.125

[16] A. O. Akinwunmi, E. A. Olajubu and G. A. Aderounmu, "A trustworhty model for reliable cloud service discovery," International Journal of Computer Applications, vol. 87, no. 16, 2014. https://doi.org/10.5120/15293-3962

[17] H. M. Alabool and A. K. Mahmood, "Trust -Based Service Selection in Public Cloud Computing Using Fuzzy Modified VIKOR Method," Australian Journal of Basic and Applied Sciences, vol. 7, no. 9, pp. 211-220, 2013. ISSN 1991-8178

[18] H. K. Chen, C. Y. Lin and J. H. Chen, "A multi-objective evolutionary approach for cloud service provider selection problems with dynamic demands," in European Conference on the Applications of Evolutionary Computation, Springer. https://doi.org/10.1007/978-3-662-45523-4_68

[19] S. Telenyk, P. Bidyuk, E. Zharikov and M. Yasochka, "Assessment of cloud service provider quality metrics," in International Conference on Information and Telecommunication Technologies and Radio Electronics (UkrMiCo), IEEE, 2017. https://doi.org/10.1109/ukrmico.2017.8095422

[20] C. Anglano, M. Canonico and M. Guazzone, "FC2Q: exploiting fuzzy control in server consolidation for cloud applications with SLA constraints," Concurrency and Computation: Practice and Experience, vol. 27, no. 17, pp. 4491-4514, 2015. https://doi.org/10.1002/cpe.3410

[21] P. Jamshidi, A. Ahmad and C. Pahl, "Autonomic resource provisioning for cloud-based software," in In Proceedings of the 9th International Symposium on Software Engineering for Adaptive and Self-Managing Systems, ACM, 2014. https://doi.org/10.1145/2593929.2593940

[22] S. S. Wagle, M. Guzek and P. Bouvry, "Cloud service providers ranking based on service delivery and consumer experience," in IEEE 4th International Conference in Cloud Networking (CloudNet), IEEE, 2015. https://doi.org/10.1109/cloudnet.2015.7335308

[23] Y. Wang, Q. He and Y. Yang, "Qos-aware service recommendation for multi-tenant saas on the cloud," in EEE International Conference on Services Computing (SCC), 2015. https://doi.org/10.1109/scc.2015.33

[24] S. Wang, Z. LiuQibo , S. Zou and F. Yang, "Towards an accurate evaluation of quality of cloud service in service-oriented cloud computing," Journal of Intelligent Manufacturing, vol. 25, no. 2, p. 283–291, 2014. https://doi.org/10.1007/s10845-012-0661-6

[25] M. Alhamad, T. Dillon and E. Chang, "A trust-evaluation metric for cloud applications," International Journal of Machine Learning and Computing, vol. 1, no. 4, pp. 416-421, 2011. https://doi.org/10.7763/ijmlc.2011.v1.62

[26] S. K. Garg, S. Versteeg and R. Buyya, "A framework for ranking of cloud computing services," Future Generation Computer Systems, vol. 29, no. 4, pp. 1012-1023, 2013. https://doi.org/10.1016/j.future.2012.06.006

[27] S. Khaddaj, "Cloud computing: service provisioning and user requirements," in 11th IEEE International Symposium on Distributed Computing and Applications to Business, Engineering & Science (DCABES), 2012. https://doi.org/10.1109/dcabes.2012.76

[28] L. Zeng, B. Benatallah, A. H. Ngu, M. Dumas, J. Kalagnanam and H. Chang, "Qos-aware middleware for web services composition," IEEE Transactions on software engineering, vol. 30, no. 5, pp. 311-327, 2004. https://doi.org/10.1109/tse.2004.11

[29] A. F. M. Hani, I. V. Paputungan and M. F. Hassan, "Renegotiation in service level agreement management for a cloud-based system," ACM Computing Surveys (CSUR), vol. 47, no. 3, p. 51, 2015. https://doi.org/10.1007/978-3-642-40861-8_9

[30] A. N. Toosi, R. N. Calheiros and R. Buyya, "Interconnected cloud computing environments: Challenges, taxonomy, and survey," ACM Computing Surveys (CSUR), vol. 47, no. 1, p. 7, 2014. https://doi.org/10.1145/2593512

[31] J. Abawajy, "Determining Service Trustworthiness in Intercloud Computing Environments," in the 10th International Symposium on

Pervasive Systems, Algorithms, and Networks (ISPAN'09), 2009. https://doi.org/10.1109/i-span.2009.155

[32] D. Bernstein and D. Vij, "Intercloud security considerations," in IEEE Second International Conference on Cloud Computing Technology and Science (CloudCom), 2010. https://doi.org/10.1109/cloudcom.2010.82

[33] M. Whaiduzzaman, A. Gani, N. Badrul Anuar, M. Nazmul Haque and I. Tanzeena Haque, "Cloud service selection using multicriteria decision analysis," The Scientific World Journal, 2014. https://doi.org/10.1155/2014/459375

[34] R. V. Rao, "Improved multiple attribute decision making methods," Decision Making in Manufacturing Environment Using Graph Theory and Fuzzy Multiple Attribute Decision Making Methods, pp. 7-39, 2013. https://doi.org/10.1007/978-1-4471-4375-8_2

[35] S. Klutho, "Mathematical Decision Making: An Overview of the Analytic Hierarchy Process," 2013. [Online]. Available: www. whitman. edu/mathematics/SeniorProjectArchive/2013/Klutho. pdf. [Accessed 25 5 2017].

A Ubiquitous and Configurable Wrist-Worn Sensor Node in Hazardous Gases Detection

Mostafa Haghi[1,*], Kerstin Thurow[1], Regina Stoll[2]

[1]*Center for Life Science Automation (celisca), Rostock University, Rostock 18119, Germany*

[2]*Institute of Preventive Medicine (IPM), Rostock University, Rostock 18055, Germany*

A B S T R A C T

The monitoring of environmental parameters is a critical topic in occupational medicine. With advances in technology, during the recent years, the wearable sensors are widely used to evaluate the ambient parameters. In this paper a compact, small, light-weighted, and configurable sensor node in three physical layers for prolonged chemical ambient monitoring is described. The device is designed and demonstrated by using MLMS (Multi-Layer Multi-Sensor) concept. The device application is not only limited to hazardous gas (CO & NO_2) monitoring but physical ambient elements (air pressure, air humidity and temperature) and motion tracking are tracked, too. The device consist of integrated (ambient physical and motion tracking) and add-on sensors (hazardous gases) and enables centralized data processing. The integrated sensors are distributed on the middle physical layer on both sides (host platform). The host platform also includes micro controller and antenna for data transmission to the smartphone. The gas sensor along with the gas sensor driver board form the gas sensor node which for easy gas exposure is located at the top of the device and stuck at the top of host platform. The gas sensor driver is designed universally to be compatible with 3-lead gas sensors with low power consumption. This layer is replaceable and can be substituted with one more target gas (e.g. NO_2 instead of CO). The sensor configuration and initialization is performed via sending commands from the smartphone to the device by the wearer. In addition, the device is equipped with the notification system (beeper and vibrating motor) for early user warning in an abnormal detected. The notification driver layer is located at the bottom of the proposed device. All physical layers are stuck on top of each other appropriately through a board to board connector. The collected data from integrated and add-on sensors are transmitted in various time interval to the smartphone for screening and logging. Filter design and calibration process are also briefly described in this work.

1. Introduction

1.1. Ambient air monitoring and wearable devices

Measuring hazardous and toxic gases such as nitrogen dioxide (NO_2) and carbon monoxide (CO) in the environment (indoor/outdoor) and also meteorological parameters (atmospheric temperature, humidity and pressure) is usually expensive. Increasing number of elderly, and costly healthcare monitoring systems in hospitals from one side and high degree of penetration of wearable devices (applied to wide range of applications) in particular healthcare and tele-monitoring on the other side, attract much attention as a potential solution. This paper is an extended version of presented paper in [1]. The work is extended in terms of filter design, calibration process and providing more details. Furthermore, the device is configurable in the new version; switching between different gas sensors is possible through sending commands from the smartphone to the device.

Information on air-quality in urban environments due to cost of equipment and space is typically measured only in limited number of sites [2]. In recent years, wearable devices and technologies are being popular and demand is increasing for the application in ambient air monitoring. Especially employees who are working in chemical, biological or medical laboratories can be exposed to gases which might affect the individual performance and health.

*Mostafa Haghi, Center for life science automation, F. Barnewitz-Str. 8, 18119 Rostock, Germany, +49-3814987806 & Mostafa.Haghi@celisca.de

The form factor (wearability), enabling technology, accuracy, low cost and prolonged ambient monitoring are some of the most important criteria that can push and increase the penetration and popularity of the wearable devices in the environmental parameter monitoring. The non-linear response of the chemical sensors in ambient air monitoring, is significantly affected by the environmental conditions and the applied heating pattern for the sensor initialization [3], especially in the regime of the on-demand sensing including off-time periods [4].

Electrochemical gas sensors due to the acceptable merit and criteria are the adequate candidate to be the major player in a wearable/portable gas measurement device. These merits include the very low power consumption, high sensitivity, repeatability and mostly the electrochemical gas sensors from the wide and different range of families, approximately have a linear response under controlled conditions [5]. The linear response may extensively be useful in calibration and recalibration. Although, the gas sensors are relatively costly in personalized wearable device application, but recently it is getting closer to an affordable range, especially in mass production. In some cases, (e.g. Spec sensors) the life time of sensors are expected between 5 and 10 years.

One of the major contribution of this work is multi-parameters measurement, while the solution is still compact, small and wearable. In fact, the effort is placed to draw the attention from the 2-D design to 3-D. Therefore, z axis also is added to the calculation of prototype implementation. In this new concept known as Multi-Layer Multi-Sensor (MLMs), each physical layer is dedicated a particular function and these layers are stuck on top of each other. This would shrink the x-y plane size and slightly extend the altitude [6]. This concept deploys board to board connectors to link the layers.

The second important features of this work is the continuous monitoring for several hours without need of battery charge. This is accomplished through careful electronics circuit design and hardware selection. Where, of the critical application of ambient monitoring is in personal healthcare monitoring, thus a longer observation of ambient parameters may protect the individual health status for longer. In other word, the larger data acquisition database is, the larger information may be provided for health status analysis. However, as the tool is consist of the notification system, if the air is unhealthy, the warning signal is generated for protection.

Creating a data base and data analyzing for the health status of the individual long-term monitoring has become an aim to enable an early notification of individual health status by analyzing the large amount of collected data. The prolonged monitoring is the function of power consumption. The power mostly is dissipated by the driver, sensors and data transmission. As a consequence, the sensor selection, driver electronics circuit design and frequently data transmission to the smartphone (time interval) and sensor sampling rates, must be carefully investigated. Once the total energy consumed by the device is estimated, the battery should be selected which mostly determines overall size and weight of the sensors and therefore user's convenience [7].

The gas sensor in this wearable is the most effective indicator in the final prototype in terms of form factor, power consumption and device performance. Conventional gas detectors including the most widely used (metal oxide), relatively small Pallisters require large power consumption (hundreds of milli-Watts to Watts) with slow response time (tens of seconds). The technical specifications for gas sensors, indeed for all sensors, are challenging, and typically [8]: small and light weight, low power (e.g. less than 200 mW), long life (e.g. 5 years), high selectivity (e.g. low cross sensitivities), range of operation -50 °C to +125 °C, range of operation between 5 and 95% relative humidity.

This paper is structured as follows: In section II some related research works are introduced. In part III, we concentrate on the proposed system architecture and the main components are described. In section IV, filter design, configurability and experimental results of a possible application scenario are presented to validate the efficiency of the proposed new multi-layer architecture design. Conclusion is summarized in section V.

1.2. Related works

With the great improvement in semi-conductor and wearable technology, the criteria for implementing a qualified device in the area of ambient monitoring in wearable devices are becoming more and more sophisticated to demonstrate. Multi-tasking operation, prolonged evaluation, flexibility and user friendliness are more crucial among others. In addition, flexibility in software and hardware extension for further development without starting from the scratch design, is a point that may not be forgotten. Future developments can upgrade the device with the latest technology development and move on the edge.

This section includes two general subsections: first the latest efforts by academic community is briefly described and after in the second subsection the available industrial tools in the marker are presented in this area of interest. The features, characteristics and most critical merits for each is addressed.

Crowd sensing is one of the major methodology in the ambient air screening, a sensor node by using this technique is presented by Oletic et al. in [9]. In this approach, the individuals are able to participate in the experiment by the infrastructure of a network of mobile sensors, to measure the factors. The device is a fixed structure sensor node with the constant and predetermined sensors which is powered on by a li-ion battery. The tool consists of two electrochemical gas sensors, temperature, and relative humidity and atmospheric pressure sensors. These sensors are sampled and data are transmitted to the smartphone with Bluetooth. In [10] Ritcher and et al. introduce a wearable environmental monitoring called Eco-Mini. This prototype is designed to investigate the feasibility of wearable device in ambient air monitoring in clinical studies. In this preliminary work, the wide range of toxic/hazardous gases (ozone, sulfur

dioxide, volatile organic compounds) are evaluated in a predefined mode. Furthermore, noise, humidity, temperature, and ambient light color balance are monitored. In addition, motion tracking is performed through 3-axis accelerometer. This prototype also is equipped with location (GPS). In spite of advantaged and a wide range of monitoring, this device is suffering from some features and critical requirements. This first notable issue is the fixed structure which does not give the flexibility to the user to select the sensor according to the needful. Furthermore, Eco-mini is connected to a smartphone and stores the sampled data locally through a SD card. The wearability mode of this tool is the next concern (waist-worn) which may lead to daily routing activity user interference. Sanfilippo et al. in [11] is using Cooking hacks platform to perform an investigation. In this work the physical parameters are supported to the vital signs (blood pressure and heart rate) sensors to screen the broader range of parameters which play role in health status. The whole structure is wire based. The different sensors are communicating to the platform to transmit the data, these data are sent to the smartphone through Bluetooth Low Energy (BLE). The last device from the academic community in the first subsection is a wrist-worn watch that published by Haghi and et al. [12], [13]. This watch is introduced for monitoring several physical parameters (air pressure, temperature and humidity) as well as toxic/hazardous gases (CO, NO_2). In addition, the motion tracking through 9 degree of freedom (DoF) by the gyroscope, accelerometer and magnetometer sensors is demonstrated. Tis device, deploy 3-lead electrochemical gas sensors to detect and evaluate the target gases. A notification through a beeper and vibrating motor are generated for user attention in abnormal status detection. In addition, the prototype utilizes BLE 4.1 to transmit the data to the smartphone in a predefined time interval.

In the second subsection which the attention is paid to the available industrial tools in ambient air monitoring, a brief description, advantages and disadvantages of each is outlined. In spite of costly devices in the market, these are mostly single task and parameter detector which are only produced for a particular gas detection. This makes these tools hardly affordable for the personal monitoring. However, this is not only obstacle. This gets critical when a user intend to detect several gases in different working places. Gas alert extreme device is introduced in [14]. This well-known device is an example of such devices. Gas alert is available in the market in a medium size and waist worn mode wearability. The device is already available for NO_2 and measure the concentration of up to 5 part per million (ppm). This range is extended due to the nature of CO up to 1600 ppm. It's equipped with alert notification and vibration motor as well. The disadvantages of Gas alert is lack of data storage in an external memory nor smartphone. The data are only observable through display and in real time.

In [15] Texas instrument (TI) introduces Single-Chip tool based on MSP430. This portable/ waist-worn is a single gas detector of TI and detects carbon monoxide (CO). Due to gas calibration, humidity and temperature sensors are integrated on board. Data storage in an external memory and real time data screening is of advantages in this costly TI product. This gas detector is provided in rectangular shape and is slightly larger than gas alert extreme.

The next presented industrial gas detector in this subsection is a circular single task product of TI. This evaluator is the CO gas detector [16]. This circular device with relatively high altitude that make it more portable than wearable, is capable of being connected to both Android and iOS. The gas detector has the capability of monitoring different hazardous gases by changing the sensor. Lack of display, external memory and notification system are the disadvantages. However, the device does not support any other physical parameters.

Jain et al. introduced a wearable environmental monitoring system (WEMS) [17]. In this platform OAQ for SO_2, NO_2 and O_3, in addition to UV radiation, humidity and temperature is evaluated. It is utilized a display for data monitoring and transmit the data through BLE to a smartphone. This WEMS is wrist worn. The platform is based on MSP430 TI dual processor with M4 and maximum rate of the frequency is 168 MHz WEMS consumes 6.5 mAh in standby mode and 98 mAh in watch display mode, this power consumption in sensing mode is 250 mAh.

In this work, the infrastructure approach based on the physical layers toward design and implementation of a personalized environmental parameters monitoring is introduced. To maintain the solution, small and wrist-worn the configurability through sending command from a smartphone and sensor node replaceability as the top layer is investigated. This features enable the approach to be more flexible to be applied in different applications according to the user decision.

2. Approach, Methodology and Layers' Description

Nowadays with excellent improvement in semiconductor technology, the size of sensors is shrinking continuously. Thus, the integration of numbers of sensors of different types in the same device becomes possible. Therefore, it is expected to reduce the dimension of the tools accordingly. This is one of the reason that the wearable sensors are coming to foreground and playing a more important role in different fields of applications (e.g. healthcare, smart city, smart home, agriculture and etc.).

The presented device (hereafter called Multi-Layer Multi-Sensor Environmental Monitoring-Gas (MLMS-EMG-3.2)) is working based on MLMS (Multi-Layer Multi-Sensor) concept [12]. MLMS is applied to create a wrist-worn device by shrinking the size of X-Y and extending the solution to Z vector (XYZ-42×42×25mm). In traditional approach more often the sensors and components are located side by side across the X-Y dimension, and length and width of the board is expanded. In the presented approach in this work, the X-Y plan plus Z are involved in development. To reach a wrist-worn in the design, the width and length of the device is not supposed to be expanded. In MLMS approach, each task is dedicated to a physical layer with its own sensor(s) and components. The signaling bridge between the

different layers is created through board to board connectors. In addition to this, this connector also stuck the layers on top of each other in a firm and reliable way. The gas sensor node, host platform and notification system are the three physical layers in this approach. However, the solution might be extended to further parameters measurements and more number of layers, while still being compact, small and light-weighted. Each layer specifications and features are described in the following sections.

In a large perspective MLMS-EMG-3.2 is working as a client and sends the collected data from the ambient physical and chemical sensors to a smartphone via Bluetooth low energy (BLE) for logging. The data for each parameter also are observed on the smartphone. In the following will realize that, the communication between smartphone and the wrist-worn is bidirectional. In the large and completed system, the data are transmitted through Wi-Fi to a server for a permanent storage and further analysis (Figure 1).
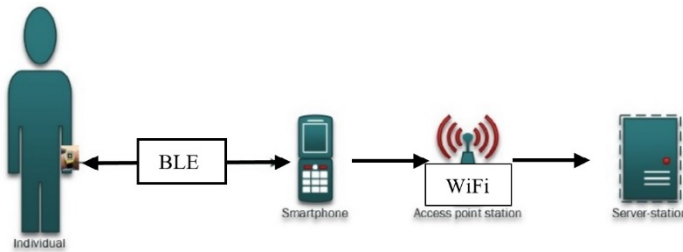
Figure 1 The general structure of whole system.

2.1. Physical Layers' Description

This approach consists of three physical layers, each dedicated to a special tasks. The layers are physically separated but are linked via board to board connectors to construct the whole device and to communicate to the micro controller and the host platform. Each layer is described as the follows:

- Middle Layer (host platform): host platform is the motherboard of MLMSEMG- 3.2. A careful host platform selection according to the requirements and criteria is mandatory to enable the integration of different sensors in this wrist worn device. The selected platform must be capable of expansion from both sides (z vector) physically. In addition, to support the number of sensors, a variety of communication protocols (ADC, SPI and IIC) is necessary. The platform should also enable the data transmission. Thus the iProtoxi host platform board based on nrf51822 (Nordic semiconductor) is the most adequate host platform for this prototype [18]. All three motion tracking sensors (gyroscope, magnetometer and accelerometer) and physical parameters measurement sensors (temperature, humidity and air pressure) are integrated on the board. These sensors are distributed on both sides of the board and allow the initialization through IIC communication. One of the extremely critical notable point is the future development of the prototype. The designers are intended to maintain the development extendable for more number of add-on sensors. This is possible with the two-side expansion in z axis through

integrated board to board connectors. On the other hand due to limited number of pins on the board, the IIC communication is chosen to be applied to the all integrated sensors for ease of use. Even the gas sensor as an add-on sensor is partially use the IIC for driver initialization and programming.

- Top Layer (gas sensor node): gas sensor node is located at the top layer stuck on top of the host platform for ease of use (replaceability) and also free exposure to air. The gas sensor node is formed of two layers: gas sensor and gas sensor driver. A hardware independent, lower power consumption and flexible sensor node is demonstrated for efficient functionality. In general the gas sensor driver is compatible with 3-lead gas sensors (with some physical and size limitation), but each sensor has its own configuration, gain and filter design. Thus, these components (gain and filter) might be added externally by the designer according to the requirements. For ease of use, the gas sensor and the gas sensor driver are soldered together to form the gas sensor node. This sensor node is the layer at the top. Where the filter design and gain of driver board is the specific parameters for each gas sensor, the general architect and structure of the gas driver is the same for gas sensors but it is customized for in these specific parameters. As the conclusion, each gas sensor is coming with its driver. Furthermore, the soldered gas sensor node, protect the sensor from frequently touch by the user and possible damage.

- Bottom Layer (notification system): The warning system is a mean to reduce the potential risk. When the monitored parameters (here toxic/hazardous gas) are exceeding the thresholds, the notification system detects an abnormal status and is activated. The notification system is consist of a driver board that activate the actuators in unhealthy status exposure when the observing parameters exceed the threshold. Beeper and vibration motor are considered as the warner means in this prototype (Figure 2). The driver is located at the bottom of prototype and both beeper and vibrating motor are connected to the board through wire. The primary idea is to target the office working of the users who are more often are working in quiet places with the beeper. The beeper is positioned in the surface of the wearable toward the user for better sound quality. The beeper is quite tangent with the device housing box. In this design in fact, micro vibration as a complementary actuator to get the attention of group of people who are working in noisy environment and my not draw the attention in risky situations. The vibrating motor is located at the bottom of the 3-d housing case and touches the bone of the forearm for haptic warning. The vibrating motor and beeper both are connected to the driver board and synchronized. The notification system is programmed in three states. In one state which all observing parameters are in the safe range, it is off and no signal is generated.

The working category of the system is two modes while depending on the degree of risk. While the threshold is exceeded but not in extremely dangerous condition, the warning signal for actuators activation is created in low frequency. This frequency

and duration of signal activation is extended in highly risky situation to attract the user attention.
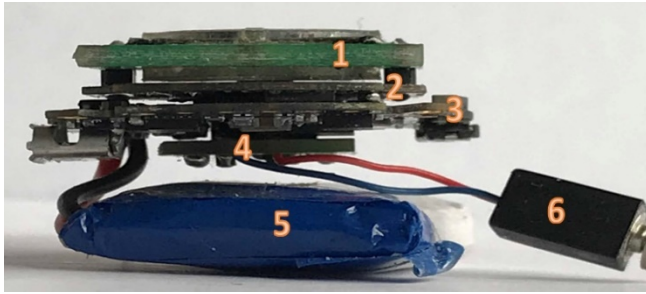
Figure 2 Platform, notification system layer, integrated and add on-board sensor of MLMS-EMG-3.2. (1)(2): gas sensor node consist of (1) sensor, and (2) gas sensor driver. (3): host platform, (4): notification driver, (5): battery, (6): vibrating motor connected to the driver.

2.2. Sensor Node's Specifications and Features

In this subsection, the most important features of the wrist-worn that provides some specific capabilities are bolded. This version of the prototype is implemented in three physical layers including one layer as the gas sensor node, host platform consists of integrated sensors and the bottom layer dedicated to the notification system driver (Figure 3, Figure 4). The MLMS-EMG-3.2 is provided as a wrist-worn in small size (42×42×25mm) and quite light (51.18 gr.). This box includes the platform, two layers at top and bottom and a LiPo / Li-Ion battery (Figure 4) in addition to beeper and vibrating motor that are located in the case's wall. Powering on/off of MLMS-EMG-3.2 is possible through a power accessible pin at the right side. In addition, the LiPo / Li-Ion battery is directly rechargeable with a USB window that is placed at the left side of the device. These specifications are limited to the physical design. The software implementation is discussed in different subsection.

- Ubiquitous device for hazardous gas detection: the flexibility, wearability, ease of use and cost of this wrist-worn, make it suitable to be applied by anyone in anywhere to detect anything (here hazardous gas). Usually, the type of hazardous gases in a specific environment are known to the user. (e.g. chemical laboratories). Many gas sensors operation at the same time (while might not be necessary), increase the power consumption and expand the size of the device by deploying number of gas sensors. One solution to reduce the difficulties to the least is to dedicate the sensor selection to the user. If the wearer can decide for the target sensor at any time and place, the efficiency of the design is improved.

gains. Whatever, the issues are pushed to the software side to be handled, the bottleneck dealing is more flexible. Consequently, the configuration is implemented in software side, filter and gain adjustment are performed manually on the hardware side. To the best of our knowledge, the best candidate which provides a complete solution is a Front End Amplifier (AFE) LMP91000 with very low power consumption. The LMP91000 is a micro-power AFE operating in the range of 2.7 to 5.25 V (in the MLMS-EMG-3.2 is used in 3.3v). The total current consumption with appropriate programming through I2C could be less than 10 μA [19]. LMP91000 amplifier converts the output current of the sensor to voltage proportionally and feeds to ADC of the microcontroller. LMP91000 is configurable for each target gas via I2C communication. In addition, a board to board connector in milli-scale also is soldered at the bottom of the sensor node. Bias $V_{ref.}$, R_{TIA}, RE are configured and unique for each gas sensor (Figure 2). Another factor for efficient performance is filter design. Depending on the sensor the filter design is varied. In general, a low pass and an integrator filter are implemented.
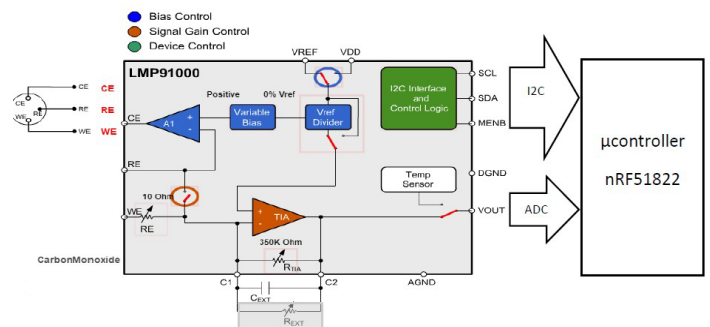
Figure 3 The internal architecture of LMP9100 and the communication with the host platform.

- Gas sensor node replaceability: to enable the sensor selection by the user, gas sensor node "replaceability" is considered as the solution. The gas sensor node that is located at the top of device through board to board connector, can be replaced by a different target gas sensor node easily. This can dramatically extend the number of gases that may be monitored (one at each time). By this feature, the device flexibility in terms of number of detected gases and wearability are improved and the power consumption is reduced. However, each gas sensor has its own configuration and requirements. The gain, filter and initialization of each sensor is customized. The number of utilized enable pins in the microcontroller are restricted due to the 24-pin board to board connector. As a conclusion, the hardware limitation is transferred from hardware side to software side for the more flexible solutions.

2.3. Device configuration and Sensor Selection

In the previous system version, the switching between the gas sensors was realized manually by the wearer. Each gas sensor was

dedicated an enable pin; the enable pin was point to the gas sensor configuration and initialization. This was demonstrated in hardware. Due to limited number of pins and future integration of other sensors (other than gas sensors), the pin utilization must be carefully performed. MLMS-EMG-3.1 version was in the scope of a one-way communication (data transmission) with a smartphone. In this version this communication is improved to bidirectional, where data are sent to the smartphone and also commands are received from the smartphone sent by the user.

the smartphone every second. In the same policy, it is started with seq. and followed by F2 and then air pressure, temperature and humidity are the parameters which are measured (Temperature also is possible to be measured through integrated sensor in humidity and pressure sensor through adequate configurations). All environmental elements are using 2 bytes divided to two higher and lower byte. Temperature data is represented also in signed 16-bit format, and the corresponding Celsius value can be computed by dividing the value with 256. Air humidity is also 16-bit value, but it is unsigned and the resulting value shows the relative humidity in percent (RH %). Air pressure uses also 16-bit representation, but the value is unsigned and the value must be divided by 32 to be converted into m bars.

Each command is sent through pressing the button on the smartphone screen. These buttons are implemented on the same screen as data observation, at the bottom.
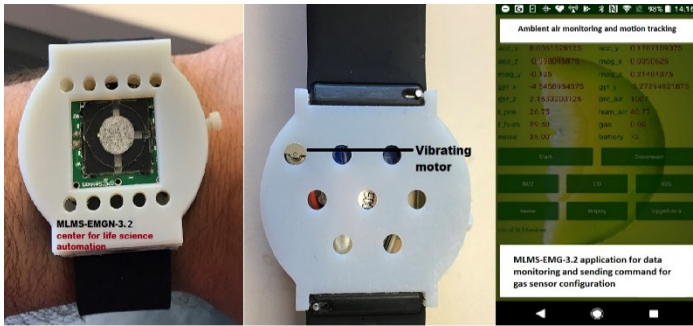
Figure 4 Top and bottom side of MLMS-EMG-3.2 and data transition to android application.

2.4. Data packet transmission to the smartphone

In MLMS-EMG-3.2 data are categorized and transmitted in 3 separate packets (Figure 5): The data that are correlated due to transient efficiency and power consumption are combined and sent in one packet. For each packet different ID's have been assigned for identification.

- F1: motion elements, 19 bytes (including 2bytes don't care): For more accuracy, this packet is sent through BLE every 20ms. A 3DoF accelerometer, 3DoF gyroscope and 3DoF magnetometer are integrated for a precision user tracking. Each sensor reserves 12 bits (1.5 byte) which are highlighted with different colors. The packet F1 is started with Seq. which is used in data number counting as well as data logging. F1 is a data ID causing identification, byte separation and recognition process in android APK. The last two bytes are reserved for further development (see Figure 6).

- F2: environmental parameters (temperature, air pressure, and humidity), 12 bytes. This packet (see Figure 13b) is sent to

- F5: Sound level, 4 bytes: The sounds packet construction is very similar to gas (see Figure 13d). However, the sound level detection is discussed in this paper and is a part of future development of the MLMS-EMG-3.2.

(c) Gas packet

(d) Sound level packet

Figure 5 Packet structure of data transmission to the smartphone.

Figure 6 Date, time stamp, sequence, ID and value of each packet

2.5. Filter design

One of the most critical and challenging parts of this work is filter design. The filter must be implemented to eliminate the
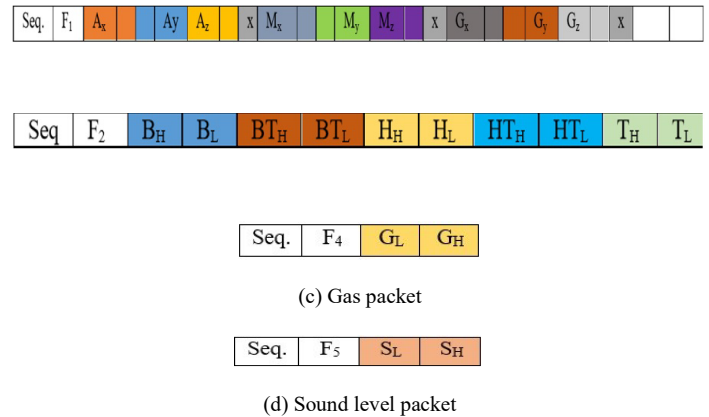
unpredicted and undesired signal noises from different sources as well as pass the real signals from the gas sensor. The filter in this project is a two-step Low Pass Filter (LPF-passive and active both are integrated in the circuit) as well as code implementation which forms a strict LPF with adequate cut-off frequency (Figure 8).

The first filter (input filter) includes a loaded resistor seen by the sensor and a capacitor (first filter including R_{1B} and C_2-passive). This resistor plays a crucial role in this circuit. It must be adjusted between 10-100 Ohm in the way that from one side it accomplishes the low pass filter design and on the other hand holds the gain in an acceptable range. R_1 is a tradeoff between response time and noise. Indeed, any noise on the WE probe is amplified by the op-amp with a gain of:

$$A = 1 + (R_{TIA\ (R2B)}/R_{1B})) \qquad (1)$$

The second filter step is carried out through a strict active LPF. The real data from the sensor passed the first LPF are fed to the amplifier which amplifies the current signal as well as filters the noises with high frequency. This filter is a feedback filter which cut-off uniquely is function of C_1 and R_{2B}. This C_1 capacitor improves the stability too. Figure 9 shows the results from first LPF and second feedback filter. The combination of both filters is simulated in Multisim simulator and results are shown in the Figure 10.

The combination of these components is working well and has been simulated in a wide range of frequency (1 Hz – 270 kHz). In higher frequencies, almost all noise signals are eliminated properly, the phase and amplitude change of the output signal vs frequency is shown in Figure 10. Each gas sensor depends on the environment parameters, sensor calibration and sensor application must be uniquely studied. It is remarkable that, first step filtering does not have to be very narrow pass band than second one. PCB design of the sensor driving board and schematic are shown in Figure 7.
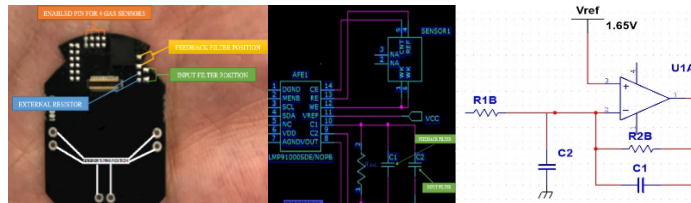
2.6. Calibration
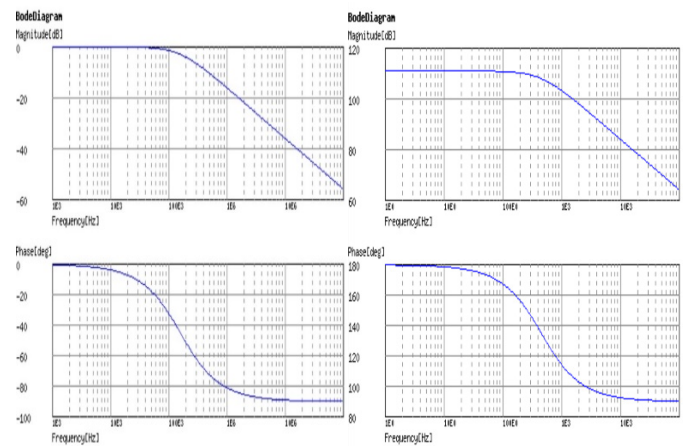
Figure 9 Results of filter application on gas sensor.

```
The target gas sensor is physically located;
The command for sensor configuration is sent to the device by the user;
The command is handled by the MLMS-EMG-3.2 and the related gas sensor node
configuration is initialized;
        VB is set;
        VB=1.65;
   Time interval of gas sensor output reading is set to 5 s;
        V output_is_read;
        While (VB-Voutput=>15 mv)
   {
   Read V output ();
        }
        Calibration process is started;
         Read temperature ();
        Read humidity ();
     Read pressure ();
Calibration is applied to the raw data;
        if (10mv <VB-V output< 15mv)
        {
        The status is detected as abnormal with degree of risk 2;
          The driver of notification system is activated;
          else if (5 mv <= VB-V output <= 10 mv)
          {
        The status is detected as abnormal with degree of risk 1;
           The driver of notification system is activated ();
          }
           The raw data are converted to ppm ();
```

Figure 10 MLMS-EMG-3.2 gas operation algorithm.

particular gas sensor, configuration and environmental parameters no more than 15 mV changes ($V_B − V_{output}$) can be achieved through gas concentration reaction with sensor membrane (otherwise it is not considered as a real signal but noise). If the V_{output} changes > 15 mV, this sample is rejected and immediately another sample is taken (Figure 11). If not, it is going through a calibration process (it's described in a separated paper). This procedure is followed by the second question on V_{output} change due to the degree of risk and two modes of notification system. As can be seen in the algorithm, the notification system configuration is strictly a function of gas concentration. At the end, raw data are

converted to ppm values. The final value of the target gas might be measured at the end of sensor stability (meanwhile in each sampling it is possible that the data reading is getting closer to the real value), but environmental evaluation status is notified to the user meanwhile (see Figure 10).

Carbon monoxide (CO) is a colorless, odorless, and tasteless toxic gas that is slightly lighter than air. The toxicity of CO is variable according to the concentration (Table 1). Long term exposition causes serious health issues. Therefore, measuring the CO concentration with a wearable device equipped with a notification system leads to an early individual notification and exposure avoidance. The focus of this work currently is on CO and NO_2 detection. The CO sensor is replaceable with NO_2 gas sensor.

Table 1 Risk of carbon monoxide in different concentrations [18].

Concentration(PPM)	Effect
0.1	Natural atmosphere level or clean air
50+	For a 50 ppm and higher CO is toxic for adults
70-75	Heart patients experience an increase in chest pain in this range. (HbCO 10%).
100	Headache, tiredness, dizziness, nausea is of signs of 100 ppm CO within 2 hrs of exposure. (Lewey & Drabkin)
200	Headache, nausea occur at this level. (NIOSH & OSHA)
400	life threatening within 3 hours
800	Healthy adults will have nausea, dizziness and convulsions within 45 minutes. Unconscious within 2 hours then death (determined in 1930)
1600	Headache, tachycardia, dizziness and nausea within 20 minutes

3.1. Experiment conditions

The experimental results presented in this paper are demonstrated under the controlled conditions in the chemical laboratory in center for life science automation (celisca) (air humidity: 59.1%, temperature: 22.7 °C, and air pressure: 1020.3 mbar). All results are produced from the prototype described above in the same laboratory and under the same conditions. For calibration, reliability and accuracy test three different gas sensor node layer were applied with different calibration factor (CF). MLMS-EMG-3.2 and BW clip as the calibrator were frequently exposed to the target gas and depending on the gas, the results were recorded in the specific time interval. The test container was a small closed transparent one. These experiments were experienced for 200 times a regular time interval within a week.

3.2. Device configuration

CO and NO_2 are two toxic gases from different families. CO is a reducing gas. When the gas sensor is exposed to CO, the internal resistor is reduced and consequently the output voltage is increased proportionally to the gas concentration. From these statement, it is understandable that, to detect the maximum range of gas concentration and avoid saturation (wider span), the LMP91000 of the sensor driver should be configured and set to the minimum initial voltage value. The deployed gas sensor is

overloaded at 5,000 ppm. To cover this relatively large range, the LMP91000 is configured to:

$$V_B = \frac{1}{5} V_{ref} \qquad (2)$$

Where: $V_{ref} = 2.8\ v$

$V_B = 0.56\ v$, indicates a gas concentration of 0 ppm. With $\mathbf{V_{ref}}$ (external source voltage) and $\mathbf{V_B}$, the difference $V_{ref} - V_B = 2.24v$ should cover the range of 0 to 5,000 ppm in CO concentration. According to the Table I, the CO measurement at very high concentrations (>1,700 ppm) does not help in individual rescue due to instant danger. In this paper, the experiments are limited to 1,700 ppm (majority of CO detectors are limited to 300 ppm). Consequently, R_{TIA} and also variable bias are configured. In Figure 2, the current flowing through R_{TIA} is called I_{sense}. Therefore, the V_{output} is a function of:

$$V_{output} = V_B\ (0\ ppm) + R_{TIA}.I_{sense} \qquad (3)$$

The ADC is reading the V_{output} generated by the gas sensor. This value is in the range of:

$$V_B < V_{output} < V_{ref} \qquad (4)$$

Generally, in design, calculation and calibration of the gas driver sensor the polarity of I_{sense} is a significant factor in (2). For reducing gases including CO, this polarity is always positive in the presence of target gas, thus V_B is expected to increase with the gas concentration. The opposite scenario is the case for oxidizing gases (e.g. NO_2). The sign of I_{sense} is negative and the respected generated voltage is reduced from the V_B. Consequently, V_{output} is proportionally smaller at higher concentrations. MLMS-EMG-3.2 has been exposed to different gas concentrations from 0 to 1,620 ppm. Within this range, performance, response time, recovery time and accuracy of the device are evaluated. The resolution of CO is categorized into two sections: 7.5 ppm (for CO concentrations < 50 ppm) and 5 ppm (for CO concentrations up to 1,620 ppm). All configurations, setting values for two groups of gas sensors, calibration process and conversion of each gas sensors are embedded in source code. Initialization for each target gas sensor is realized by sending the related command from the smartphone. During the experiments, several significant observations were perceived which are summarized as the follows:

- At high concentrations, when the different volume of gas is released into the container, the experiment for the gas with higher concentration is reaching to the some specific value in a faster rate rather than a test with lower concentration and volume of gas (both sensors reach to this point at the different time slopes). To differentiate between experiments, the time interval for the recording of the concentration must be at the same time interval, but before saturation state. In observation sections figures, number of portions are seen in each figure with the slop rate of response time or recovery time within the test (Figure 6, Figure 11, and Figure 13).
- During the measurement two status were observed in data recording: stable state and temporary unstable state. Stable state is a safe level that the recorded values are constant and reliable. In spite of stable state, the second state is consist of

gas values which are quickly recorded and disappeared. In fact, in this status a swinging values of measurement is recorded. Approximately every 50 ppm, the stable state was observed. This is experienced more often in higher concentration due to high volume of released target gas. In a normal exposure of target gas in low concentration, the response time might be as small as 15 second or less, while when the gas sensor is exposed to high concentration, the gas sensor does not have enough time for stability, therefore, the dynamic value updating in the form of swinging values are monitored. During the experiment the response time of 47 s was recorded for some tests, while this time might be more for the lower released gas volume. Halt time is increased with time and gas volume intensity and only occurs within stable points (Figure 12).
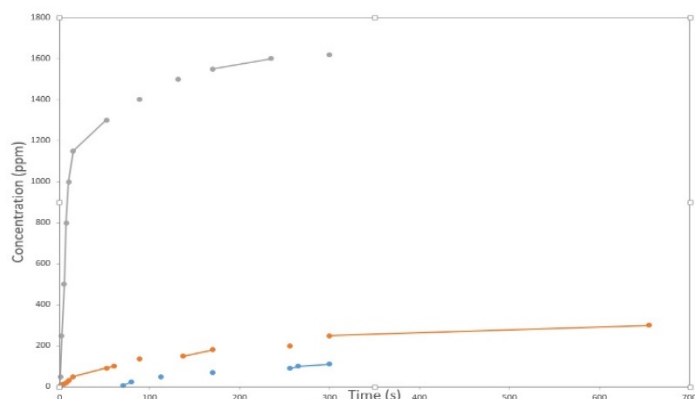
Figure 11 Response time for 110, 300 and 1620 ppm (CO).
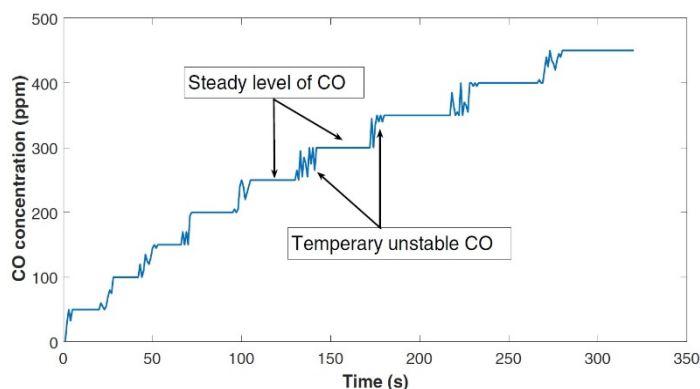
Figure 12 Stable level and halt time for 1,620 ppm (CO).

- The rate of response time is reduced gradually with the time (Figure. 8). The fastest response time is seen at the beginning of each experiment for the first measurement test. The rate of flowing gas into the container is constant for all experiments. To increase the concentration of the gas in experiments, the

- "Recovery time" for a gas sensor is defined as the essential required time for the sensor to reach from a non-zero ppm to zero ppm and is ready for the next experiment. The observation of several experiments in the laboratory, indicate after removing the sensor node from the gas exposure, for small period of time (seconds) the measured gas value is seen the same with the last change. Then a quick and large value drop is occurred. This drop, if followed with a sharper reduction of recorded values. The evaluation is continued to reach zero ppm but with slower rate than the previous stage. This pattern was repeated for three different gas concentrations (several experiments for each gas concentration), this is depicted in Figure 13.
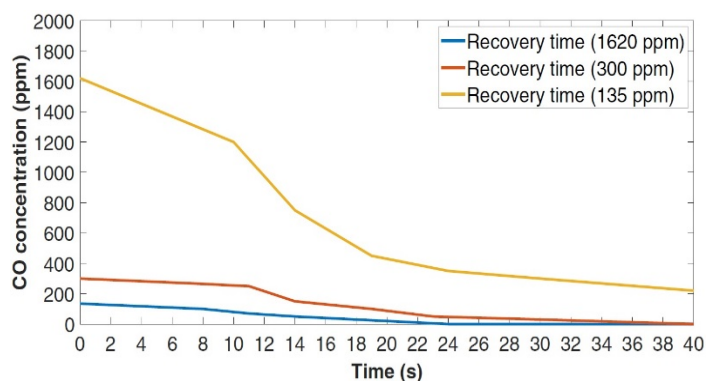
Figure 13 Recovery time for 135,300 and 1620 ppm (CO).

- Table 2, exhibits a limited number of recorded samples during the experiments under the different gas concentrations. These data are validated in conjunction well-known BW CLIP [21]. During the experiments MLMS-EMG-3.2 and BW CLIP were located into a container under the quite the same conditions. The maximum recorded error in the experiments is 6.5 %. Care must be taken that, the accuracy of the measurements are strongly the function of the "when the data are read". The authors would recommend for the average time of five minutes for the data assessment. The mean time of each test in our experiences were five minutes, however this time in some individual tests could be extended up to ten minutes. In the majority of the cases, the recorded gas concentration reaches to 75-80 % of the final value at the first five minutes.

Table 2 Experiment results of calibrator and MLMS-EMG-3.2.

MLMS-EMG-3.2 (ppm)	45	67.5	105	115	150	302
Calibrator (ppm)	42	65	111	126	152	300

4. Conclusion

In this work a compact, small and light weighted wrist-worn device for hazardous gas measurement was introduced. This prototype is based on physical layer design, named MLMS approach. Host platform including the physical ambient air sensors, and motion tracking sensors, is the middle layer. The gas sensor node consisting of the gas sensor and sensor driver is located at the top and the notification driver which is activated in abnormal status is located at the bottom of the host platform. MLMS-EMG-3.2 is designed for a large range of applications, due to particular feature that the gas sensor node is replaceable with some other target gas sensors. The target gas sensor is configured through sending commands from the smartphone by the user. By using the commands from the smartphone to activate the target sensor, the separate enable pin for each gas sensor is eliminated. This leads to a reduction the complexity and restrictions in hardware; instead a more flexible solution is presented in software side. Future developments of this prototype will include noise (sound level) detection.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

References

[1] Haghi, Mostafa, Kerstin Thurow, and Norbert Stoll. "A Multi-Tasking, Multi-Layer and Replaceable Wrist-Worn Environmental Monitoring Sensor Node." In 2018 5th International Conference on Control, Decision and Information Technologies (CoDIT), pp. 25-31. IEEE, 2018. 10.1109/CoDIT.2018.8394781

[2] E. of the European Parliament, of the Council of 21 May 2008 on ambient air quality, and cleaner air for Europe, 2008. http://ec.europa.eu/environment/archives/enlarg/handbook/air.pdf [Accessed on August 13, 2018].

[3] I. Kousuke, and J. Watson. *The Stannic Oxide Gas Sensor Principles and Applications*. CRC press, 2017. Town is missing

[4] D. Oletic, V. Jelicic, D. Antolovic, and V. Bilas, "Energy-efficient atmospheric co concentration sensing with on-demand operating mox gas sensor," in SENSORS, 2014 IEEE. IEEE, 2014, pp. 795–798. DOI: 10.1109/ICSENS.2014.6985119

[5] R. S. Istepanian, E. Jovanov, and Y. Zhang, "Guest editorial introduction to the special section on m-health: Beyond seamless mobility and global wireless health-care connectivity," IEEE Transactions on information technology in biomedicine, vol. 8, no. 4, pp. 405–414, 2004. DOI: 10.1109/ACCESS.2015.2437951

[6] M. Haghi, K. Thurow, and N. Stoll, "A multi-layer multi-sensor wearable device for physical and chemical environmental parameters monitoring (co & no$_2$)" in Information and Digital Technologies (IDT), 2017 International Conference on. IEEE, 2017, pp. 137–141. DOI: 10.1109/DT.2017.8024285

[7] M. Haghi, K. Thurow, and R. Stoll, "Wearable devices in medical internet of things: scientific research and commercially available devices," Healthcare informatics research, vol. 23, no. 1, pp. 4–15, 2017

[8] I. Stassen, N. Burtch, A. Talin, P. Falcaro, M. Allendorf, and R. Ameloot, "An updated roadmap for the integration of metal–organic frameworks with electronic devices and chemical sensors," Chemical Society Reviews, vol. 46, no. 11, pp. 3185–3241, 2017. DOI: 10.1039/C7CS00122C

[9] D. Oletic and V. Bilas, "Design of sensor node for air quality crowd sensing," in Sensors Applications Symposium (SAS), 2015 IEEE. IEEE, 2015, pp. 1–5. DOI: 10.1109/SAS.2015.7133628

[10] R. R. Fletcher, N. M. Oreskovic, and A. I. Robinson, "Design and clinical feasibility of personal wearable monitor for measurement of activity and environmental exposure," in Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE. IEEE, 2014, pp. 874–877. DOI: 10.1109/EMBC.2014.6943730

[11] F. Sanfilippo and K. Y. Pettersen, "A sensor fusion wearable health monitoring system with haptic feedback," in Innovations in Information Technology (IIT), 2015 11th International Conference on. IEEE, 2015, pp. 262– 266. http://doi.ieeecomputersociety.org/10.1109/INNOVATIONS.2015.7381551

[12] M. Haghi, K. Thurow, and N. Stoll, "A three-layer multi-sensor wearable device for physical environmental parameters and NO$_2$ monitoring," in Smart Systems and Technologies (SST),2017 International Conference on. IEEE, 2017, pp. 149–154. DOI: 10.1109/SST.2017.8188686

[13] M, Haghi, R. Stoll and K. Thurow, 2018. A Low-Cost, Standalone, and Multi-Tasking Watch for Personalized Environmental Monitoring. *IEEE transactions on biomedical circuits and systems*, (99), pp.1-11. 10.1109/TBCAS.2018.2840347

[14] http://site.jjstech.com/pdf/GAExtremeManual.pdf. [Accessed on August 13, 2018].

[15] http://www.ti.com/lit/ug/tidub41/tidub41.pdf. [Accessed on August 13, 2018].

[16] http://www.ti.com/lit/ug/snoa922/snoa922.pdf. [Accessed on August 13, 2018].

[17] H. Cho, "Personal environmental monitoring system and network platform." *Sensing Technology (ICST), 2015 9th International Conference on*. IEEE, 2015. DOI: 10.1109/ICSensT.2015.7438496

[18] https://iprotoxi.fi/ [Accessed on 10October 02, 2018].

[19] http://www.ti.com/lit/ds/symlink/lmp91000.pdf. [Accessed on August 13, 2018].

[20] https://www.spec-sensors.com/wp-content/uploads/2016/05/Carbon-Monoxide-Risks-at-Low-Levels.pd. [Accessed on August 13, 2018].

[21] http://www.honeywellanalytics.com/en/products/BW-Clip. [Accessed on August 13, 2018].

Developing Students' Motivation for Learning through Practical Problems in School

Mikhail Rodionov[*,1], Zhanna Dedovets[2]

[1]*Department of Computer and Mathematical Education, Penza State University, Russia*

[2]*Department of School of Education, The University of the West Indies (UWI), Trinidad and Tobago*

A R T I C L E I N F O	A B S T R A C T
	This paper is an extension of work originally presented in the 19th International Conference on Computer Supported Education and Information Technology. This paper identifies motivational factors that ensure the initiation and effective implementation of mathematical activity. These factors are in subordination to each other, forming a hierarchical dependence. At the heart of this hierarchy is the practical need for solving problems from real-life practice. In order to actualize this need, a number of approaches are proposed, each of which manifests itself differently at various stages of teaching mathematics at school. At the first stage, the task material is intended, in the main, only to stimulate the consideration of certain mathematical problems. It also initiates to some extent the activity procedures inherent in reality through observation and experiment. At the same time, the emphasis in teaching is on solving problems of calculation, measurement, tracing, construction, cutting, etc. At the next stage, the main emphasis is on the possibility of using a mathematical tools in the study of related disciplines. It introduces elements of mathematical modelling of real-life states and processes, which can be carried out on the basis of solving various textual problems. At the final stage, the dominant focus the vocational guidance function of practice, the mathematical knowledge for successful implementation of future professional activity. Such awareness can be provided, in particular, with the help of pseudo-real applications from the relevant professional field. In this paper the authors also provides examples supporting each stage.

1. Introduction

Successful motivation to learn mathematics cannot be achieved by a simplistic and monolithic approach, since mathematical activity is an innately multidimensional phenomenon. It requires the discriminate attention to a range of motivational factors-namely practical need, creative need, adequate and accurate language facilities, the need for proof, and aesthetic satisfaction.

These can be shown as a triangular hierarchy, mirroring the classic triangle of needs first conceptualised by A. Maslow Figure 1.

At the base of this triangle lies the practical need for solving pressing problems within the realm of day to day human activity. The solution of each of these tasks presupposes the need to generalize the empirical material accumulated in the course of each practical activity, a generalised extraction of certain characteristics common to many objects and phenomena. This is not possible without a definitive leap in the development of human thinking, which marks the emergence of a creative need for the discovery of new facts and patterns that do not belong to the sphere of immediate utility.
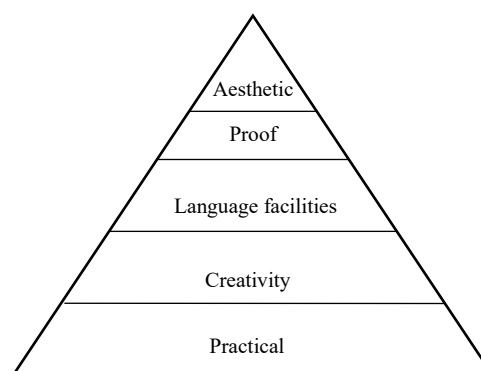
Figure 1 Hierarchical model of mathematical activity needs

[*] Mikhail Rodionov & Email: do7tor@gmail.com

The historical accumulation of schemes and decision algorithms available for an elementary study of problems necessitated the clarification of terminology and the development of more precise and, at the same time, more symbolic means of presenting these problems than could be achieved using natural language. This encouraged the logical organization of mathematical knowledge in the form of a harmonious deductive theory. At the same time, there was a reassessment of the means for checking the discovered regularities, which, in turn, led to the emergence of the need for their proof using system predetermined postulates and axioms.

Further improvement of mathematical education was achieved through, in particular, attempts to minimize the set of initial assumptions and interpretations of various mathematical theories in such a way that the derivation of new regularities would be seen as elegantly simple, unforced and economical. These attempts were a manifestation of a new aesthetic need to create the perfect expression of the hitherto ad hoc description of motivational factors.

The transition to more advanced levels of the described hierarchy does not invalidate the discrete functioning of the "lower order" needs. On the contrary, these needs are also improved and enriched by entering into productive interaction with the "higher" motivational factors.

These factors are simultaneously present in any mathematical activity. This is especially true for creative needs, which are ideally realized at the initial stage of posing the problem, and then at the subsequent stages of choosing the path of the solution and its justification. The need for proof is clearly expressed at the stage of justifying the decision, and the need for effective language facilities is needed at the stage of formalizing the initial problem. Practical need is especially evident at the stage of problem selection and at the stage of correlating the already obtained result with the limitations provided by the specifics of this problem. Finally, the aesthetic need can potentially be actualized at all these stages, even after solving the initial real situation, by seeking the possibilities of expanding derived and valid mathematical regularity.

The presence of a particular need of the researcher is only a necessary condition for the productive course of mathematical activity. What is also required is the possibility of "objectifying" and internalising this need into consciously understood motivated activity. This in turn increases practical capacity which in turn can increase conceptual understanding in a virtuous circle.

2. Practice as a source of mathematical creativity

The practical need for the realization of mathematical activity stems from the specific nature of the subject of mathematics as distinct from other scientific disciplines and is embodied in the nature of the interaction of this discipline with reality and public practice. Most scientific disciplines clearly relate to some form of the movement of matter, or to the sphere of individual and social practice, which they study with the help of a variety of methods (including mathematical ones), but, as a rule, remain within their subject area. Mathematics does not set as its immediate goal the analysis of any specific phenomena and processes, but as its raison d' etre specifically distinguishes the quantitative relations and

spatial forms inherent in all subjects and phenomena without exception, and considers them as the purpose and object of its research. Such research is always carried out on the basis of formal approaches "potentially admitting the most diverse material incarnations, and consequently, applications". In other words, mathematics can be seen as a "universally applicable scientific method", a kind of generalized "working scheme" for research, description and cognition of nature [1, 2].

This approach has always been associated by specialists in the field of the methodology of science with the general philosophical question of the reasons for the effective value and universality of mathematical theories in solving problems arising in the course of human interaction with the environment. Regular dependencies, expressed by mathematical laws, can be embodied in the intrinsic nature of the objective external world, and we only discover these dependencies through experience and experiment. At the same time, the deductive, speculative nature of mathematical knowledge can reflect its relative isolation from other spheres of human activity. In this latter case the applicability of mathematics to practice can seem bafflingly remote. The first perspective brings to the forefront the external source of the moving forces of mathematics, while the second focuses its attention on the internal needs of the development of this science and its systemic representation.

In actual mathematical activity, both tendencies are almost never presented in isolation. The practice of pure mathematics can subsequently find important practical applications, while results initially assumed to be solely applied do not in fact find any practical applications. So, for example, mathematical logic, which previously mattered only to the persons dealing with the problems of justification, have in recent decades begun to be regarded as an applied science closely connected with computational mathematics. The set-theoretical concept, traditionally perceived as the theoretical foundation of all modern mathematics, is now directly used for the analysis of phenomena of the most diverse sciences - from biology to linguistics [3].

At the same time, it is possible to give examples of so called applied theories, such as the "paint brush theory", which have no value from either a practical or a theoretical point of view. An interesting example of this nature is suggested by J. Stuart. This author tells about a man who, from general mathematical considerations, derived a very complex formula, filled with constants e, c, h for calculating the radius of the universe. And it was only after many years that scientists decided to obtain by using this formula a specific value of the radius. It turned out that it is equal to 10 centimeters [1].

R. Courant and G. Robbins emphasized that discoveries which simultaneously meet theoretical and practical needs are of particular importance for the further improvement of mathematical knowledge [4]. Thus, for example, the transition from natural to rational numbers served a theoretical need to remove restrictions on the performance of the corresponding arithmetic operations, and also the practical necessity for numbers suitable for the results of measurements of quantities.

In actual mathematical activity, external and internal stimuli, as a rule, are difficult to distinguish. Practice itself often influences mathematics, not only because it immediately requires extensive

and profound mathematical knowledge, but also because it can create insights into the unraveling of the mysteries of nature and of the properties and activities of otherwise mysterious configurations. In addition, it is the development of practice certainty, and the need for unambiguous, consistent and correct interpretation of certain natural patterns which has been one of the main motives for developing relatively uniform principles for the systematization and organization of mathematical knowledge, which in turn led to their qualitative transformation and to the formulation of the deductive method.

Let us consider in more detail the manifestation of the stimulating role of practice in relation to mathematical science. The most transparent role is manifested when developing special mathematical methods for solving specific problems that arise in peoples' real life activities. Among such problems in the early stages of the development of mathematics are the problems of land surveying, the calculation of the volume of vessels, the practical calculation, the calculation of time and the prediction of natural phenomena. Somewhat later, practical changes and developments in, for example, trade, construction and agriculture, combined with new challenges in areas such as astronomy, geography, mechanics, and optics-accelerated the predominant development of computational methods.

A general description of the realization of the relationship between mathematics and practice in similar cases presupposes a sequence of steps, as reflected in Figure 2 below.
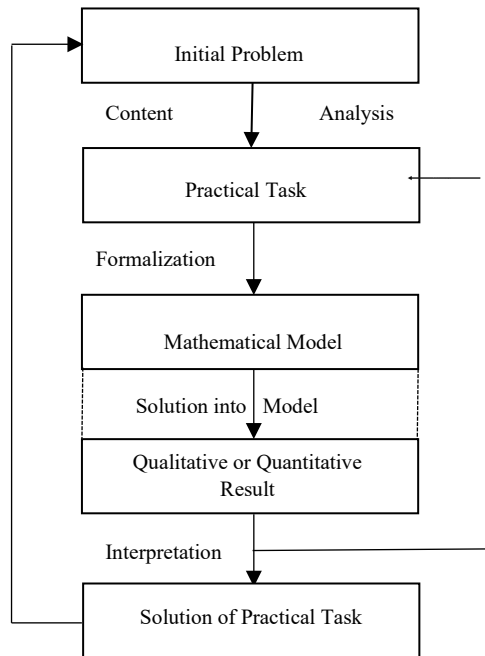
Figure 2 The realization of the relationship between mathematics and practice

In accordance with the above Figure 2, a person's activity in the investigation of one or another side of the real world begins with a meaningful analysis of the initial general problem situation, leading to the formulation of a more specific concrete practical problem. This analysis includes a refinement of the meaning of initially insufficiently defined input parameters; checking the completeness of the initial information and introducing, if necessary, missing data; choosing the range of accuracy and

numerical data corresponding to the nature of the problem, and making a preliminary assessment of the value of the problem being solved in relation to the initial problem situation [5].

At the next stage, a mathematical model of the problem is constructed, reproducing the features of the structure and properties of the original in the language of mathematical terms and symbols. Next, the model is transformed, with the aim of obtaining specific numerical values. After correlating these values with the initial situation, a conclusion is made about the plausibility of a given result and the possibility of using the method of obtaining it in solving problems of a similar nature. As an example of this possibility, one can indicate the application in electromagnetism and optics of mathematical methods originally intended for the theory of elasticity [2].

The dominant motivational factor in any considered case is the urgent need to obtain a solution of this particular practical problem irrespective of other possible areas of application of the constructed mathematical model. In such a situation, the effect of this motive essentially ends with the result. A relatively recent example of such a "direct" impact of practice on the development of mathematics is that of linear programming, which arose on the basis of a number of particular problems (optimization of material consumption, organization of transport, etc.) [3, 5, 6, 7].

It should be noted that in the course of the historical development of mathematics, the limiting scope imposed by the way a task was framed has often had a negative effect on the improvement of mathematical knowledge. For example, ancient scientists, in solving the problem of "incommensurable" values, could not overcome the Pythagorean numerical traditions and build mathematics on the basis of pure axiomatic geometry. This resulted, in the words of R. Courant, in one of the "strange wanderings in the history of science", which for two millennia delayed the "evolution of the idea of number and alphabetic calculus" [1].

A more positive effect of practice on the development of mathematics occurs when mathematical thinking goes far beyond what the posed practical task directly demands, passing successively through a series of steps of abstraction of the concepts and methods from their originally concrete and material prototypes. At the same time, the increasing abstracted purity of mathematical theories simultaneously increases their applicability, so that the range of problems under consideration is broadened and generalized, which, in turn, facilitates the transition of any given theory to a new stage of abstraction. This regularity can be represented schematically in the form of a system of "embedded" plane figures embodying the corresponding levels of abstraction of the practical problems (Figure 3).
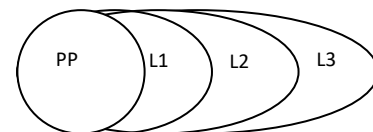
Figure 3 The levels of abstraction

We will consider this Figure 3 on the basis of one of the key directions in the development of mathematical science.

In solving the problem of the listing of any objects, each of the individual properties that are available for each of these objects can begin to play the role of a "unit". A search of these "units" leads to the establishment of a certain sequential order and the emergence of the concept of the ordinal, and then of the quantitative number. The implementation of the simplest operations on numbers convinces us of the need to extend the notion of a number, which has as its final result the creation of a single logical concept of arithmetic.

At the next level of abstraction, it is appropriate to introduce symbols, which imply any quantitative numbers. Thus, we move from arithmetic to algebra and then to mathematical analysis, which unlike algebra studies variables not only in discrete but also in continuous form [8].

The rapid development of theoretical natural science in the late 19th century meant mathematics had also to develop to a new, higher stage of abstraction. In addition to matters of quantitative symbolic content, the mathematical operations themselves are considered to be variables.

The highest level of abstraction of modern mathematics, as already noted above, does not mean its separation from practice. On the contrary, abstract mathematics, operating with such abstract concepts as a group, a set, an abstract space, has managed to explain and express the most complex processes and phenomena of 'reality'. The application of mathematics to practice is at this level generally realized directly in indirect conjunction with natural science disciplines. As an example, the practice and development of topology produced changes in differential geometry, enriched the theory of relativity and, through the theory of critical points in the calculus of variations stimulated the development of homological algebra, and, through the theory of sheaves and cohomology, algebraic geometry. Thus topology has moved from an esoteric category of mathematics to one of its central unifying and basic facets [3].

During the transition through levels of abstraction, mathematics extracts additional information hitherto implicit in its structures, while providing itself with the possibility of continuing applicability to pressing practical demands and also strengthening its internal potential. It should be noted that the initial practical need does not disappear without a trace in the course of its immediate satisfaction. It implicitly participates in the construction of a chain of internal motives for the improvement of mathematical knowledge, reinforced from time to time by direct external impulses. These impulses are due to the classical and ongoing tensions within science between the newly obtained results of experience and observations and the basic laws underlying the corresponding theoretical concept. The solution of these contradictions assumes either the search for new, still unknown methods, or the amendment of the already adopted basic provisions with the aim of increasing the accuracy of their wording through the use of second, third, etc. degrees of approximation. For example, in the early 20th century mathematicians realized that the theory of quadratic forms developed by D. Hilbert was not entirely suitable for solving the problems of rapidly developing quantum mechanics. Awareness of this fact served as a stimulus for J. von Neumann to improve this theory. By refusing the traditional "binding" of a quadratic form to a specific algebraic record,

Neumann was able to produce a more general definition which avoided the limitations of the Hilbert approach. This improved theory of quadratic forms was able to provide answers to the very real and specific demands of modern physics [3, 5].

Another possibility of the stimulating effect of practice on mathematics is the transfer of concepts, representations and modes of activity which are characteristic of natural and humanitarian disciplines, to the field of "pure" mathematical science. This possibility stems either from the absence of concepts in mathematics, yet in which language one can describe certain real processes and phenomena, or from the presence of certain discrepancies in the interpretations of "related" objects of study in mathematics and corresponding fields of human knowledge. A change in the viewpoint of mathematical methods makes it possible, in a number of cases, to stimulate the development of a corresponding section of mathematical content. For example, the use of a rigorous definition of the limiting transition in the sense of Augustin-Louis Cauchy in the study of real processes seems at first sight to be impossible, since the consideration of any physical quantities reduced "beyond some reasonable boundaries" is completely meaningless. In connection with this, in physics, the so-called "practical infinitesimal quantities" are considered, and treated as actually infinitesimal. The apparent contradiction is resolved on the basis of a scientifically grounded possibility of using the definition of a limit transition not only for infinite continuous processes, but also for particular cases on bounded sets. Despite this possibility, the practical treatment of infinitesimal quantities has found its application in the concept of nonstandard analysis that goes back to Leibniz, which in a number of cases substantially simplifies the classical exposition of it [4, 6, 9].

Among the existing "channels of influence" of practice on mathematical activity, one more should be mentioned. As we know, practice is the source of plausible reasoning based on intuition, experiment, analogy and constructive induction. These arguments, unlike the proofs, do not provide the reliability of the mathematical theory. However, in them, according to R. Courant and G. Robbins, can be found the real essence of any mathematical discovery, even if it belongs to its most abstract areas [1]. Many examples of how plausible reasoning can lead to the discovery of certain mathematical regularities are given in the well-known work of G. Polya [10]. It is important to note that this author directly relates the application of these arguments to motives that are starting points for the promotion of relevant hypotheses and their proofs.

3. The motivational role of practice for mathematical education in school

The assimilation of a mathematical theory requires consideration of empirical concepts directly related to practical activity. These concepts can also be fully understood, refined and used in practice as meaningful interpretations of abstract theoretical concepts. It is necessary in this respect to know about the stages in the assimilation of educational material. At the first stage, following observation and experiment, a basis of understanding of a fragment of the mathematical theory is formed. Then, in the course of understanding the whole system of empirical concepts and interrelations between them, knowledge ascends to the theoretical level. Finally, mathematical concepts and methods

of activity produce some concrete, meaningful interpretations that make it possible to intensify the students' desire to apply the acquired theoretical information in practice. This process must of course take into account the overall (and especially age related) context of school based mathematical education.

At the earliest stage, practical activity mainly performs a stimulating function in the study of mathematical content, resulting from the tendency inherent in the child's initial consciousness to "cling" to specific and directly "tangible" facts and situations [4, 5]. At the same time, the study of mathematics is a process of empirical cognition, in which observation and experiment (calculation, measurement, drawing, construction, etc.) play a major role. The main motivational factor here is the inherent desire of any person to connect the acquired material with their own life and practical experience.

In the following stages of training, the stimulating role of practice, although it ceases to be dominant, nevertheless retains its role as an important means of motivating consideration of a fragment of content and the sparking of initial interest. At the same time, the mathematical fact begins to act not so much as a straightforward step of generalization of empirical material realized by direct teacher guidance, but rather as a result of solving a purely mathematical problem, specially formulated during the consideration of the corresponding practical problem. The possibilities for such work are significantly increased when connecting the material of related subjects, especially physics. Thus, for example, the study of the propagation of light reflected from a specially placed mirror in its path leads to the question of choosing the path by which light travels the shortest distance. An attempt to answer this question leads to the formulation of a purely geometric problem. The solution of this problem is realized, as is well known, on the basis of the symmetry method. Thus, the physical problem appears here as a carrier of motivation for the application of the symmetry method in solving geometric problems.

In another case, observations of sunset followed by plotting the time dependence of the moment of sunset from the date of observation can be used to motivate the introduction of the trigonometric function y=sin x and to reveal some of its features, making it easier for schoolchildren to assimilate such concepts as the domain of function, monotonicity, zeros and the periodicity of the functions [11]. Here, empirical activity does not end with the very introduction of the concept of trigonometric function, but "permeates" the entire content of the topic, periodically providing additional support to the motivational mechanism involved in the initial situation. The main didactic condition for the effective implementation of such situations is the provision during the educational process of the possibility of their detection, awareness and successful resolution.

The participation of schoolchildren in the process of the emergence of new concepts by abstracting and generalizing the phenomena of the real world, significant though it is, by no means exhausts the stimulating potential of practice in the study of mathematical material. In particular, an essential role in the realization of this potential is played by an emerging awareness that knowing a particular fact for solving an important problem, whilst complete in itself, can also lead to the further development

of the problem or to proving the theorem. The proposed problem or theorem should always be linked to students' previous, empirically based experience.

For example, at the primary level students meet challenges to find the sum of certain numbers in a number sequence:

$$1+2+3+\ldots+99+100$$

To find the sum students change the order of numbers in the sequence and group them accordingly:

$$\frac{(1+100)+(2+99)+\ldots+(100+1)}{2} = \frac{101\times100}{2} = 5050$$

At the senior level, students can apply this fact using geometric material. Here students must find out the number of lines connecting a certain number of points. After some practical actions with two, three, four points, students determine how many lines can be drawn through 100 points. As a result of discussions, students establish that the 100th point can be connected with another point by 99 lines, the 99th point by 98 lines, the 98th point by 97 lines, and so on. This allows a transition from a geometrical problem to an algebraic problem. Further analysis helps to find the number of lines passing through n points. To solve this problem students need to find the sum of the first n natural numbers, i.e. finding the sum of arithmetic progression corresponding to a number sequence.

In the given example, students were stimulated to learn the progression from manipulating geometric objects to formulating a geometrical problems and then converting it into algebraic problem. The use of geometric material gives pupils a subjective feeling of novelty. At the same time it allows them to engage in practical activities.

At the next stage of mathematical preparation, the motivational role of practice is expressed in the realization of its worldview function. Such an implementation is possible through the demonstration of the application of the studied mathematical content in related courses and other school disciplines and consideration of the history of the emergence and evolution of scientific concepts and methods. It also develops familiarity with the elements of mathematical modeling of real states and processes, underlying the mastery of applied mathematical ideology [7, 11, 13]. In addition, an understanding of the role of mathematical knowledge as an important component of human culture becomes one of the leading motivational factors, creating a conscious desire by students to use the acquired material in related subjects and real life practice.

Textual problems are a traditional means of demonstrating the practical importance of mathematics as it is studied in school. In solving them, students become acquainted with the basic l method of cognition of reality via the concept of mathematical modeling of the initial real life situation. They learn to choose which model and how to construct and apply it. They thus also learn how to analyse and interpret the quantitative, graphical or qualitative results they obtain.

Textual problems form part of the very first mathematics lessons, thereby implicitly preparing junior students for a future explicit understanding of the concept of modeling. As a rule, here they act as 'pseudo-real 'problems, presupposing exactly as much

data as necessary for the solution, and providing for an unambiguous and exact answer. The process of solving such problems is, as it were, "monologic". It does not require the distinct special definitions and refinements which characterise applied problems of a tangibly external nature [12, 14]. Accordingly, the motivational effect of such tasks is mainly determined by the artificial actualization of problematic situations, often of an entertaining nature, the plot of which to some extent correlates with the real non-mathematical experience of students. The presence of such an effect implies compliance with the following minimum set of requirements.

1. The given tasks must correspond to the real characteristics of the non-mathematical objects described.

2. The task plot should be relatively "close in spirit" to the student, reflecting the significant aspects of his or her experience.

3. The set of mathematical tools mastered by schoolchildren should be sufficient to solve the problem at a level corresponding to this stage of mathematical preparation.

As student mathematical understanding develops, so the possibilities of realizing the worldview function of practice are substantially increased. This is primarily because the set of the subject tools (the types of equations studied, the functional dependencies, algebraic expressions, etc.) are substantially enriched by each of them having the potential to be understood as a mathematical model of some real phenomenon or process. The study of related disciplines (physics, chemistry, geography and others) provides the school mathematics curriculum with a significant number of practical applications and motivations. Mastery of the intrinsic conceptual qualities of mathematical modeling of real processes, studied in related fields of knowledge, provides an understanding of the general possibility of applying mathematical knowledge and skills, and thereby develops enthusiasm for such application. At the same time it becomes the basis for the formation of educationally cognitive motivation both in relation to mathematics itself and in relation to other affected disciplines [15].

It is important to identify a number of conditions for the effective implementation of this 'worldview' function of practice, which apply to all stages of the school based mathematics curriculum.

1. Maximum correlation between the mathematics curriculum and other school disciplines, with close consideration of real life practice. This correlation depends on careful planning and timing, a unity of approach to the formation of concepts common to these courses; consistency of terminology, notation, systems of units of measurement, and also the correspondence of operational structures to the solution of typical problems.

2. Purposeful development of mathematical intuition, implying the introduction of features specific to applied activity in the teaching of mathematics [15, 16]. This includes the basic skills and techniques used in solving practical problems (selection of the necessary data, estimation of the result, methods of approximate calculations, etc.), and the cohesive consistency of the steps characteristic of applied activity (analysis of the real situation, formulation of the problem, choice and construction of its

mathematical model, interpretation of the real meaning of the result).

3. Clear demonstration of the origin and development of mathematical concepts and methods because of related knowledge and real life needs and experience.

All this is best achieved within a common ethos and a culture of dialogue within each and across all parts of curriculum. This produces mutually enriching learning within and across disciplines. It is also a matter of positive and sensitive approach. Factors such as emotional tone, confidence in the cognitive abilities of the interlocutor, and mutual support are important examples. Tasks of a practical nature provide great opportunities for developing an iterative dialogue of learning because of their initial empirical uncertainty. The process brings clarity, relevance and priority. Students create a virtuous circle of learning and motivation.

As an example, let us consider a practical problem depicting a real situation associated with the construction of a cottage.

Problem 1.

The cottage has width $a = 5$ m and length $b = 8$ m. What is the size of the mansard if the distance from the attic flooring to the top of the mansard is $h = 3$ m?

After discussing the meaning of the terms related to the given situation (mansard, attic flooring, top of mansard) students find out that the length of the mansard is the length of the cottage. To find other sizes students analyze the situation and answer the following questions.

1) What geometric shapes correspond to the images of the façade of the cottage and the cross section of the mansard? The answer is a rectangle inscribed in a triangle.

2) When drawing variants of the specified geometrical configuration, when does the mansard becomes the most spacious? Check your answer by calculation. The answer is the rectangle sides are slightly different from each other.

3) How to mathematically characterize the most rational form of the cross section of a mansard? Its area is maximum.

4) What problem can you formulate using this fact? You need to find the maximum cross-section of the mansard.

The problem requirements are not imposed on students by the teacher. They are requirements which develop as an intellectual acquisition as a result of discussion.

Next, some preliminary quantitative estimation of an expected result should be carried out. The estimated value of this result should be determined. Rough approximation shows that the result is equal to the area of the cross section of the mansard which has a triangular shape (15 m^2). More precisely, this value is determined through a drawing using students' visual and intuitive reasoning. This value equals half the cross-sectional area of the mansard approximately (7.5 m^2). This process allows students to control the progress of the search.

At the next stage of solving the problem, students' constructive team work continues. They build a mathematical model. Using the

geometrical drawing students formulate a mathematical analog of the practical problem:

Find the sides of a rectangle inscribed in the biggest triangle. The triangle has base *a* and height *h*, (Figure 4). Introduce the notation: *LD=y; CE= h; DM=LG=x; NK=a*

Then students construct an analytical model of this practical problem using similarity of the triangles *NCK* and *LCD*:

$$S = \frac{a}{h}(h - x) \times x,$$

Where *S* is the area of the rectangle *LDMG*.
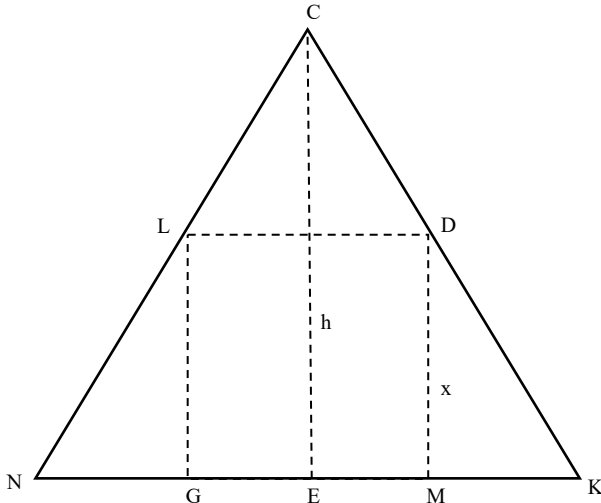
Figure 4 The mansard

An investigation of the relationship between the area of the rectangle and the second and the third factors (*a/h* constant) leads to the mathematical question: at what value of x is the product of *(h–x)x* a maximum *(x<h)?* This product can be presented in the form $-x^2 + hx$ and corresponds to the quadratic function

$$f(x) = -x^2 + hx. \tag{1}$$
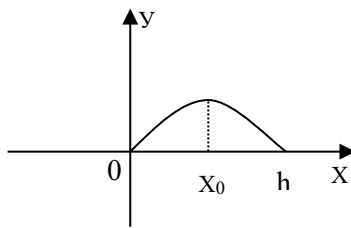
Figure 5 Graphical model

In this case only a graphical method is possible. Students construct a new graphical model of this practical problem. This model is a part of the parabola which opens downward (Figure 5). It leads to the next mathematical question. At what argument value is the function *f(x) = –x²+hx* a maximum?

Clearly the argument value corresponds to the abscissa of the parabola vertex. From the concept of symmetry or the formula this value is $x_0 = \frac{h}{2} = 1,5$

We come to different interpretations of the results.

1. An initial situation.

The maximum cross-section of the mansard is when the height of the mansard is equal to half the distance from the attic flooring to the top, i.e. *1.5 m*. The area of the cross-section is

$$S = \frac{a}{h} \times (h - x) \times x = 3,75 m^2 (S < 7.5 m^2)$$

Maximum mansard capacity is $V = S \times b = 30 m^3$

2. A geometrical model.

The side of the rectangle inscribed in the biggest triangle is the middle line of this triangle.

3. An analytical model.

The product of two positive factors has maximum value if they are equal (the sum of these factors is constant).

4. A consequence.

Among all rectangles with the same perimeter, a square has the maximum area.

5. A graphical model.

The relationship between an area of a rectangle inscribed in a triangle and its side is a part of a parabola. The abscissa at the top of the parabola corresponds to the side when an area is maximum. The abscissa of the points of intersection of the parabolas and the coordinate axes corresponds to the side when an area is minimal, etc.

As a result of this work a teachers inform the students that the practical problem considered is one from a wide range of problems where one needs to find maximum and minimum. These problems are of great practical importance. It is useful to give students homework: choose similar problems from their life experience.

This example illustrates all the major stages of solving practical problems.

Among these stages the key element is the choice of the basic model of the studied situation. Practice shows that students have challenges in choosing this model without prompted training activity organized by the teacher. For this to be effective, the teacher must help students develop particular skills. For example, students must establish the similarity between various explanations, to estimate the outcome of each of them in a specific situation, whilst evaluating different approaches. This kind of work can be organized through specially chosen sets of practical problems.

Problem 2

What kind of measuring instruments do you need to determine the area of the steel plate in the form of an equilateral triangle measuring a cm?

Most students propose to use the formula: $S = \frac{a^2\sqrt{3}}{2}$ for the area of the triangle. It is sufficient to measure the side of a triangle with a ruler. The content of the school physics course allows them to determine the area through the volume and density of the material of the steel plate: $S = \frac{V}{n} = \frac{m}{p \times h}$.

Students should weigh this steel plate and find the appropriate value of density in a table. After this activity students get identical results using different formulas and different measuring instruments. The relationship between mathematical and physical reasoning then becomes clear to them.

Students can see that those two approaches produce an approximate measurement. Then they have to explain the equality: $\frac{m}{p \times h} = \frac{a^2 \sqrt{3}}{2}$. The obtained "motivational impulse" can be used by the teacher to clarify the range of possibilities of the methods used. A teacher gives students steel plates of different forms (a circle, a rectangle, ellipse etc.) and discusses the benefit of different approaches to solve practical problems. The practical approach is optimal for any given activity. Students may see plates of various shapes. Also, they can use the weighing method to determine the characteristics of geometric shapes. This method was successfully applied by Archimedes for the volume of a sphere. During this activity students recognize how a practical method relates to a mathematical method. This method allows students to consider from a given problem how to find areas and volumes of geometrical shapes in general. As a result of this activity students will have long-term motivation in learning future mathematical analysis. It is important to note that in the above (and indeed other) practical problems mathematical activity is not a closed and separated structure; it is a natural component of the universal system of knowledge about the world [9].

4. Students' orientation in the modern world through solving practical problems

At the last stage of teaching, the role of practice in providing mathematical orientation becomes dominant. Practice gives students a new motivation and understanding of how mathematical skills are needed to fully participate in the modern world and for the successful implementation of future professional activity. However, this is not always easy. One challenge is that many modern scientific fields operate in such a way that their types of models do not smoothly integrate with the traditional mathematics curriculum. More subtle approaches might be then needed to impart an understanding of the practical application of mathematical knowledge to future professional development and in relation to environmental and other extracurricular activities.

Let us consider an example from the field of medicine. In studying the exponential function in the school, students' attention is drawn to the traditional formula expressing the laws of growth ($y=e^{kx}$) and decay ($y=e^{-kx}$). At the beginning of the 20th century American scientists revealed the law related to the latter formula.

It reflects the approximate dependence of the area of protracting wounds from the time when the wound becomes sterile. This dependence can be traced with a special device, a planimeter. A planimeter is used for approximate measurement of the surface area bounded by lines. The perfect curve of wound healing is described by the formula: $S=S1e{-kt}$, where $S1$ is the wound area at the initial time. The perfect curve is also called the prediction curve. A prediction curve is compared with an actual curve (Figure 6). The wound is infected if the observed wound area is larger than the area defined by the perfect curve. If a wound heals faster than the perfect curve shows, this indicates the appearance of secondary ulcers. A wound is healing well if the prediction curve is the same as the actual curve.

Such examples, as shown in our monitoring of the teaching-learning process, significantly enhance the general attractiveness of mathematics for students and across disciplines.
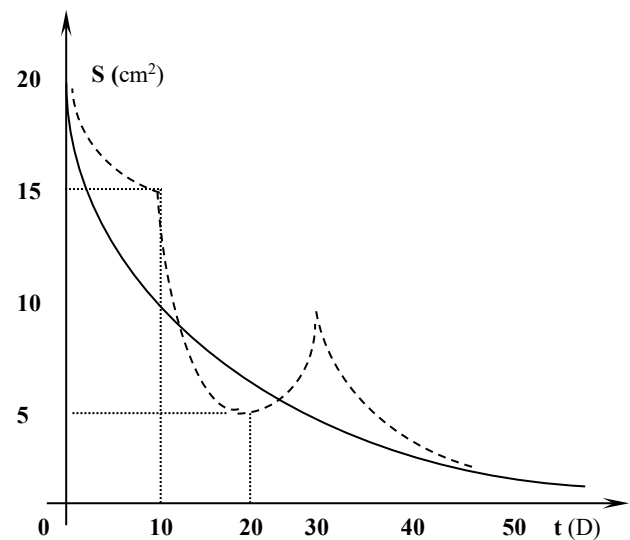
Based on the previous discussion, the motivational characteristics of practical problems can be presented in a systematic form in Table I.

Table 1: Systematic form of practical problems

Stage	Role of practical problem	Motivating factor	Objective	Instructional techniques	Students' activity
1	Stimulate students' interest	Intention to link learning material with their own life experiences	Deriving of mathematical relations; understanding concepts, theorems, algorithms and their application	Organization of empirical support for students' activity	Using students' life experience, applying it to solve non-standard word problems
2	Develop students' world outlook	Intention to apply mathematical material which was studied in other related school disciplines and real-life practice	Demonstration of practical application of mathematics in various branches of human knowledge	Forming of modeling method	Using plausible reasoning to solve practical problems related to other school disciplines
3	Students' orientation in the modern world	Intention to apply mathematical tools for association to their environment	Awareness of importance of mathematical knowledge and skills for further education and profession	Understanding of the features of mathematical models and their use for practical activity	The development of research with the involvement of the appropriate mathematical tools

This table demonstrates the implementation of the motivational role of practice and instructional techniques related to students' activity. The majority of the characteristics reflected in Table I are not tied to a particular stage of the teaching process. Mathematical knowledge and skills are enriched when students move to the next stage, which reinforces the stability and depth of their motivation.

5. Conclusion

The study of the nature of mathematical activity identifies a number of key factors which help to creatively regulate and stimulate motivational processes. These include capacious and precise 'linguistic' means of expressing mathematical regularities, the need for their justification and the development of a mathematical apparatus that enables a solution to any given problem of an applied nature. These factors can be described in hierarchical sequence [17, 18].

The initial link in this sequence is the practical need for solving specific pressing problems from the field of real human activity by the application of tailor made mathematical activity. In particular, it is expressed in the fact that any given practical task can stimulate the development of certain mathematical methods which in turn facilitate useful generalizations that would later extend those methods to a whole range of practical problems. This can itself become the starting point for relatively long-term mathematical research and open up the possibility of applying the developed mathematical theories to as yet unsolved practical problems. It can also help specify the prerequisites for the application of the developed mathematical tools to future stages of the development of science.

All these aspects should be taken into account when studying the school based mathematics curriculum. They offer a rich set of motivational opportunities. The particular motivational mechanisms which are applied will always initially be mindful of age range. In the initial stage of teaching mathematics, practical activity is basically a directly stimulating function in which the main role belongs to observation and experiment (real tasks for computing, measuring, plotting, constructing, etc.).

Motivational potential is then enhanced by demonstrating how a specific method for solving a specific problem can be used as the launching pad for the further development of the problem and for proving a theorem. The intention in so doing is to correlate this activity with the students' previous real life experience. Later, the motivational role of practice is realized through the recognition of the role of the studied mathematical content in the deployment of related courses and also through consideration of the history of the emergence and evolution of scientific concepts and methods. Further it develops understanding of the various elements of mathematical modeling of real states and processes which underpin the mastery of applied mathematical ideology. Textual tasks are at the core of the mechanism for such an implementation. All this continues at the most senior levels, but this is now combined with a growing awareness of ecological orientation – of the importance of the mathematic repertoire for general involvement in the world and in particular for the successful implementation of future professional activity. This function can be performed using real-world applications using a sufficiently serious mathematical tools.

The implementation of these mechanisms, reflecting the specifics of the implementation of the motivational role of practice in relation to mathematics, requires careful and precise correlation with each student's educational activity. The relevant material is presented in detail in our textbooks and articles.

References

[1] R. Courant, H. Robbins, What is Mathematics? An elementary approach to ideas and methods: London, Oxford University Press Incorporated, 1996.

[2] M. Kline, Mathematics. The Loss of Certainty: New York, Oxford University Press, 1980.

[3] M. Rodionov, The Formation and development of students motivation: Saransk, MGPI, 2012.

[4] M. Rodionov, Z Dedovets, " Practical Problems as Tools for the Development of Secondary School Students' Motivation to Learn Mathematics" in 19th International Conference on Computer Supported Education and Information Technology, New-Your, USA, 2017 https://waset.org/publications/10008026/

[5] G. Polya, Mathematics and Plausible Reasoning: Induction and analogy in mathematics: Princeton, New Jersey: Princeton University Press, 1954.

[6] L.C. Karpinsky, H. Y. Benedict, J. W. Calhoun, Unified Mathematics: Boston-New York-Chicago, D. C. HEATH & CO., 1918.

[7] T. Larkin, D. Budny, "Learning styles in the classroom: approaches to enhance student motivation and learning" in 6th International Conference on Information Technology Based Higher Education and Training, Santo Domingo, Venezuela, 2005. https://ieeexplore.ieee.org/document/1560310/

[8] M. B. Balk, V.A. Petrov, "About the mathematization of problems arising in practice" Mathematics at school, 3, 55-57, 1986.

[9] N. A. Tereshin, Applied orientation of school mathematics: M. Education, 1990.

[10] A. Ovezov, "Applying mathematical reasoning" Mathematics at school, 4, 45–52, 1998.

[11] A. D. Myshkis, M. M. Shamsutdinov, "Methods of applied mathematics" Mathematics at school, 2, 12-14.

[12] I. I. Blekhman, Mechanics and Applied Mathematics. Logic and applications of mathematics: M. Science, 1983.

[13] P. T. Apanasov, N. P. Apanasov, Collection of math problems with practical content: M. Education, 1987.

[14] H. Freudenthal, Mathematics as an Educational Task: Dordrecht-Holland, D. Reidel Publishing Company, 1973.

[15] W. Sawyer, Prelude to Mathematics: M. Education, 1972.

[16] I. Stewart, Concepts of Modern Mathematics: New York, Dover Publications, 1995.

[17] A. N. Tikhonov, D. P. Kostomarov, Stories about applied mathematics: M. Science, 1991.

[18] G. M. Wozniak, "The motivation problem in education" Mathematics at school, (2), 9-11, 1990.

ASTES

A Study on Improving Security and Efficiency using Dynamic Ownership Management in Client-Side Deduplication Environments

Won-Bin Kim, Im-Yeong Lee*

Department of Computer Science & Engineering, Soonchunhyang University, 31538, Republic of Korea

A R T I C L E I N F O	A B S T R A C T
	Data deduplication technology is used to improve the spatial efficiency of cloud storage. This technology is used for storing data on a cloud and omitting data uploading if the data are already present. However, various security threats may occur during the deduplication process. These security threats include poison attacks and user identity exposure through ownership. In addition, in an environment in which ownership changes in real time, there is a problem in renewing ownership information that has already been issued. Therefore, various studies have been conducted to solve these problems. In this study, a poison attack, real-time ownership management, and ownership anonymization are provided through MLE and dynamic ownership management.

1. Introduction

The rapid development of information communication and technology (ICT) in the recent times has led to many changes in data storage environment. According to a Dell EMC report in Korea, the amount of data produced in 2013 was 4.4 trillion GB, and this value is expected to increase to 44 trillion GB by 2020. As the amount of data produced increases, there is a demand for storage media with sufficient capacity to accommodate the data. Accordingly, in the future, the basic size unit of the storage medium will exceed terabytes and will be of the order of petabytes. However, portable hard disk drives (HDDs) and universal serial bus (USB) memories, which have been used in the past, have to be carried along always, and there exists the risk of losing them.

Cloud storage is a storage service that is available remotely over a network. It provides an environment wherein multiple users can access the storage simultaneously. Therefore, it is necessary to accommodate the data of a large number of users, and it involves maintenance and expansion costs, such as the costs for periodic storage space expansion, because it is necessary to ensure availability always. However, in general, much of the data stored by users are the same. Therefore, some of the storage space of the cloud storage is wasted in storing the same data repeatedly. To solve this problem, a data deduplication technique is proposed.

Data deduplication is a technique that reduces the amount of data stored in the data storage by preventing duplication of the stored data[1]. Because most of the data stored in the data storage

are stored as the same data repeatedly, storage space is wasted. Data deduplication technology allows confirming that the data to be added are stored when the data are added to the storage. At this time, if the data to be added are already stored, the data are not stored, and the ownership of the data is given to the user. Therefore, the use of data deduplication technology can prevent repeated storage of the same data and improve storage space efficiency.

The cloud storage is a remote server. Therefore, data can always be leaked because of internal or external threats. The data stored in the storage must be encrypted so that the contents of the data can't be accessed by unauthorized users. However, common encryption techniques can't be applied simultaneously with deduplication because they don't know whether the two encrypted data originated from the same source. To solve this problem, a secure data deduplication technology using various technologies such as convergent encryption (CE) has been developed.

The initial technique for secure data deduplication was developed to enable simple deduplication of encrypted data. However, during the process of secure data deduplication, various security threats such as poison attack and ownership forgery attack occurred. In addition, a number of techniques were studied to solve such threats, but these techniques created additional problems such as the inconvenience of ownership management and excessive operation. In this paper, we propose an improved method of secure data deduplication. This paper is an extension of work originally presented in 2017 4th International Conference on Computer

*Corresponding Author. - Im-Yeong Lee, Email: imylee@sch.ac.kr

Applications and Information Processing Technology (CAIPT) [2].

Figure 2. Server-side deduplication
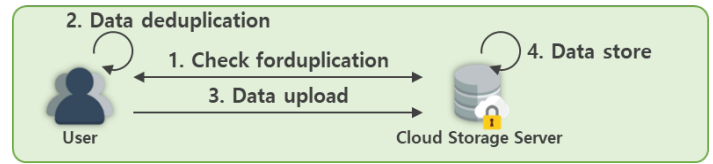
Figure 3. Client-side deduplication

Figure 1. Data deduplication

2. Related Works

2.1. Data Deduplication

Data deduplication is a technique that prevents the same data from being repeatedly stored. To achieve this, when data are stored, it is necessary to check whether the same data are already present in the storage system or device as shown in Fig. 1. Therefore, with data deduplication a comparison of the data is conducted to check whether the same data are already stored. To do so, the data source is hashed and compared, and the data are stored according to the comparison results. Various methods can be applied during this process, and various types of systems can be designed according to the method used.

2.2. Client-Side Deduplication (CSD)

Data deduplication uses a method for determining whether the data to be uploaded have already been stored. During this process, it is the responsibility of the storage server to determine whether the data already exist. However, the process of removing redundant data may take various forms. In the initial data deduplication method, all data to be uploaded are transmitted to the server, and the deduplication process is conducted in the server, as shown in Fig. 2. This is called server-side deduplication. However, because all data including redundant data are transmitted, a large amount of data transmission traffic occurs irrespective of the ratio of redundant data. In addition, bottlenecks may be incurred when data are uploaded from many different users
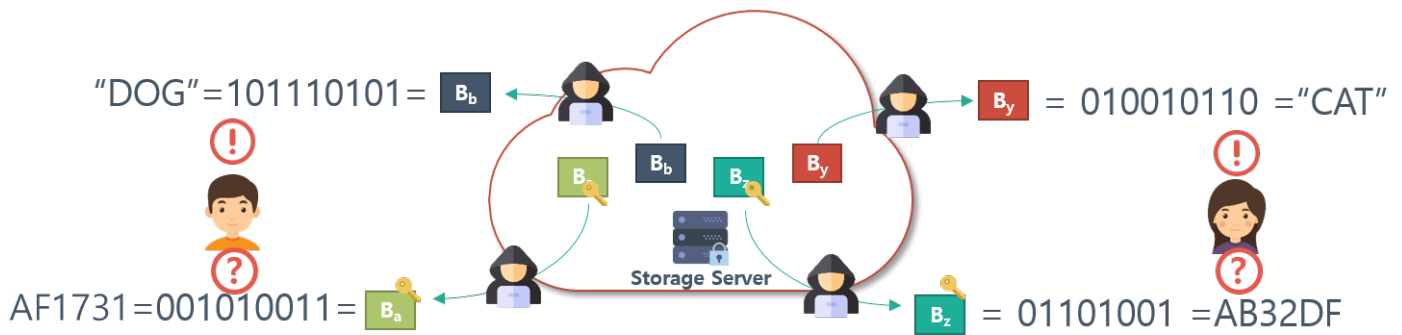
Figure 4. Necessity of data encryption
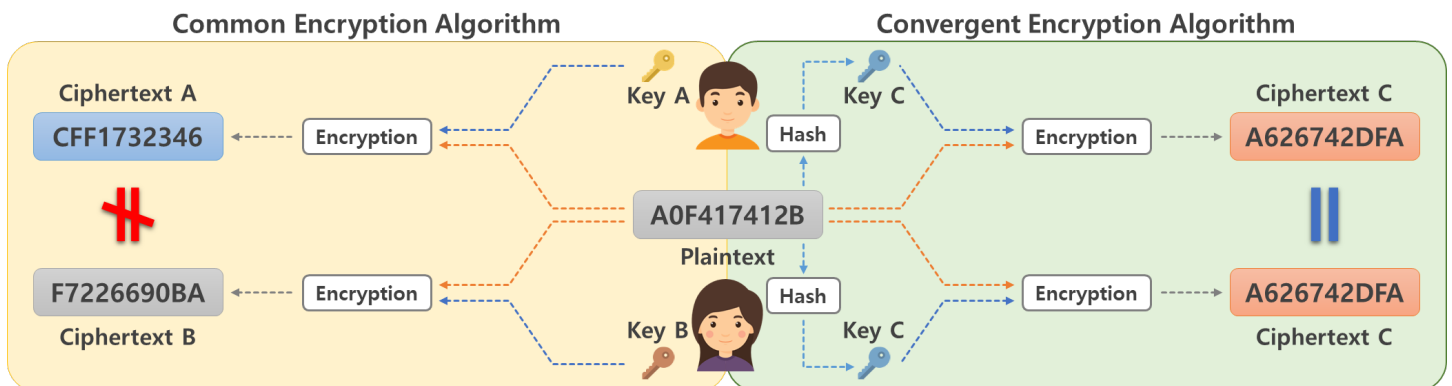
Figure 5. Comparison of common encryption and convergent encryption
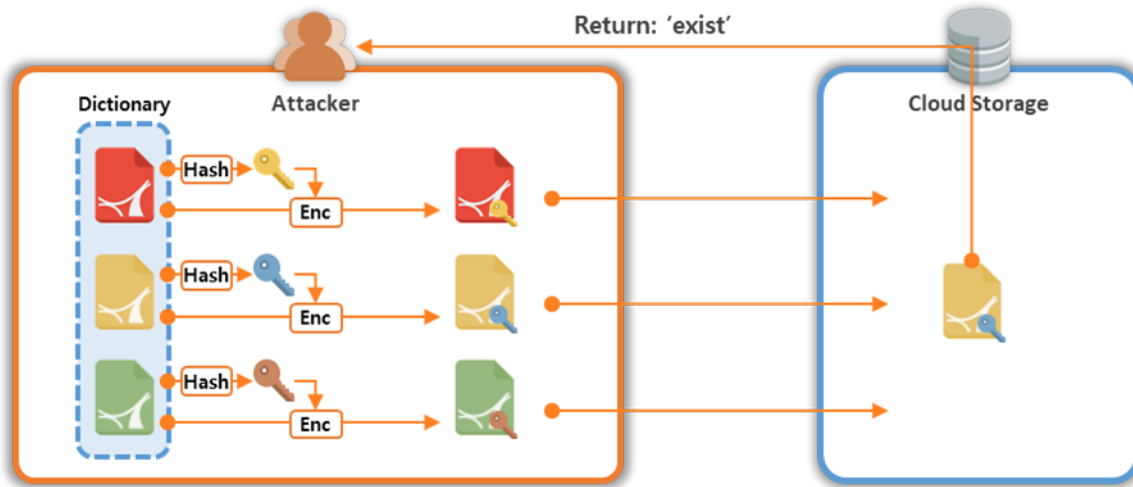
Figure 6. Dictionary attack scenario

concurrently. Therefore, client-side deduplication (CSD) technology has been proposed to solve this problem.

CSD is a method for determining redundancy by comparing a list of stored data with a list of identifiers from the data to be uploaded by the client to the server, as shown in Fig. 3. The server determines whether the data are redundant and transmits a list of non-duplicated data to the user. The user conducts deduplication using the duplicated data list received from the server, and transmits only the deduplicated data to the cloud storage server. CSD can reduce the amount of data transmission traffic and load on the server side. In this study, we used the CSD method [1-3].

2.3. Convergent Encryption (CE)

Data deduplication is conducted by comparing whether two data are the same. Therefore, we use hashed data for the data comparison. However, cloud storage is provided at a remote server, and thus has an honest-but-curious attribute. Therefore, threats continuously occur from insiders or from outside the system, and an encryption technique is thus required to prevent the data source from being known even when the data are leaked, as shown in Fig. 4. However, even if encrypted data are the same as the original data, the result depends on the encryption key used for encryption, and thus a data comparison based on the data deduplication technique cannot be conducted. To solve this problem, convergent encryption (CE), shown in Fig. 5, has been proposed [4].

CE is a technique for hashing a data source so as to generate hash data of a fixed size and using the source as an encryption key. Therefore, even if different users encrypt the data, the same encryption key and ciphertext are always generated. M. Storer proposed a technique for the deduplication of secure data using this approach [3]. First, key generation and encryption are conducted through the CE process. Thereafter, it is determined whether the data have been duplicated using the identifier of the encrypted data, and the data are additionally stored according to the result. Therefore, encryption and deduplication can be conducted concurrently. CE is used in most data deduplication environments, and various types of technologies have been proposed to utilize this technique. However, CE incurs a threat of poison attacks

caused by different data and their identifier, as well as dictionary attacks that infer the data sources.

2.4. Dictionary Attack

A dictionary attack is an attack in which a ciphertext is decrypted, or related information is obtained, using pre-existing data. CE uses a method for hashing the data source to generate an encryption key, and encrypting it using the encryption key. Therefore, if an attacker knows the data source, the attacker can find the encrypted data and the encryption key, as shown in Fig. 6. Therefore, the attacker attempts an attack using a data list that is guessed as the data source. First, the attacker obtains the key by hashing the data that are guessed as the data source. When encryption is conducted using an encryption key and a data source, password data are generated. By comparing the generated secure data with a ciphertext maintained in storage, it is possible to confirm that the data generated by the attacker matches the original stored data. Such threats are quite serious in environments where data are predictable [5].

2.5. Poison Attack

A poison attack is a threat that occurs when the data stored do not match the actual data, as shown in Fig. 7 [6]. When a poison attack occurs in data M, two types of threats can be created:

- **Loss of data source:** If the user who acquired the ownership by performing a subsequent upload to the data where the poison attack occurred deletes the data source from the local storage, the original data can't be acquired again.

- **Damage due to malicious code or modified data:** Failure to detect tampering of data when downloading poison attack data may result in damage due to malicious code contained in the tampered data or property damage due to tampering with sensitive information.

To protect user's data from poison attacks, user need to be able to check whether stored data and metadata are generated from the same data. The RCE used in this study compares the tag with the data source and performs integrity verification after downloading
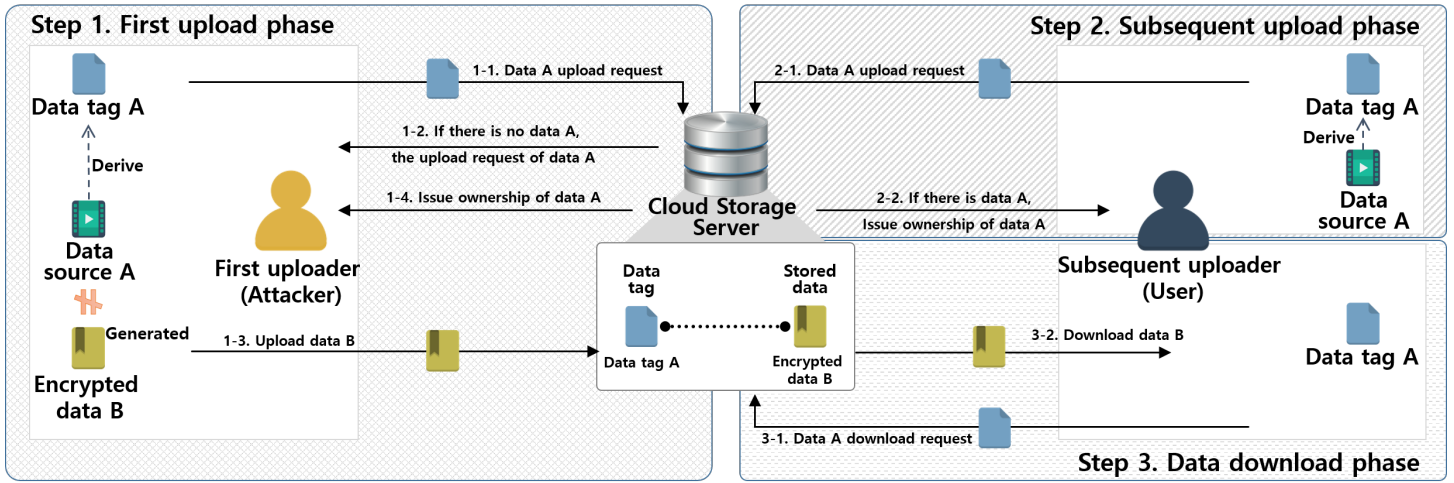
Figure 7. Poison attack scenario
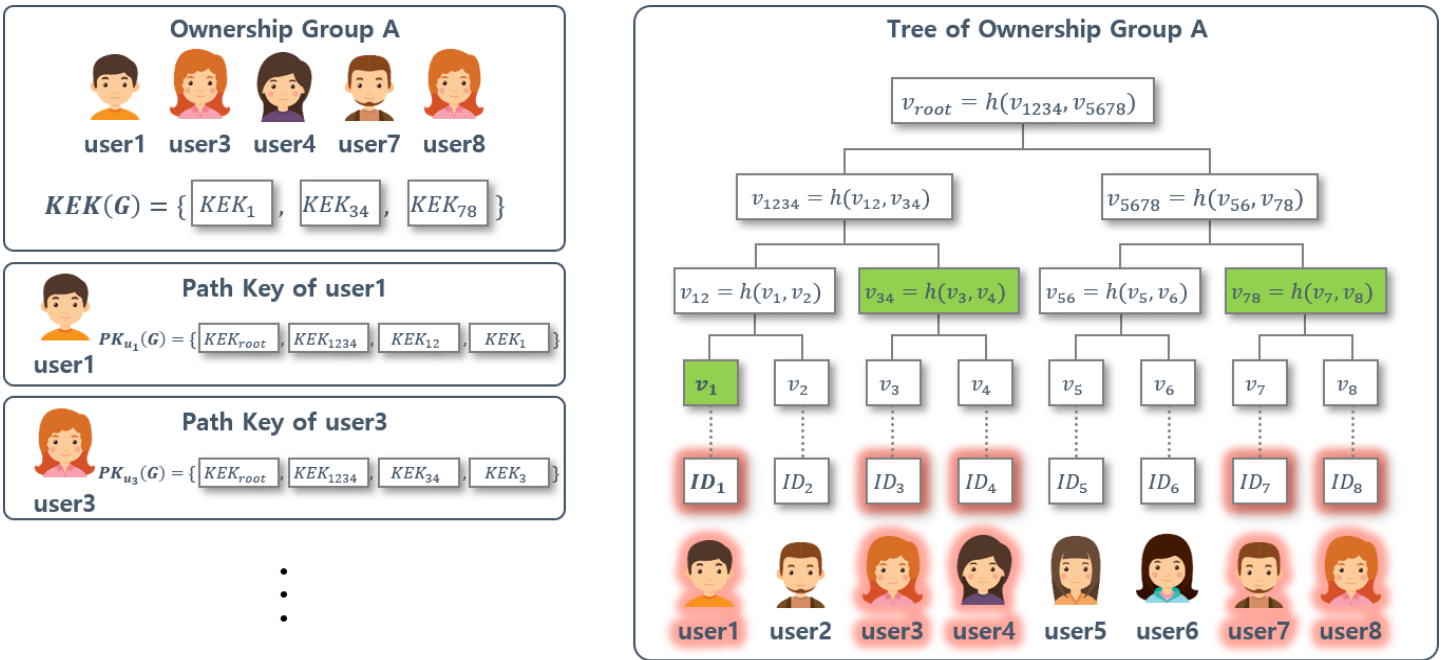
Figure. 8. Method of Ownership Management with Re-encryption

the data. However, since the direct association between the tag generated from M and the cipher text C can't be found in the data uploading stage, it is vulnerable to a poison attack.

2.6. Message-Locked Encryption

Message-locked encryption (MLE) offers four types of cryptographic techniques for the deduplication of encrypted data using technology proposed by Bellare et al. [7]. As a common MLE approach, data source M generates an encryption key K by computing K←H(M). Four types of MLE have been developed: CE, hash and CE without a tag check (HCE1), hash and CE with a tag check (HCE2), and randomized convergent encryption (RCE). The CE and HCE1 methods conduct only encryption and decryption. The HCE2 method conducts encryption and decryption along with tag integrity verification. The RCE method is based on HCE2, and the key generation entropy is improved through the generation of a randomized key in a one-time pad

during the key generation process. In this study, a protocol design using RCE was applied.

2.7. Proof of Ownership (PoW)

In client-side deduplication technology, if the uploaded data are duplicated, ownership is issued without uploading the data. However, a problem arises if an attacker attempts to upload data by forging a data identifier. In this case, the attacker can take ownership of the data without the data source. To solve this problem, a technique is proposed to verify data ownership without transferring the data source. This technique is referred to as proof of ownership (PoW) [8].

2.8. Dynamic Ownership Management

A variety of data are stored on a cloud, and each data group has its own ownership group. However, if a change occurs, such as the issuing and discarding of ownership, a problem arises in that all

ownership of the ownership group must be renewed. In general, it is relatively easy to issue ownership to a new user. However, if ownership information for a particular portion of data changes, it is difficult to change the ownership previously issued to the user. Therefore, dynamic ownership management technology is necessary to manage changes in ownership information in real time. In this study, we use proxy re-encryption for dynamic ownership management.

2.9. Hur et al.'s scheme

Hur et al.'s scheme was proposed in 2016 [9]. The authors proposed a method for achieving dynamic ownership management in a secure data deduplication environment. This scheme addresses the problem of identifying the data owner using the ownership information of the user by providing anonymity of ownership. As a way to provide anonymity of ownership, a general method for confirming ownership through the ownership group was developed. However, this approach complicates the ownership management because the ownership of the entire group must be changed at the time of ownership issue and renewal. re-encryption was proposed to solve this problem, along with a dynamic ownership management technology providing better efficiency. However, this scheme is based on a server-side deduplication environment, is continuously affected by the redundancy rate of the data, and always involves the same number of computations. Therefore, in the present study, we researched an improved technology by applying the idea of Hur et al.'s scheme.

In Her et al.'s scheme, ownership management is performed using the Merkle Hash Tree (MHT) for dynamic management of data ownership. In the leaf node of the MHT, the identification information of the users is located, and the MHT is configured as shown in Fig. 8 using this identification information. Ownership group G_i contains a list of users having ownership of data M_i. Also, $KEK(G_i)$ is configured as shown in the left side of Fig. 8 so that users included in G_i can be included in the minimum number of nodes. Therefore, user1, user3, user4, user7, user8 belong to group G and $KEK(G)$ consists of KEK_1, KEK_{34} and KEK_{78} in Fig. 8. Also, for each user, a node included in the user's own identifier path from the root node is provided as a path key(PK). In Fig. 8, PK_{u_1} (G) is PK of user1. And it includes $KEK_{root}, KEK_{1234}, KEK_{12}$, and KEK_1.

The ownership manager generates $KEK(G_i)$, which is a list of KEK of the ownership group G_i, and provides $PK_{u_1}(G_i)$ to the user. Then, after generating the group key GK_i of the group G_i, the data M_i is encrypted to generate the cipher text $C_i = E_{GK_i}(M_i)$. Finally, the group key GK_i is encrypted with $KEK(G_i)$ to complete the update of the ownership group for the data M_i.

Through the above process, the user belonging to the ownership group can obtain the data $M_i = D_{PK_{u_1}(G_i) \cap KEK(G_i)}(C_i)$ by using the $PK_{u_1}(G_i)$ held by the ownership group and the $KEK(G_i)$ provided by the ownership manager.

3. System Requirements

3.1. System configuration

The system structure required in this study is shown in Fig. 9. Users access the cloud storage, which includes storage and metadata servers. Therefore, the user stores the metadata server,
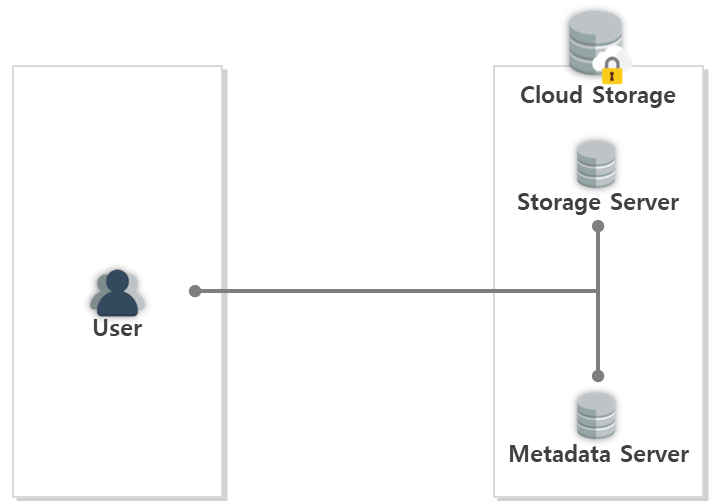
Figure 9. System configuration

data tag, and verification data, and the encrypted data are stored in the storage server.

3.2. Security requirements

The six security requirements are poison attack resistance, dictionary attack resistance, confidentiality, integrity, anonymity, and efficiency.

- **Confidentiality:** Data stored in cloud storage must be encrypted and archived in case of data leakage. Encryption should also be considered to enable data deduplication.

- **Integrity:** Data stored in cloud storage should be preserved without modification. If the data stored in the cloud storage are modified, the user should be able to detect the deformation when downloading the data.

- **Anonymity :** Ownership information stored in the server can only be used to verify the user's ownership, and information that can identify the user's identity should not be included.

- **Efficiency:** In the case of data deduplication, a large number of operations and data transmissions can occur. In this case, the computation and data traffic efficiency must be provided during the process because the benefits of the data deduplication are offset.

- **Resistance to poison attack:** A tag is derived from the data source and must be able to verify the integrity of the original data. This prevents poison attacks that may occur owing to the different tag sources and data stored in the server.

- **Resistance to dictionary attack:** An attacker should not be able to attack a dictionary. Therefore, an attacker should not be able to derive data or data encryption keys through the CE using an analogy with the data source.

4. Proposed Scheme

The proposed technique is based on Hur et al.'s concept of dynamic ownership management. In the proposed method, the data upload request phase is initially applied, and the first and subsequent data upload phases are conducted depending on whether the data are duplicated. If a poison attack is detected in the
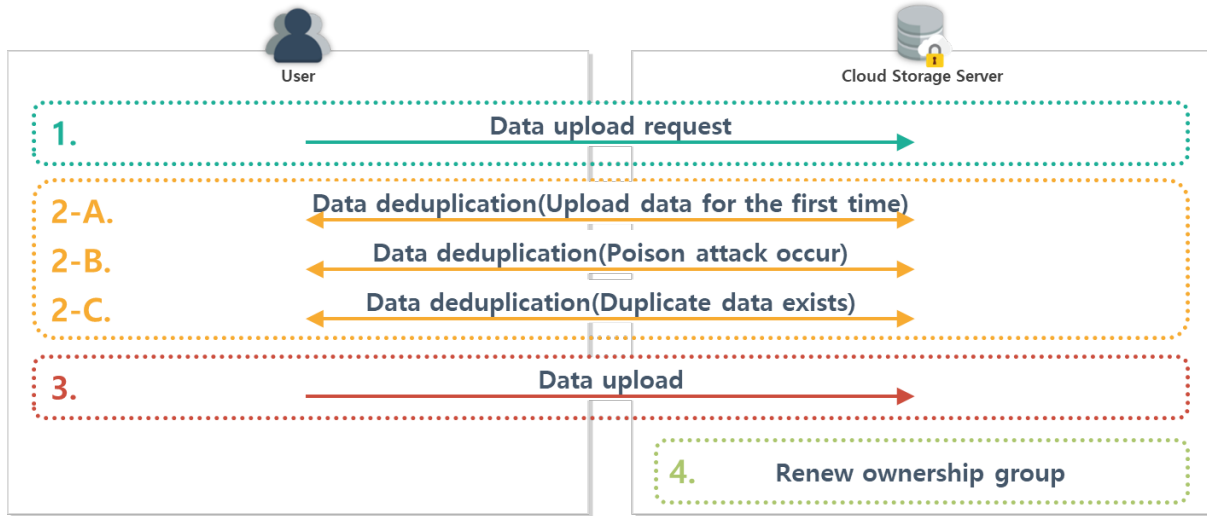
Figure 10. Overview of uploading data in the proposed scheme
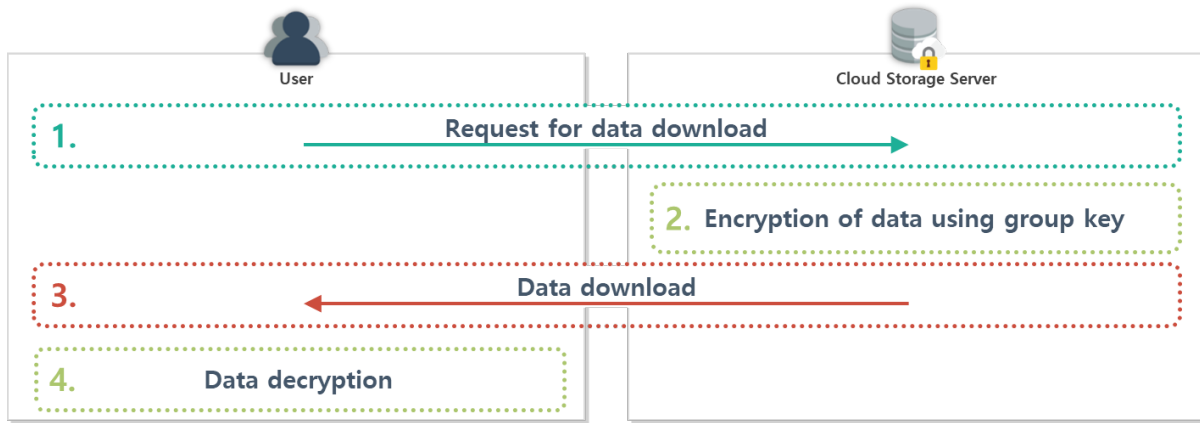
Figure 11. Overview of downloading data in the proposed scheme

$$K_i \leftarrow H(M_i) \tag{1}$$

4.1. Data Upload Request Phase

This step is a common step in the data upload process. At this stage, the user uses the identifier of the data to be uploaded, and requests the server to confirm whether the corresponding data exists, as shown in Fig. 12.
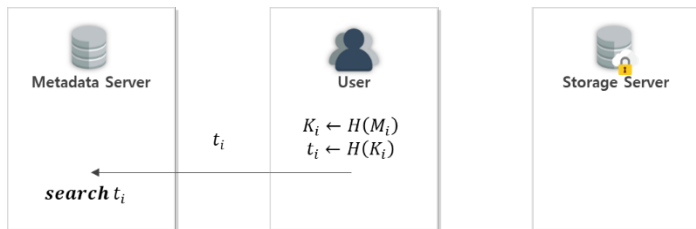
Figure 12. Data upload request phase

The user converts data M_i to be uploaded into a hash algorithm for generating K_i and t_i. The generated t_i is then transmitted to the metadata server to request confirmation of storage.

4.2. Data Upload Phase (First Upload)

This step is conducted when the corresponding data do not exist as a result of the data upload request phase, as shown in Fig. 13.

The metadata server determines that there are no data that a user has requested to upload and transmits the result to the user.

A user who receives a response from the metadata server generates L_i, C_i^1, C_i^2, and C_i^3.

$$L_i \overset{\$}{\leftarrow} \{0,1\}^{\lambda(K)} \tag{3}$$

$$C_i^1 \leftarrow E_{L_i}(M_i) \tag{4}$$

$$C_i^2 \leftarrow L_i \oplus K_i \tag{5}$$

$$C_i^3 \leftarrow H(M_i \parallel L_i) \tag{6}$$
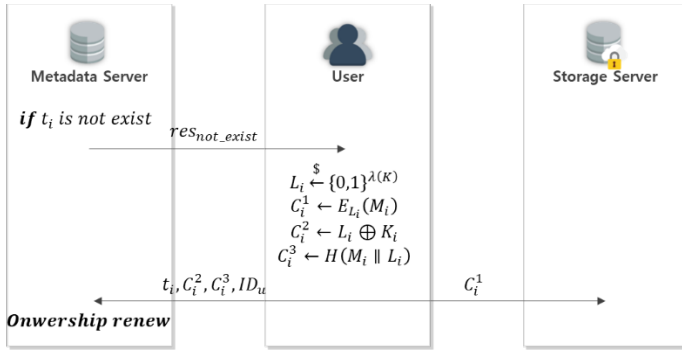
Figure 13. Data Upload Phase (First Upload)
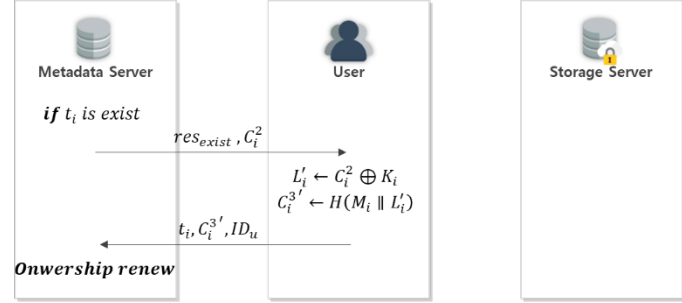
Figure 14. Data Upload Phase (Subsequent Upload)
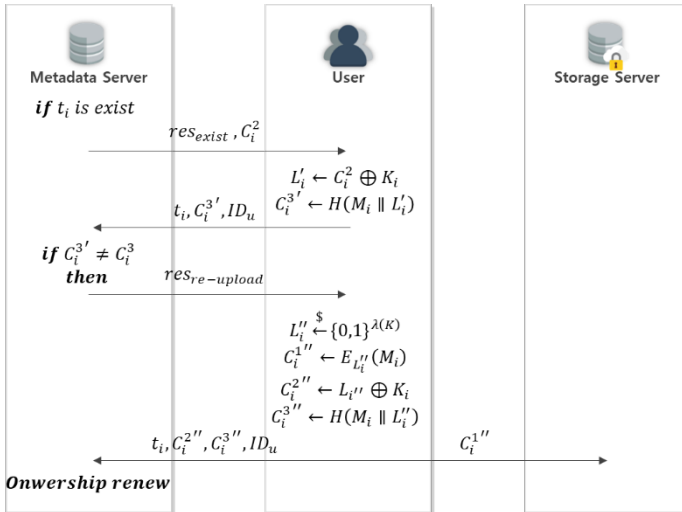
Figure 15. Data Upload Phase (when a Poison attack occur)

The user obtains L_i by calculating data C_i^2 obtained from the metadata server. Then, $C_i^{3'}$ is generated using the obtained L_i.

$$L_i' \leftarrow C_i^2 \oplus K_i \qquad (7)$$

$$C_i^{3'} \leftarrow H(M_i \| L_i') \qquad (8)$$

The user sends the created t_i, C_i^3 and ID_u to the metadata server. The ownership group renewal is then conducted.

4.4. Data Upload Phase (when a Poison attack occur)

This step is conducted when it is determined that a poison attack has occurred on the data uploaded during the data upload phase (subsequent upload), as shown in Fig. 15.

The metadata server determines that there are data that the user has requested to upload, and transmits the result and C_i^2 to the user.

The user obtains L_i by calculating data C_i^2 obtained from the metadata server, as shown in (7). Then, $C_i^{3'}$ is generated using the obtained L_i, as shown in (8).

The user sends the created t_i, C_i^3, and ID_u to the metadata server.

The metadata server determines that a poison attack has occurred when the user-uploaded $C_i^{3'}$ is different from C_i^3 stored in the metadata server.

The metadata server then requests the user to re-upload the data. Ownership group renewal is then conducted. A user who receives a response from the metadata server generates L_i'', $C_i^{1''}$, $C_i^{2''}$, and $C_i^{3''}$.

$$L_i'' \overset{\$}{\leftarrow} \{0,1\}^{\lambda(K)} \qquad (9)$$

$$C_i^{1''} \leftarrow E_{L_i''}(M_i) \qquad (10)$$

$$C_i^{2''} \leftarrow L_i'' \oplus K_i \qquad (11)$$

$$C_i^{3''} \leftarrow H(M_i \| L_i'') \qquad (12)$$

The user sends the generated t_i, $C_i^{2''}$, $C_i^{3''}$ and ID_u to the metadata server and transmits $C_i^{1''}$ to the storage server. Then, the ownership group renewal is conducted.

4.5. Ownership Group Renewal Phase

This step is executed after the data upload phase ends. Include the user who has been issued ownership in the ownership group. In this process, the metadata server encrypts and archives the GK_i using the added KEK. Fig. 16 shows the formation and management of the user's tree and ownership group.

The owner group G_i of data M_i has a key $KEK(G_i)$ for encrypting the group key GK_i. The $KEK(G_i)$ term is a KEK list of the minimum nodes that can include all user nodes included in G_i. For example, when $G_{data1} = \{u_1, u_3, u_4, u_7, u_8\}$, $KEK(G_{data1}) = \{KEK_1, KEK_{34}, KEK_{78}\}$ is established. The group update procedure is as follows.

The metadata server applies a data deduplication process with the user, and issues ownership of data M_i to user u_2.

The metadata server includes the user in ownership group G_i of data M_i. Then, $KEK(G_i)$ is updated. The added KEK is assumed to be KEK_2.
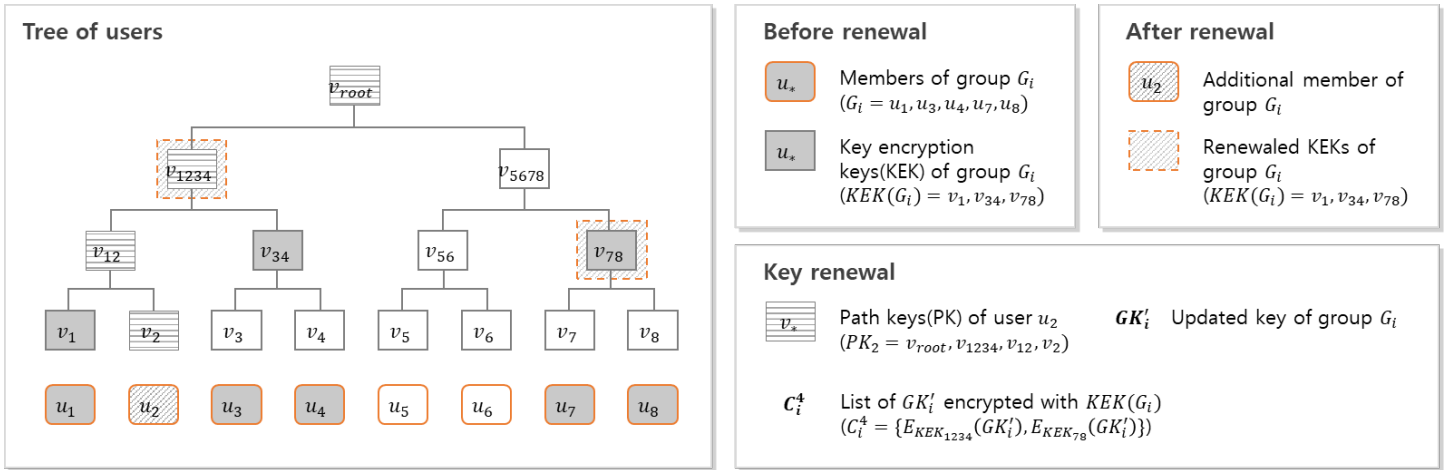
Figure 16. Method of renew ownership
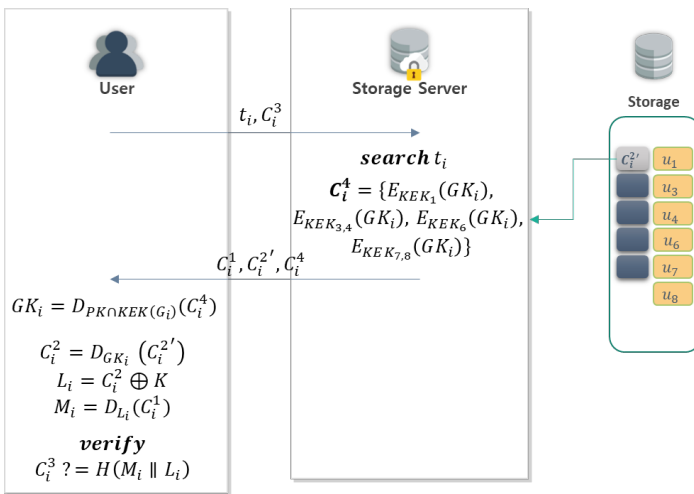
Figure 17. Data download phase

$KEK(G_i)$ is composed of KEK, which can cover all users included in group G_i with the minimum KEK by combining the added and existing KEK.

The metadata server encrypts C_i^4 with GK_i, and encrypts GK_i with KEK_{1234}.

4.6. Data Download Phase

This procedure corresponds to Fig. 17, and is conducted when a user who has been issued ownership of data M_i requests data to be downloaded from the server.

The user transmits identifier t_i of the data to be downloaded and verification data C_i^3 to the metadata server.

The server searches for data corresponding to t_i, and then transmits $C_i^1, C_i^{2'}$, and C_i^4 stored in the server to the user.

The user obtains M_i using $C_i^1, C_i^{2'}, C_i^4$, and the user's own PK_u and k_i.

$$GK_i \leftarrow D_{PK \cap KEK(G_i)}(C_i^4) \qquad (13)$$

$$C_i^2 \leftarrow D_{GK_i}(C_i^{2'}) \qquad (14)$$

$$M_i \leftarrow D_{L_i}(C_i^1) \qquad (16)$$

$$L_i \leftarrow C_i^2 \oplus K \qquad (15)$$

The user compares C_i^3 and $C_i^{3'}$. This allows verification of the data integrity.

$$C_i^3 \, ? = H(M_i \parallel L_i) \qquad (17)$$

5. Analysis of Proposed Scheme

In this paper, we propose a more efficient and secure method for deduplication of encrypted data based on Hur et al.'s dynamic ownership idea [9]. Therefore, it has similar features to the Hur et al. method in certain areas, but differs in the following ways.

Confidentiality: Cloud storage is continuously exposed to data breaches. Therefore, data stored in the cloud must be encrypted and stored. However, because data deduplication and data encryption have opposite characteristics, we used CE to apply them concurrently. CE is a technique for obtaining an encryption key K by hashing the data source M as in (18). Therefore, the same encryption key and ciphertext are always generated for users who own the same data source. Therefore, using CE, it is possible to deduplicate the encrypted data. In this proposed scheme, data encryption is conducted using RCE mode of MLE based on CE.

$$K \leftarrow H(M) \qquad (18)$$

- **Integrity:** The user who has downloaded the data can verify whether the downloaded data have been transformed through the RCE tag validation. With this process, the user obtains data $chunk_i$ and conducts a hash operation using the encryption key L_i to create C_i^3, as shown in (19), which can be verified.

$$C_i^3 \, ? = H(M_i \parallel L_i), \qquad (19)$$

- **Anonymity :** This method is based on Hur et al.'s approach. When a user requests a data download, the cloud storage server sends the encrypted group key to the user. A user with legitimate ownership can decrypt the key and can verify that

Table. 1. Comparison with other technologies

				Hur, et al.	Kim, et al.	Proposed Scheme
Deduplication Location				Server Side	Client Side	Client Side
Poison Attack Resistance				△	○	○
Proof of ownership				○	X	○
Dynamic Ownership Management				○	X	○
User Anonymity				○	X	○
Compu-tation	Upload	First upload	User	$2H + 1SE + 1\oplus$	$3H + 1SE$	$3H + 1SE + 1\oplus$
		Subsequent upload	User	$2H + 1SE + 1\oplus$	$3H + 1SE$	$3H + 1\oplus$
		Poison attack occurred	User	$2H + 1SE + 1\oplus$	$3H + 1SE$	$4H + 1SE + 1\oplus$
	Download		User	$1H + 3SE + 1\oplus$	$1H + 1SE$	$1H + 3SE + 1\oplus$
	Ownership group renewal		Server	$1H + 3SE + 1\oplus$	$1H + 1SE$	$1H + 3SE + 1\oplus$

○: Offer; X: Not offer; △: limited offer;
H: Hash algorithm; SE: Symmetric key encryption; $\oplus$: XOR Operation;
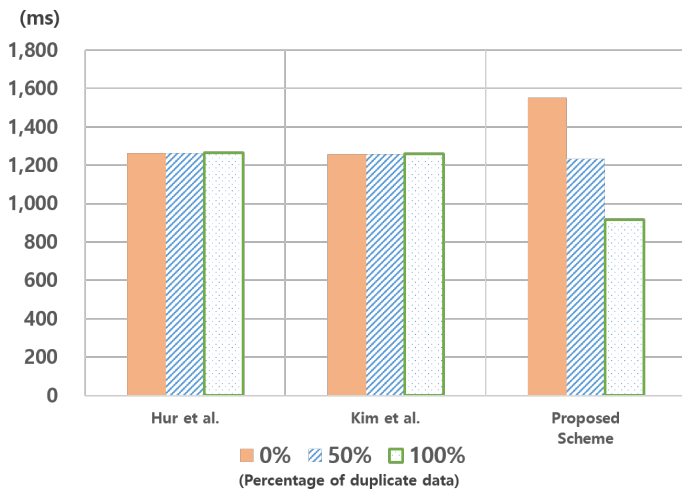
Figure 18. Computation time of 250k blocks
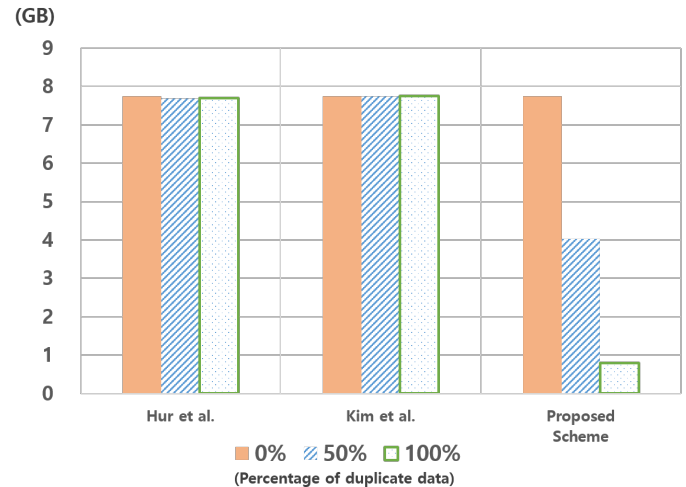
Figure 19. Data traffic of during deduplication of 250k blocks

it has legitimate ownership without identifying the user. Accordingly, the user has ownership of the corresponding data through this method, and thus downloads and decrypts the data through the process shown in (13-16).

- **Efficiency:** The proposed method shows that the total computational load decreases as the duplication ratio increases, the cause of which can be observed in Table 1. With the proposed method, one more hash operation per block is generated than in Hur et al.'s and Kim et al's method during the upload process when the first upload and poison attack occur[11, 9]. However, in the proposed scheme, the $3H + 1SE$ operation is omitted when redundant data exist. Therefore, the total number of operations can be reduced as the ratio of redundant data increases. In addition, the proposed scheme uses client-side deduplication. This method can conduct data uploads and deduplication with less computational and transmission overhead than server-side deduplication methods when uploading duplicated data, as shown in Fig 18, 19[10].

- **Resistance to poison attack:** In Hur et al.'s scheme, when a poison attack occurs, its occurrence can be checked at the download phase. However, data lost owing to a poison attack cannot be recovered. Therefore, the proposed scheme detects the poison attack in the subsequent upload phase to prevent data loss. In Hur et al.'s scheme, when a poison attack occurs, its occurrence can be checked during the download phase. However, data lost from a poison attack cannot be recovered. Therefore, the proposed scheme detects a poison attack during the subsequent upload phase to prevent a data loss [8].

- **Resistance to dictionary attack:** CE is vulnerable to dictionary attacks. Basically, because CE has a structure for obtaining an encryption key from a data source, the encryption key can also be obtained if the data source can be guessed. Therefore, to achieve resistance to a dictionary attack, data encryption key acquisition through data guessing should be made impossible. For this purpose, the proposed scheme conducts data encryption using RCE mode of MLE based on

CE. In RCE mode, the key obtained by hashing the data source is not used as a data encryption key, but is used as a key "K" for encrypting the data encryption key "L." The formula for this is shown in (8),(21–23). Therefore, even if a data source is guessed, it is possible to resist a dictionary attack, which is an attack achieved through data guessing, because it does not obtain the data encryption key directly, but acquires a key capable of decrypting the data encryption key.

$$L \xleftarrow{\$} \{0,1\}^{\lambda(K)} \qquad (21)$$

$$C^1 \leftarrow E_L(M) \qquad (22)$$

$$C^2 \leftarrow L \oplus K \qquad (23)$$

6. Conclusion

This study was conducted to improve the efficiency and security of data deduplication. Cloud storage wastes space owing to the redundant storage of the same data. Data deduplication has been proposed to solve this problem. Data deduplication is a technology that saves storage space by preventing the same data from being stored. Therefore, the uploaded data are compared with the data stored in the existing storage to make sure the data are the same. However, cloud storage always incurs a threat of data leakage, and thus data encryption should not be used to identify the data source. Therefore, cloud storage requires the data to be encrypted and archived through deduplication. However, CE was proposed to deal with the conflicting characteristics of data encryption and deduplication technologies. Because CE is a technology for generating an encryption key by hashing the data source, the same ciphertext is always generated when encrypting the same data source. It therefore becomes possible to deduplicate the encrypted data. However, CE poses a threat of a dictionary attack, and client-side deduplication poses a threat to a poison attack. Therefore, we use RCE mode of MLE to solve this problem. MLE is a CE-based encryption technology, and can prevent data-source guessing. In addition, the verification data generated in RCE mode can be used to identify the occurrence of a poison attack. In addition, when data deduplication is conducted, the ownership information of the user is updated. Because cloud storage is an environment used by multiple users, such ownership information is frequently updated. Therefore, ownership management difficulties may occur, and the identity of the user may be exposed when using such ownership information. To solve this problem, we applied ownership management technology using a binary tree based on Hur et al.'s scheme. Through this, this paper is safe from poison attack, dictionary attack, proof of ownership, ownership management and ownership anonymity. In addition, by applying a client-side deduplication environment, the number of computations and amount of communication can be reduced as the data redundancy ratio increases. As a result, the proposed scheme is effective against various security threats, and reduces the amount of computational traffic as the data redundancy ratio increases.

Acknowledgment

References

[1] Kim, K. W., Joo, Y. H., Eom, Y. I. "Technical trends for cloud storage data deduplication", Proceedings of Symposium of the Korean Institute of Communications and Information Sciences, 2012, pp. 228–229. DOI: 10.22648/ETRI.2018.J.330107

[2] Kim, W. B., Lee, I. Y., Ryou, J. C. "Improving dynamic ownership scheme for data deduplication." Computer Applications and Information Processing Technology (CAIPT), 2017 4th International Conference on. IEEE, 2017. DOI: 10.1109/CAIPT.2017.8320671

[3] Storer, M. W., Greenan, K., Long, D. D., Miller, E. L "Secure data deduplication", Proceedings of the 4th ACM International Workshop on Storage Security and Survivability, ACM, 2008, pp.1–10. DOI: 10.1145/1456469.1456471

[4] Douceur, J. R., Adya, A., Bolosky, W. J., Simon, P., Theimer, M. "Reclaiming space from duplicate files in a serverless distributed file system." Distributed Computing Systems, 2002. Proceedings. 22nd International Conference on. IEEE, 2002. pp. 617-624. DOI: 10.1109/ICDCS.2002.1022312

[5] Bellare, M., Keelveedhi, S., Ristenpart, T. "DupLESS: Server-Aided Encryption for Deduplicated Storage." IACR Cryptology ePrint Archive 2013 (2013): 429.

[6] Kaaniche, N., Laurent, M. "A secure client side deduplication scheme in cloud storage environments." NTMS 2014: 6th International Conference on New Technologies, Mobility and Security. 2014. pp. 1-7. DOI: 10.1109/NTMS.2014.6814002

[7] Bellare, M., Keelveedhi, S., Ristenpart, T. "Message-locked encryption and secure deduplication." Annual International Conference on the Theory and Applications of Cryptographic Techniques. Springer Berlin Heidelberg, 2013, pp. 296-312. DOI: 10.1007/978-3-642-38348-9_18

[8] Halevi, S., Harnik, D., Pinkas, B., Shulman-Peleg, A. "Proofs of ownership in remote storage systems", Proceedings of the 18th ACM conference on Computer and communications security. ACM, 2011, pp. 491-500. DOI: 10.1145/2046707.2046765

[9] Hur, J., Koo, D., Shin, Y., Kang, K. "Secure Data Deduplication with Dynamic Ownership Management in Cloud Storage", IEEE Transactions on Knowledge and Data Engineering, 28(11), 2016, pp. 3113-3125. DOI: 10.1109/TKDE.2016.2580139

[10] Kim, K., Youn, T. Y., Jho, N. S., Chang, K. Y. "Client-Side Deduplication to Enhance Security and Reduce Communication Costs." ETRI Journal 39(1), 2017, pp. 116-123. DOI: 10.4218/etrij.17.0116.0039

Proposed System of New Generation LMS Using Visual Models to Accelerate Language Acquisition

Imad Hasan Tahini*, Alex Dadykin

Belarusian National Technical University (BNTU), Minsk, Belarus

A R T I C L E I N F O	A B S T R A C T
	Language skill is a rule-like operation, based on generalized connections. The main property of language skills is awareness. They are formed with conscious mastery of the language means of communication (phonetic, lexical and grammatical). Language acquisition is a complex process which includes a large number of different parameters. Therefore, the study and improvement of language learning and teaching require creative collaboration between experts from different domains. We propose to transfer knowledge about the structure of the language from the verbal to the visual form, thereby creating the opportunity to use them as an indicative basis for planning, managing, controlling and correcting the training of primary language skills both by the teacher and by the student himself. The proposed method allows to develop the ability to organize sentences to convey meaning by means of Visual Models and to describe the sequential steps to choose the most effective ways for working with audio to improve listening skills. The use of visual tools for the analysis of abstract systems and theories makes it possible to discover new patterns and simplify their understanding. The aim of our study is to put principles for building a new effective system using new methods for acquiring language skills.

1. Introduction

This paper is an extension of work originally presented in the Sixth International Conference on Digital Information, Networking, and Wireless Communications (DINWC), 2018, Beirut, Lebanon [1]. It presents the use of new methods to facilitate teaching and learning. The new results were obtained by using a Structural Visual Method and Visual Models and tools of systems analysis and information technologies. Distinguishing feature of this method is using colors to encode meanings in structural diagram.

The 21st century is the century of the dominance of information and information technology. They are rapidly conquering the world, penetrating into all spheres of human activity, which led modern society to a general historical process called Informatics. This process consists in the free access of any citizen to information, penetration of information technologies into scientific, industrial, public spheres, high level of information services. To accelerate the acquisition of language for adults requires a new approach that differs about the traditional goal-

setting, the selection of new education type, and is realized in a practical way.

Our research singled out among the teaching methods "central" - a support for the language and gives a practical experience of adult learners who, in the process of learning, analyze and understand, generalize and evaluate it. This is the basis for selecting the content of education, the choice of new methods and the organization of the pedagogical process as a whole.

Specificity of the organization of the pedagogical process by using new technology in the education of adults is due to the fact that, firstly, adults have educational needs, life experience and independence, which have grown from their practical activities and are vital for them; secondly, physiological and psychological features that affect the process of perception, assimilation and assessment of knowledge and skills, processing and adaptation of information; Thirdly, adults are directed to the rapid realization of acquired knowledge, skills and qualities; fourthly, all the educational activity of adults is determined by temporary, spatial, professional, domestic, social conditions. At the same time, it is necessary to take into account the contradictions inherent in the life and work experience of adult learners, the "plus" is that experience serves as a basis for a profound understanding of the problems

*Corresponding Author: Imad Hasan Tahini, Beirut, Lebanon,
Email: imad.tahini@live.com

under study, but "minus" in the existing stereotypes of thinking, established ideas that impede the perception of the new not appropriate. The effectiveness of the adult education process will be high when a person is put in the position of a researcher who independently searches for a solution and is able to coordinate it with others.

The goal of this article is to use the visual modeling system that helps you to learn how to build English sentences without thinking about rules and theoretical aspects. This work is under way to introduce this approach into training applications, programs and new type of LMS.

The visual model comes here to the rescue, allowing you to analyze English grammar not in words, but with the help of models, visual images. This model is abstract forms, colors, arrows, images. They are similar to the subway scheme. All schemes allow you to control meanings using a visual system that is much more powerful than a verbal one.

This method helps to understand the grammatical of English. In just one session you get a holistic view of the dynamics of English grammar. And this is important for an adult. Of course, this does not exclude the need for intensive training. But now it becomes clear what to train. In addition, today this article has a unique grammar training system, similar to which not met anywhere.

This paper is organized as follows, Section 2 we do the analysis of the research methods about possible training to get a new proposed way of competency. Section 3 we display the methods for improving adult's foreign language skills as well as the methodology of the study. Sections 4 we create the structure of new LMS. Section 5 we describe a stage of designing and development a new system for adults. The conclusion is discussed in the last section.

2. Research of possible training ways

Analysis of the model of the psyche [2], in spite of its simplicity, allows us to see interesting patterns.

The first conclusion is that the only thing that a person is able to manage and what it is possible to teach is activities and ways of doing it. As Galperin [3] showed, the training of mental activity necessarily includes the stage of its implementation in the form of external, physically performed actions.

Therefore, the acquisition and assimilation of information is only an intermediate stage in the process of forming the skills of performing certain activities and should not be an end in itself of the learning process, as often happens.

The second conclusion - motivation and interest are the most important component of the learning process, and without them the process cannot be effective.

The following is that human behavior is controlled only by a small degree of consciousness and thinking, and there are at least several different behavioral control loops, often unconscious and inaccessible to direct impact and change. The human psyche is a very complex and complex phenomenon, and any of its models, including the most detailed and complex, cannot reflect all its properties and ensure full compliance. Therefore, several simple

models can bring more practical benefits than attempts to create a complete and all-encompassing model.

Let's consider known models of obtaining skills (competence) and trace the ways of their obtaining on the model of the psyche.

The model (Figure 1) for obtaining unconscious competence (skill), (usually attributed to Abraham Maslow [4], which is not true). It is widely known in the online learning environment and motivational and psychological training. It is structurally identical with the model of the psyche and is directly derived from it, and vice versa [2].
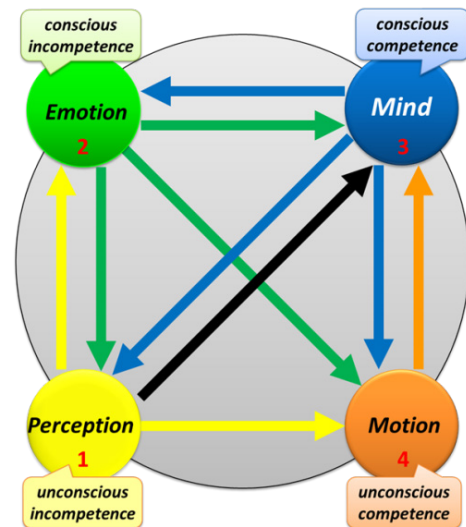
Figure 1. Way competency (skill).

1. A person observes the activities of other people, but in the absence of motivation he does not realize the need for such a skill for himself.

2. Awareness of the need causes emotion and the desire to receive this skill.

3. A person studies information and acquires knowledge, which leads to the ability (conscious competence) to perform the necessary actions.

4. Training and practice lead to the formation of a skill that allows performing actions without conscious control, automatically. Practical applications for real problems lead to better skills and increased skill.

This is the classical way of conscious learning. It is declared in most educational systems and organizations, but, unfortunately, in the absence of motivation for students, a formal approach to the subject of instruction and a minimum of practice, it degenerates into a formal educational path typical for formal educational structures (path 1-3 in Figure 1, shown by the black arrow).

The knowledge obtained in this way is not supported by practice, causes a minimum of interest and does not imply subsequent practical use. Therefore, after a formal check and receipt of marks, they are usually very quickly forgotten, and this all ends.

In Figure 1, we can observe other ways that lead to the desired skill, and correlate them with known or new training approaches.

So, direct way 1-4 is a way of imitation, copying. Very simple and effective, as it involves specially designed for this brain mechanisms - mirror neurons. Unfortunately, it is used mainly only with young children and primary school students. Although in some areas it can give better and faster results in comparison with the theoretical approach and for adults.

The path of imitation, reinforced by motivation and interest (1-2-4), we called modeling. It is well represented in the well-known classical studies (for example, in the concept of social training Bandura [5]).

Path 2-4 is a path of trial and error, possible in the absence of sources of information and a sample for modeling. He is also the path of insight or insight in the implementation of new, previously non-existent activities.

Also on the scheme, you can observe more exotic, but at the same time more efficient ways. For example, as in the approach of Galperin [3] 2-1-4-3-4: motivation - demonstration - independent execution in reality - pronouncing out loud - talking to oneself - full automation of the skill.

Of special interest is the path of visual modeling proposed by the authors of this article [6] (1-2-1-4). In this case, the verbal, abstract description of the mode of activity is replaced by a schema, a visual model, which serves as an indicative basis for the activity being mastered see Figure 2.
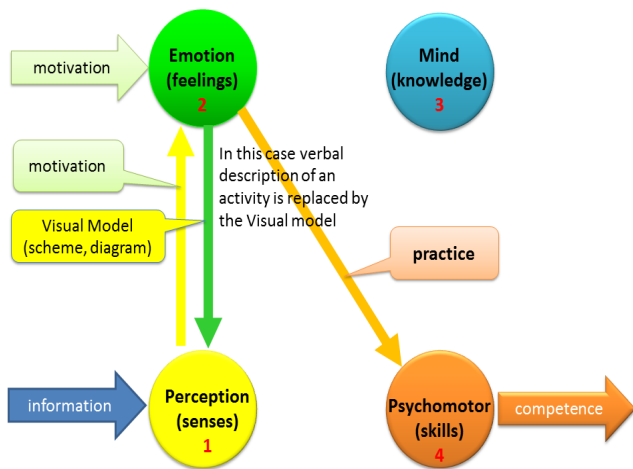
Figure 2. Proposed way of competency.

Such an approach has recently been rapidly gaining popularity in connection with the development of computer systems, visual interfaces, visualization tools, infographics and methods of visual thinking.

3. Methods for Improving Adult's Foreign Language Skills

The objective of this project is to elaborate a new inductive methodology of language skills on the basis of the Structural-Visual method (SVM) and the Visual-Auditory Shadowing method (Nakayama & Mori, 2012) [7]. The Structural-Visual Method is a new inductive language learning methodology, based on mapping of the structure of linguistic knowledge in a graphic form using color to encode the most common patterns [2]. The models thus obtained replace the textual explanations (rules) in the formation of the corresponding skills. The Visual-Auditory

Shadowing method (VAS) is also an inductive language learning methodology which facilitates learning of phonological knowledge (pronunciation) and ideographical knowledge (spelling) [8].

3.1. Improving listening skills

Shadowing is the best method to develop a phonological loop to form the skills of understanding by using communication of auditory images with speech movements. according to Tamai (2005) [9], Shadowing is "an act or a task of listening in which the learners track what he/she heard in speech and repeats it as accurately as possible while listening attentively to the in-coming information (p.34)", the shadowing method is effective in improving phonological loop process, and it leads to improvement in listening skills.

VA shadowing method is a combination of auditory shadowing and visual shadowing. Auditory shadowing requires so-called online processing of auditory input, which requires listening and repetition at the same time. Visual shadowing requires learners to read text aloud employing online processing [10] according to Nakayama (2017), VA shadowing method better improves learners' listening skills, compared to shadowing method alone.

By using auditory images allow one to understand different voices and accents, it is played an important role to implement the concept of "Repetition without repetition", where each presentation of the same material is produced by different voices, with different speed, tone, intonation and accent, is of interest.

For more rapid formation of sound filters and pronunciation patterns, it is proposed to use the technologies Text-to-Speech, voice recognition, audio file decoding and other achievements of computer technology. The need for multiple repetitions can be used to simultaneously develop both speech skills and linguistic competencies. For this purpose, a combination of the VA shadowing technique, enhanced by the use of computer technologies, and the technique of Structural-Visual modeling.

3.2. Improving language skills

In the study of language, the language is the subject of research, and at the same time the research tool, which is already methodologically incorrect. The researcher is produced in the native language of the researcher, and the grammar structure of this language inevitably mediates the way and structure of his thinking. Therefore, language implicitly is also a method of investigation. The result of the research is usually a scientific publication, which is a language product - text. Thus, in the study of thinking, language, language activity and teaching methods, language is also a subject of activity, an instrument of activity, a mode of activity and the result of activity. What cannot but lead to confusion, contradictions and excessive complexity due to logical closure and looping. To eliminate this confusion, it is suggested to make a description of the structure of the language from the sphere of the same language and apply the Visual Meta-language. Moreover, information about the structure and laws of processes is not coded by words and terms, but by the parameters of abstract visual objects and figures-by the color, shape, size, relative location, boundaries, etc., as well as by special signs and symbols.

Particularly productive was the use of this approach to explain to the student the structure of the language being studied and the principles of constructing sentences. As you know, conscious practice is much more effective than mechanical copying and repetition. This is repeatedly confirmed in the studies and underlies the theory of successive formation of mental actions Galperin [3] and developed on the basis of its methods of practical development of skills in various fields of activity [3,11]. But in the field of language teaching, this approach does not work because of the above contradiction - the language must simultaneously be both a subject of educational activity, and an instrument of this activity, and the way of this activity, and the result of this activity. If the student does not know how to express an idea and form a phrase in the language he is studying, he cannot do this because of this ignorance. And if he got the knowledge how to do this, with the help of rules and terms, he still cannot do this because the speech area of the brain, which should implement the act of speaking, is busy with the language information on how to carry out this activity.

The proposed visual approach to the coding of grammatical information allows us to translate the orientation basis of activity, information on how to properly perform this activity, from the language, verbal form to the visual, visual. Releasing this speech zone from the functions of planning and controlling utterances and creating conditions for easy and unhindered performance of speech activity. Language activity itself is meaningless. Language is an instrument for thinking and communicating, and manifests its properties only in such activities. Therefore, the study of language as an end in itself is also unjustified and illogical. It is necessary to teach not language, but activity through language, that is, thinking and communication, which is emphasized in the documents of the Council of Europe [12].

Psychological science did a great deal in the 20th century to study human activity, and to find the patterns of obtaining skills for its implementation (Skinner [13], Bandura [5], Leontiev [14], Galperin [3], etc.). The obtained data made it possible to develop sufficiently detailed theories and, on the basis of them, to create systems for accelerated training of specialists in specific industries (the army, special services, certain spheres of production and technology, large corporations). The closed nature of these areas and the lack of interest in disseminating such experience have led to the fact that these studies are unknown to the overwhelming majority of specialists engaged in the development of similar systems in other industries and countries. This forces them to use outdated, inefficient models and approaches developed for other conditions and tasks and applied in pedagogy because of the conservatism of educational systems.

Galperin [3] in his theory of the step-by-step formation of mental actions pointed to the need for an indicative stage in mastering the skill of mental action, emphasizing the importance of auxiliary tools that facilitate this orientation - the so-called schemes of the Basis of Activity.

A similar approach in Western pedagogy is widely known as Instructional scaffolding. This theory was developed around the same time by (Ninio & Bruner) [15].

Both these theories are based on Vygotsky's idea of a zone of proximal development [16]. Additional support tools and instructions can be provided through the sensory, motor and verbal channels.

A similar phenomenon was discovered in the experiments of Zhinkin [17] in the middle of the last century, and the low effectiveness of grammatical rules in the mastering of oral speech was repeatedly declared by many linguists and psychologists (Krashen [18], Pinker [19] and etc.).

To resolve this discrepancy, suggest applying SVM. SVM in linguistics is a layout of the structure of linguistic knowledge in graphical form using the color of visual objects to encrypt the most common patterns. The obtained models thus replace the textual explanations (rules) in the corresponding skill formation. For this it is proposed to use another part of the psyche - a nonverbal visual system. The one that is usually located in the right hemisphere of the brain.

The higher efficiency of this method compared to other Visual Aids follows from the features of the functioning of the human visual system, shown in recent studies by Kozlovsky [20]. In them, the conceptual model of Baddeley's work memory [21] was confirmed at the physiological level, in particular the part where the presence of visual working memory is declared, which functions independently of verbal working memory ("visual-spatial matrix").

The visual approach to learning foreign language removes the main contradiction of the grammatical method - grammar rules block speaking. If the student does not know how to say the phrase in English, he cannot say anything, because he does not know how to do it. But if you learned the rule how to do it, you still cannot say a phrase. Because that area of the brain, which is responsible for speaking, is occupied with this rule.

SVM show the structure of the sentence in a visual way, freeing the speech zone from the functions of planning and control of speech in the studied language.

The priority direction is the development, research and application of visualization, visual representation of information, visual thinking in different spheres of activity. LingvoMap - a visual representation of the structure of the English sentence and a wide range of tools based on them. The Structural-Visual Method (SVM) is a method of visual structuring of knowledge in the domain by means of color coding of the properties of basic elements. The method allows you to make the complex simple by showing what other people are just telling about.

SVM –it is a set of simple schemes of 3 types, showing:

- Diagrams - the structure of relations between elements of the English language.
- Dynamics - the structure of the Russian language and its relations to the described processes and phenomena.
- Models - the structure of the English proposal and how to build it.

This is the main training material for training on Visual-Body Modeling of the structure of the English language.

Visually - The body method of modeling the structure of the English language radically changes the very approach to the study of English grammar.

The structure of the division of the grammatical system into its component parts is changed, as is the sign system of the description of this structure.

Verbal rules are replaced by diagrams:

- Abstract scientific names of forms and phenomena are replaced by specific visual parameters of the graphic elements of the schemes - color, shape, location.
- Instead of searching for direct links between grammatical forms of two languages that are extremely complex, ambiguous and not formularized, an intermediate sign system is introduced in the form of abstract graphic images uniquely related to real situations in a certain context.
- Instead of classifying grammatical phenomena in form, another structure is introduced, where the elements are linked by the general logic of development of processes and meaning.
- Instead of analyzing abstract sentences from educational texts, specific physical actions of the teacher and student are performed, which are carried out in reality.

Logical operations to understand structure and connections are carried out not on the verbal plane, but with the help of figurative and subject-active thinking.

We developed the LingvoMap - a visual model of the structure (grammar) of the English language.

This technique [22] replaces:

- Complicated rules - easy and comprehensible schemes.
- Abstract theory – practical and comfortable tools.
- Formal book phrases - usual physical actions.
- Hundreds of pages of verbal description - a complete and harmonious system, located on one sheet.

Without going into the theory, we can create phrases of any complexity.

This will allow you to understand how the English language is organized in just a few hours. And hundreds of hours of saved time you can send to the practice of phonetics, vocabulary and conversation.
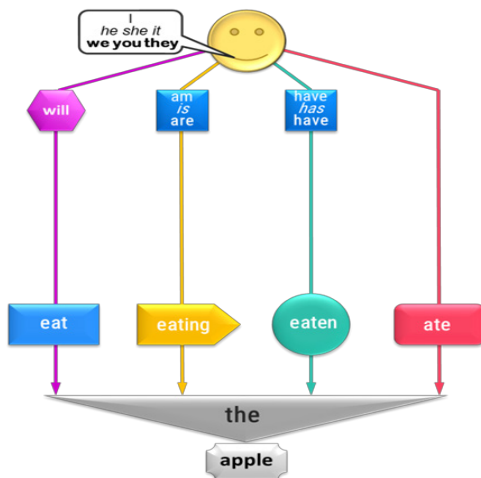
Figure 3. Model of a simple narrative sentence for a verb of action.

Models contain the essential set of temporary structures and development of the kinds of supply required for this level of

language capability or as per the educational programs of this training course. For instance, the least difficult model for building a basic account sentence in 4-time forms for an activity verb appears in Figure 3.

Figure 4 shows a more detailed model that allows you to build on it proposals of different forms (narrative, interrogative, negative) and for more times.
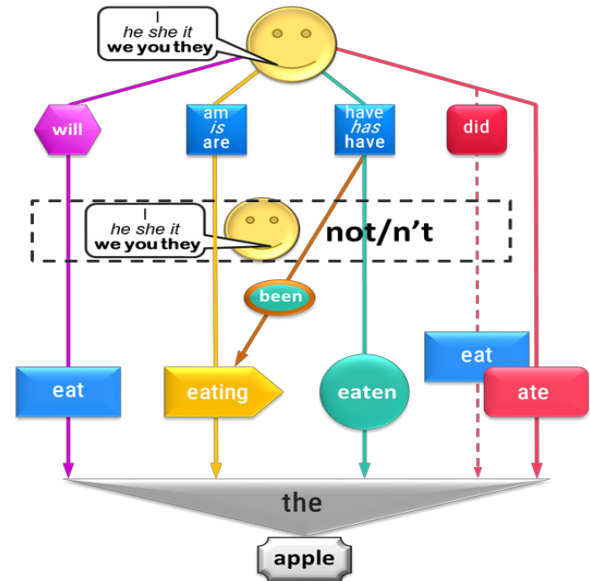
Figure 4. Model for different forms of sentences with action verbs.

In Figure 5 shows the full model of all types of active forms of active voice. It can be applied at higher levels of grammatical competence to systematize knowledge and understand the complete structure of a system of times.
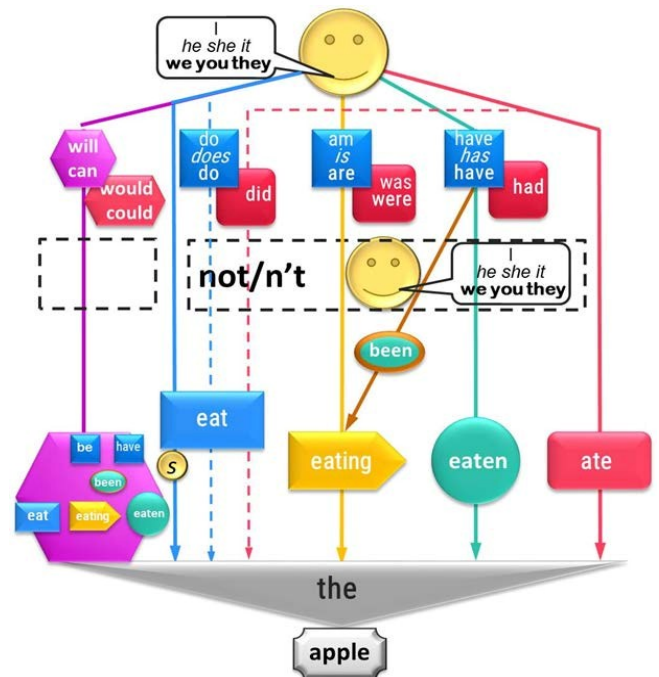
Figure 5. Full model of all active lien times.

Also, models have been created to study individual grammatical topics – passive voice, modal verbs, impersonal verbs, conditional sentences, and many other difficult to understand topics.

3.3. Training skills

The research method is to conduct classes with control groups of trainees in various methods of teaching foreign languages. Based on the results of each lesson, a detailed statistical analysis of the results is carried out, dynamic learning curves for each student are displayed, coefficient tables are established and the speed of the formation of listening comprehension skills is determined, speech levels from the initial to the threshold level of spontaneous speaking are determined in accordance with the CEFR [12] scale.

The peculiarity of the proposed approach is the logical interconnection of the entire language system, the economics of the time of mastering the material and the dynamics of the use of language structures. The achievement of results should be achieved by maintaining the level of the student's effort at a sufficiently long interval of time (months, years), which plays a decisive role in foreign language learning, "Figure 6".

The research will enable to achieve the following goals:

1. Identify the mathematical patterns of the formation of basic language skills; test some hypotheses of linguists about the nature of these processes (Krashen [18], Pinker [19], etc.).

2. Integrate models of neurolinguistics and cognitive psychology into pedagogical practice and methodological development.

3. Create effective computer tools for measuring, managing and controlling basic language skills.

4. Create a framework for the effective interaction of representatives of computer technologies and the humanities.
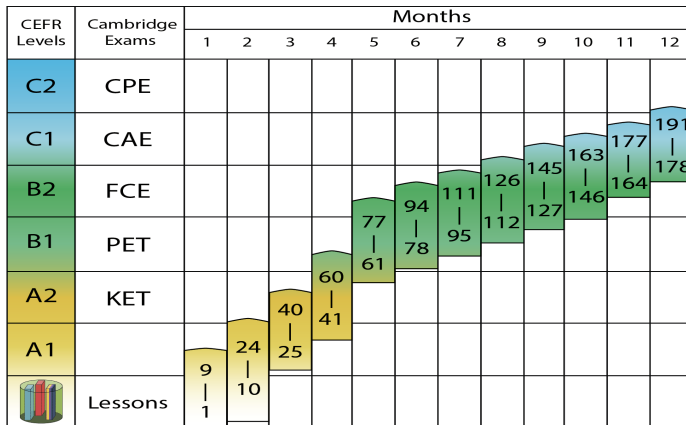
Figure 6. An approximate comparison of the number of lessons in E-AMS with CEFR levels and the Cambridge Common English exams.

After passing the training, without going into the theory and without using formal rules, you will be able to:

- Correctly formulate your thoughts in English;

- To see the logic of application of different designs, their interrelation and correspondence to different contexts;

- Make phrases of any complexity, in all persons, pledges, times and forms;

- Use as learning materials not abstruse textbooks and tedious texts, but your favorite films, serials, songs, books;

- Freely transform any sentence from these sources into hundreds and thousands of others grammatically correct;

- Independently control the correctness of their statements in conversations, letters, texts, translations and exercises and not depend on teachers, teachers, translators.

Training, like any other theoretical preparation, requires the material to be fixed with practical exercises in the proportion of "1 hour of theory - 10 hours of practice"

Without long training, the knowledge and skills you have gained will not do you any good. Language is a skill and all its aspects should be brought to full automaticity!

The fact that new tools allow you to get a skill easier and faster does not eliminate the need for practice, but only increases its effectiveness.

3.4. Methodology of the Study

Learning activities for mastering the language are also infinitely complex and have an unlimited number of parameters and connections. Therefore, to manage this activity it is necessary to apply an abstract scientific model with a minimized number of parameters. In [23] we distinguish the following with minimum set of events.

Set of events (Single event): A single event is one and the same event, the development of which in time is described by the same verb in different forms.

For example:

I will eat an apple. I am eating the apple. I have eaten the apple. I ate the apple.

The minimum set of vocabulary: We confine ourselves to the most minimal set of vocabulary, sufficient for conducting measurements.

In Figure 3 a minimal set of vocabulary is used:

- A complete set of personal pronouns;

- Minimal set of the same verbs (1-8 pieces);

- Minimal set of similar objects (for example, fruits) (2-10 pieces);

- Service words, without which it is impossible to do (articles, auxiliary verbs);

The minimum set of grammatical: To measure the speed of acquiring a skill, let's select the minimum unit of the grammatical skill - this is the ability to compose a phrase according to one template (one form of the sentence for one-time form and one person).

For example, the first template:

This is an apple. This is a pear. This is a peach.

Second template:

I am eating an apple. I am eating a pear. I am eating a peach.

Changing one parameter of the template, you can measure the speed of acquiring lexical skills (learning words) - as in the examples above or grammatical skills on an established set of types:

-Change by persons (conjugation):

I am eating an apple. He is eating an apple.

She is eating an apple. It is eating an apple.

You are eating an apple. We are eating apples.

You two are eating apples. They are eating apples.

-Change in the stages of development of the event (times), initial dynamics [23] (4 elements):

I'll eat an apple.

I am eating the apple.

I already have eaten the apple.

I ate the apple.

-Change in the form of the object (pragmatics):

He is eating a peach.

He isn't eating a banana.

Is he eating a peach? Yes, he is.

Is he eating a banana? No, he isn't.

Who is eating a peach?

What is he doing?

What is he eating?

As we can see, this grammatical set forms 8 x 4 x 7 = 224 variants of constructions for describing a single-type event.

4. Structure of New LMS for IT Development

In the modern era of information technology development, the information process penetrates into all spheres of human activity, including education. There is an improvement and mass dissemination of modern information technologies (IT).

The relevance of the study of the problem of using IT in the teaching of foreign languages is that information technologies have high communicative capabilities, promote the development of knowledge and skills of speaking and listening skills, actively include them in educational activities and effectively develop the skills of communicative competence. All this is necessary for a successful life in the modern world. It should also be noted the enormous popularity of the Internet and computer technologies among young people.

The purpose of the study is to consider the possibilities of using IT in the teaching of EFL. IT include various software and hardware devices and devices that operate on the basis of computer technology, as well as modern means and information exchange systems that provide collection, storage, storage,

production and transmission of information which help to facilitate of production new type of software for the self-teaching the languages.

The use of computer technology contributes to the removal of the psychological barrier of the student on the way to the use of a foreign language as a means of communication. One of the manifestations of this barrier is the so-called fear of error. Trainees note that when using computer technology, they do not feel uncomfortable, making mistakes, and receive fairly clear instructions on how to overcome them.

It should be noted that ITs are not only a means of supplying material, but also a means of control. They provide high quality presentation of the material and use various communication channels (text, touch, graphic, sound, etc.). New technologies allow individualizing and intensifying the learning process. The trainee can choose his educational route and move along it at a convenient pace. A differentiated approach creates the conditions for successful activity of each student, provoking positive emotions and, thus, increasing his educational motivation. Another positive aspect of the use of IT in the learning process is its ability to make the student's assessment more objective (assignments with pre-determined evaluation criteria make it possible to avoid the subjectivity of the assessment).

To control the process of formation of language and work skills of adults, the authors recommend a conceptual solution in the form to create a training system [24]. The generalized structure of the new generation LMS is shown in Figure 7.
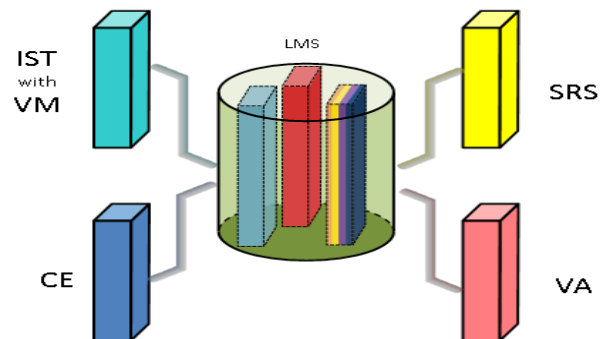
Figure 7. LMS Generalized structure.where: IST with VM – Interactive Speech Trainers with Visual Models, SRS – Speech Recognition System, CE – Continuous Evaluation,VA – Virtual Assistant

This new training system [8] will integrate the following points:

1) Methodical principles substantiated in the works of Bandura [5] and Galperin [3].

2) Unification of visual models and interactive visual-auditory tracking, generate a synergistic effect, both in the first stage of mastering a foreign language and in a "barrier to overcome" phase.

3) Use the achievements of the field of information technology as a tool to ensure the implementation of learning objectives with continuous monitoring of the current situation and obtain the results of learning assured in a limited number of steps.

4) Further development of the main components of the LMS is carried out in the following interrelated areas.

5) It is proposed to use SVM in combination with the modified VA method. Using a new type of LMS provides:

- Step-by-step deceleration and acceleration of the speed of speech in a programmatic way to facilitate the formation of correct pronunciation.

- Application of reference sound templates and speech simulators.

- Repetition without repetition – use for each approach of a material changed by some parameter (voice, tempo, height, lexical or grammatical transformations).

Setting up a subsystem for continuous assessment and real-time learning management will allow us to create logarithmic support for the learning curve and to compensate for the basic requirements of the learning curve to deteriorate to expected loss of efficiency.

The use of modern information technology with the use of effective skills acquisition models allows to reduce the impact of various psychological obstacles and thus accelerate learning and improve its success by conveying a synergistic effect to all phases of language skills [24], especially "barriers to overcome" (Figure 8).
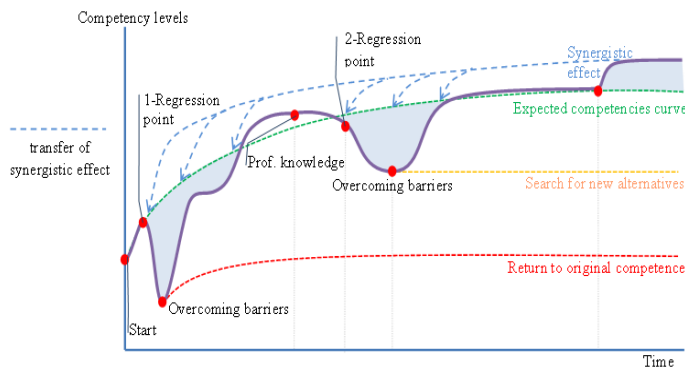
Figure 8. The learning curve with the transfer of synergistic effect.

Training organized in this way will provide the process of controlling the formation of speech skills to a threshold level, allowing the transition from studying the language to improve the process of use. Measuring and verifying the digital parameters of this process will allow access to the data needed to build an automatic language acquisition control system, which in fact opens a new direction to research the construction of the LMS.

5. The Proposed Approach

The result of our analysis is a link between theory and practice, IT and humanities field, science and social processes to create a new generation of Learning Management System (LMS), this system currently under construction, so the requirements of this software are collected; now we are going to a stage of designing and development.

The main contributions of this paper are:

- Working with the real-time voice processing and voice recognition that provide user inputs and system outputs to reach a correct pronunciation.

- We present an API for JavaScript specifically designed to support building such design for "LingvoMap" (Map for construction a sentence (SVM)).

- The effectiveness of this approach by conducting a several prototypes of "LingvoMap" developed using the proposed API requiring well-timed inputs and outputs in real-time.

- This prototype of software runs on both Desktop and Android platforms, which corroborate that cross-platform, web LMS to accelerate learning for learner inputs provided via voice and recognition processing and view the action images.

The software developers formulated several requirements for this system that will be developed in accordance to:

- Availability is the ability of the system to locate and access to training components from a remote access point, and to supply them to other points.

- Adaptability is the ability to adapt the curriculum according to the needs of organizations and individuals.

- Efficiency is the ability to increase productivity, reducing the time and costs of delivering instructions.

- Longevity is the ability to meet new technologies without additional and expensive refinement.

- Interoperability is the ability to use learning materials regardless of the platform on which they are created.

5.1. Prototype of Platform with Speech Recognition System

In order to achieve the prototype [25], we consider the following steps:

Database Design: It is a relational data tables contain the information about lessons to extract the information to the system and the result about any training lesson will be stored in database.

The implementation of the LMS Platform: The implementation of the Platform will be using API for C# and JavaScript, as the HTML5 standard emerged, new opportunities for the development of robust and efficient APIs that support voice processing, and this powerful, widely adopted standard includes interactions with different media, protocols and programming languages [26]. An HTML5 page can process voice and recognition captured directly from devices possibly available on the user's hardware. In addition, WebGL programs consist of control code written in JavaScript and special effects code (shader code) rendering in an HTML, use developments element to draw WebGL graphics for visual programming of the (SVM) flexibility with the introduction of programmable visual effects and these interact with the web page by means of scripts.

Adapt the speech recognition system in the LMS system: We suggest a mechanism to facilitate interaction between the user and the platform through an interface that uses voice recognition. In our platform, the template consists of an interactive image and SVM representation, which reflects the internal structure of the issue of voice recognition. This module will be integrated into the server and also linked to the ASR system. To do this, the following operations are required:

- Voice-to-text: The process used for adapt the voice recognition system is the ASR. We will focus specifically in API Google Speech to Text (STT) Service because it is a cloud computing system and does not compromise the performance of the local computer.

- Text-to-Speech: To convert Text to Speech (TTS), we propose to use the cloud services of Google. Using Google Translator, however, is a less complex process compared with ASR.

5.2. System Architecture

In "Figure 9" we see the system architecture and explained as follow:

Information Request and Display Module (IRDM): In this module the images that represent the action of sentence are displayed as action image and SVM, and this module allows the user to control by the options of program through the keyboard and computer screen, this information is sent to a Webserver. Once information is processed, the requested are shown to the user as data lesson.

Knowledge Data Formation Module (KDFM): The module consists of a database schema that contains all the relational tables, and is used to describe and represent area of knowledge such as data lessons. Database contains the stored procedures to provide a way to create the content of lesson under user requests depend on the selected options. Web server contains the API for the client part to transfer the received data from database server into a graphical interface for client, the connection from web server to database using data access layers these layers contain connection manger and business layers which contain a set of objects (Classes) that return the data as dataset to be used by web services and JavaScript and WebGel with the use the services of Google for speech and recognition.

Speech Recognition Module (SRM): This module allows to the learner to interact in our Platform by using voice. For instance, the system returns the information to the SRM in order to be converted to speech using TTS (Text-to-Speech). The user speaks for information in the microphone and such consult is converted to text in the ASR system. In web server the java script functions check from this input data if it is correct to the displayed text at IRDM. Finally, the feedback of correct speaking will be saved in database for evaluation.
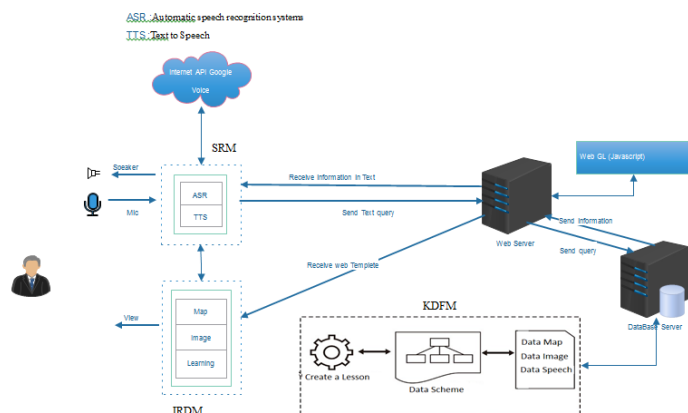
Figure 9. System Architecture

5.3. Working of Software

An educational object of LMS is any educational material that can be displayed in a web browser (for example, texts, pictures, voice and Map, web pages), as well as any combination thereof, intended for educational purposes and collected together in a special way. In addition, the browser must be implemented and enabled for the JavaScript language. Thus, this software standard describes database for training materials contains:

- Educational materials in the form of data in relational tables.
- Arbitrary dynamic content: JavaScript code and other objects that can be displayed inside the browser. Dynamic content can inform the LMS of the student's progress.
- All study materials are structured (i.e. break into lessons or level of Lessons).
- The description of the material flow sequence is indicated. For example, a specific text should be provided to the learner only after he has read other texts or passes the test.

The main flow of events: begins when the user intends to work with training courses containing SVM-objects and voice recognition. The system should offer one of the following options:

1. Load data lesson from a database.

2. Load the image and "LingvoMap".

3. The simulator should speak, and the learner should listen and vice versa, the learner should speak, and the simulator should understand it and give feedback.

4. The result of training will be saved in database to display the level of training at learning curve.
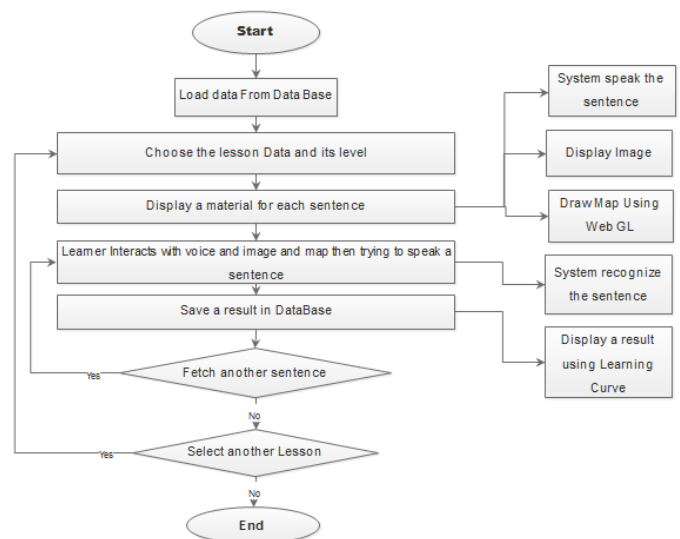
Figure 10. Flow Chart of a lesson

In the IST (Interactive Speech Trainers), the data will be loaded from database into the interface, we set the interface to select the lesson then choose the level of each lesson, so the level will be determine the change of model of structural visual model, each level of lesson contain the structure of (SVM) start from simple then going to full model of (SVM) due to the structure of lesson and its words, so the data of lesson will be set from back-end content management system (CMS), in "Figure 10" we see the start

workflow of the process of the lesson , the material of a lesson's level contain set of sentences with its components that will be generated automatically on running the system (Images ,Sound, SVM), the user will interact with these material by hearing and speaking a sentence and save the feedback in Database to display these results in learning Curve.

The User Interface of the webpage is shown in Figure 11, which let the user login to the system. When the login is a success then the user can access to material that contains a list of lessons for the user to select the target lesson.

Visually, the description of LMS can be represented as a set of lessons, each describing a certain part of learning process "Figure 6".

Figure 11. Set of Lessons.

From a list of lesson that displayed on screen, select any lesson to load the content and to display the SVM and images of lesson, "Figure 12".
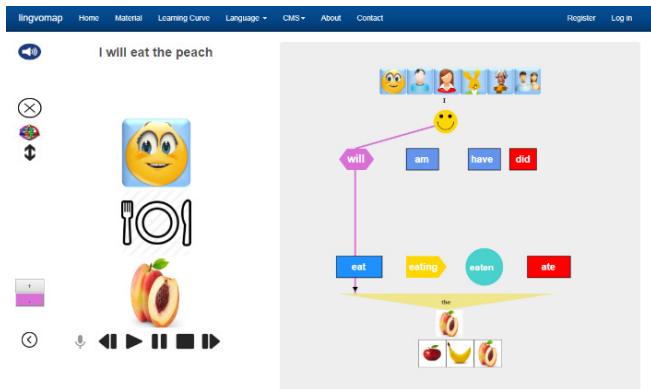
Figure 12. Interface of IST

In addition to providing training courses, the Continuous Evaluation subsystem of the lessons can inform us about the results of the training to do more evaluation by user to reach the satisfactory level "Figure 8". For example, when completing a task, the student received a certain number of points and spent a certain amount of time performing; Learning curves are adequate for evaluating performance improvement due the positive effect of learning.

5.4. Content Management system

Managing Web content is becoming a top priority for improving business performance. Careful planning and elaboration of this process is required, and timely and prompt

updating of the information content of the resource plays one of the main roles in the success of the project.

The educational content of Current software is understood as the set of educational lessons collected in sentences, SVM, action images, learning result. These content units are designed as a template in such a way they can be used repeatedly in a different lesson. For example, once the template of lesson content is created, it can then be used in any new lesson with a change of words, when a need arises.
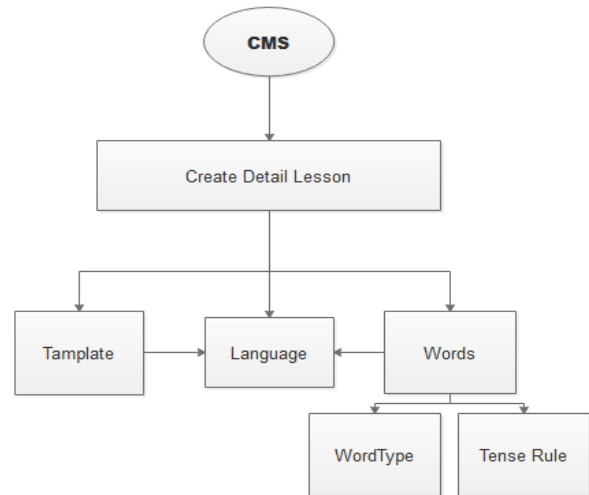
Figure 13. Flow Chart of CMS.

To Control the content and images of lesson, we need to fill the data for each lesson, in "Figure 13" we have the flow charts of lesson that composed of Template, language and words the words related to word types (Verb, object, …) and Tense Rule (Past simple, Future,…).

To create the lesson, there are main components that will be considered in order to complete the structure of the lesson, where Template, Language and words form the complete data of lesson.

Template: it is a set of structures and form type (positive sentence, negative sentence, question, …) of a sentence, structure is a set of combination of word type (Subject, Verb, Object, Auxiliary, …) by aggregation of these word type by add symbol '+' between them for example (Subject+Verb+Article+Object), by generate the function of a given template of a lesson we get a set of lesson data from this structure by replace subject by (I,He,She , …) and Verb by (eat)and Article by (the) and Object by (Apple) replace '+' by ' ' the result is a sentence (I eat the apple) , for instance the set of the structure for English language for a given lesson at a given level as (Subject+Auxiliary+Verb+Article+Object, Subject+Verb+Article+Object, Subject+Auxiliary+not+Verb+Article+Object, Auxiliary+Subject+Verb+Article+Object+?, ….).

6. Conclusion

In this paper, we have discussed the different techniques and new methods of learning to be used in current technology framework for acquisition a language and also reviewed the results obtained from previous papers on this subject.

The application of this approach explaining linguistic phenomena and teaching foreign languages, proposed in this work, allows us to translate the planning and control of speech activity into the right hemisphere and thus greatly simplify and simplify the process of teaching languages.

A prototype was created. To date, we get an encouraging result about the experiments 300 students during the training related with the development of this LMS system.

Exploratory confirmation of preparing materials, test systems and proposed educational programs components of (SVM) on a restricted gathering of pupils demonstrated outcomes like those acquired through the quick advancement of other training methods for the Galperin theory [3]. There was a decrease in preparing time to perform particular work in (3-30 times) and increment in preparing accomplishment by 10-25% to 80-95%. These are primer outcomes that require extra check and independent testing.

Based on this approach, we are developing a new generation of training and computer programs and we are working to carry out intensive testing and collection of statistical data. Visual models and pedagogical theories, which have received a new incarnation in them, open other horizons and scale prospects for improving educational technologies.

References

[1] I. H. Tahini and A. K. Dadykin, "A study of new techniques for learning management system to accelerate language acquisition using structural visual models," 2018 Sixth International Conference on Digital Information, Networking, and Wireless Communications (DINWC), Beirut, 2018, pp. 92-97.doi: 10.1109/DINWC.2018.8357002.

[2] A.K. Dadykin, V.A. Dibrova and I.H. Tahini, "The Visual Approach in Educational Projects," 4th International Conference on Education and Psychological Science (ICEPS 2017), Barcelona, Spain, February 12-14, 2017 / International Journal of Social Science and Humanity (IJSSH, ISSN: 2010-3646, DOI: 10.18178/IJSSH).

[3] P.Y. Galperin , "Psychology of thinking and teaching about the gradual formation of mental actions", Research in the thinking of Soviet psychology. Moscow, 1966.

[4] A.H. Maslow, (1987). *Motivation and Personality. (3rd ed.).* New York, NY: Harper & Row. pp. 117-118.

[5] A. Bandura & R.H. Walters, (1963). Social learning and personality development. New York: Holt, Rinehart, & Winston.

[6] A.K. Dadykin, VA Dibrova, Generalized visual model of the structure of the educational process // Modern trends in secondary adult education: Materials of III International. scientific method. Conf., Minsk, October 21. . 2016 .: 2 hours - Minsk RIVSH, 2016 - Part 1. - P. 62-67.

[7] T. Nakayama, T. Mori, Efficacy of Visual-Auditory Shadowing, Tokyo: J-Stage, 2012, v 42, 55-68 p. ISSN I0917-3536.

[8] I.H. Tahini, T. Nakayama, V.A. Dibrova, and A.K. Dadykin, "Cognitive Psychology Models and Approaches to Develop Language Skills," 5th International Conference on Education and Psychological Sciences (ICEPS 2018), Seoul, South Korea, January 27-29, 2018, International Journal of Information and Education Technology (IJIET, ISSN: 2010-3689, DOI: 10.18178/IJIET).

[9] K. Tamai, "Risuningu shidoho to shiteno shadoingu no koka ni kansuru kenkyu [Research on the effect of shadowing as a listening instruction method]," Tokyo: Kazama Shobo, 2005.

[10] T. Nakayama, "Efficacy of Visual-Auditory Shadowing Method in SLA Based on Language Processing Models in Cognitive Psychology," Tokyo: Kaitakusha, p.110, 2017.

[11] B. Badmaev, Psychology and methodology of accelerated learning. M.: Humanité. Ed. Center VLADOS, 1998. - 272 p. ISBN 5-691-00102-7.

[12] Council of Europe (2011). Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Council of Europe.propositions of Principia Mathematica and related systems I]. Monatshefte für Mathematica und Physik. 38: 173–198.

[13] B.F. Skinner, The Technology of Teaching, 1968. New York: Appleton-Century-Crofts Library of Congress Card Number 68-12340 E 81290 ISBN 0-13-902163-9.

[14] A.N. Leontiev, "Activities. Consciousness. Personality". Moscow, 1975.

[15] A. Ninio and J. Bruner, "The achievement and antecedents of labeling," Journal of Child Language, No. 5, pp. 1-15, 1978.

[16] L. Vygotsky, "Thought and language," Cambridge, MA: MIT Press, 1934/1986.

[17] N. Zhinkin "About code transitions in internal speech," Questions of linguistics, No. 6, pp. 26-38, 1964.

[18] S.D. Krashen, "Principles and Practice in Second Language Acquisition," University of Southern California, p.202.

[19] S. Pinker, "The Language Instinct: How the Mind Creates Language," William Morrow and Company, p.483, 1994.

[20] S.A. Kozlovsky, "Psychophysiological mechanisms of preservation of visual images in working memory," Available: http://www.dissercat.com/content/psikhofiziologicheskie-mekhanizmy-so khraneniya-zritelnykh-obrazov-v-rabochei-pamyati#ixzz4zGVJphTy.

[21] A.D. Baddeley, "The episodic buffer: A new component of working memory? ," Trends in Cognitive Sciences, No. 4, pp. 417-423, 2000.

[22] I.H. Tahini, "New Structure of Learning Management System for Formation of Language Skills by Using Visual Approach", V International Scientific and Technical Internet Conference, November 18-19, 2017 Section Information Technologies in Production and Scientific Research, BNTU,2017.

[23] I.H. Tahini, A.K. Dadykin and V.A. Dibrova. "The Model of Change as the Basis of the Knowledge Structure in the Next Generation E-LMS", 10th annual International Conference of Education, Research and Innovation, 2017, Seville, Spain (ISBN: 978-84-697-6957-7 / ISSN: 2340-1095, doi: 10.21125/ 10.21125/iceri.2017.1325). – Pages: 5022-5032.

[24] I.H. Tahini, A.K. Dadykin, V.A. Dibrova, "Control System of Foreign Language Training Based on Structural-Visual Techniques," Proceedings 4th International Conference on Computer Science Computer Engineering and Education Technologies (CSCEET2017), Beirut/Lebanon: pp.149-154, 2017.

[25] I. Tahini, A. Shaparenko, A. Golikova, A. Dadykin. Management System for Forming Language Skills of Adults – New Generation LMS / EDULEARN 2018 Proceedings 10th annual International Conference on Education and New Learning Technologies, pp. 5061-5071, ISBN: 978-84-09-02709-5, July 2018, Palma de Mallorca , Spain, ISSN: 2340-1117.

[26] S. Fulton, and J. Fulton, " HTML5 Canvas," O'Reilly Media, Incorpo- rated, 2013.

Enhancing the Energy/Power Efficiency of a DC Distribution Grid for Residential Buildings via Modular Architecture of DC/DC Solid State Transformers

Faizan Dastgeer[*,1], Hassan Erteza Gelani[1], Faisal Ali[1], Zahir Javed Paracha[2]

[1]University of Engineering and Technology, Lahore, Department of Electrical Engineering, FSD Campus, Pakistan

[2]The University of Faisalabad - Amin Campus, Faisalabad, Pakistan

A B S T R A C T

DC power is apparently attempting to breach into the power system - the system which it once lost to the AC power paradigm. DC is already present in the generation, transmission and utilization sides of the system, leaving distribution as the only area where it has not shown any significant presence – this may be regarded as still in research phase. The current work aims to investigate efficiency enhancement of a DC power distribution system for residential buildings via the use of a modular approach for the DC/DC solid state transformers (SSTs). Specifically, DC distribution system shows reduced efficiency at light loads because of the lowering in efficiency of its constituent SSTs. Via the use of modular approach, a single SST is composed of smaller modules which may be turned on or off individually, for the purpose of improving the converter efficiency. In this paper, we utilize two methods of load division upon SST converter modules. As the efficiency of an SST is increased, the overall efficiency of the system increases because bulk power passes through these transformers.

1. Introduction

Electrical energy, apparently being the most easily transferable (from one place to another) and convertible form of energy is, therefore, the most common source of energy for powering our residential and commercial buildings. Fundamentally, this energy has two forms available – AC and DC; and out of these, AC has been the dominant scenario for the power grid for the many past decades. This is in spite of the fact that the earliest power system was DC in nature. However, with the advent of electronics, DC is appearing again. It is present in the form of personal computers, laptops, and various other electronic gadgets. Variable speed drive (VSD) based air-conditioners [1], [2] also require DC as an intermediate stage between supply and load. The modern trend of LED-based lighting[3], [4], for our home and office buildings requires DC power as well. DC is also present on the energy generation side in the form of Solar Photovoltaic (PV) cells and wind farms with AC/DC/AC conversion [5], [6] , and on the bulk power transfer side in the form of High Voltage DC (HVDC) power transmission [7]–[9] with practical solutions available via Siemens [10] and ABB [11]. One may say that DC is struggling

again to obtain a share of the electrical energy system.

Residential (and commercial) power distribution is one segment of the power system where DC has not yet obtained any significantly extensive appearance. DC power distribution may currently be regarded as in the research phase; wherein a number of relevant publications may be found in literature, some of the publications in the recent past are[12]–[20]. One of the important parameters of the system is its efficiency. Mathematically, efficiency may be expressed as:

$$\eta_E = \frac{E_{out\_sys}}{E_{in\_sys}} \times 100 \tag{1}$$

$$\eta_P = \frac{p_{out\_sys}}{p_{in\_sys}} \times 100 \tag{2}$$

where η_E and η_P are energy and power efficiencies of a system. For a system with constant energy demand, both the quantities will be the same. The efficiency factor has a history with DC. It was apparently the system efficiency that led to the first overcoming of DC by AC in the 1880s, and then again it was apparently the efficiency that led to the re-appearance of DC in

[*]Faizan Dastgeer, University of Engineering and Technology, Lahore, Department of Electrical Engineering, FSD Campus, Pakistan.
Email: faizandastgeer@uet.edu.pk
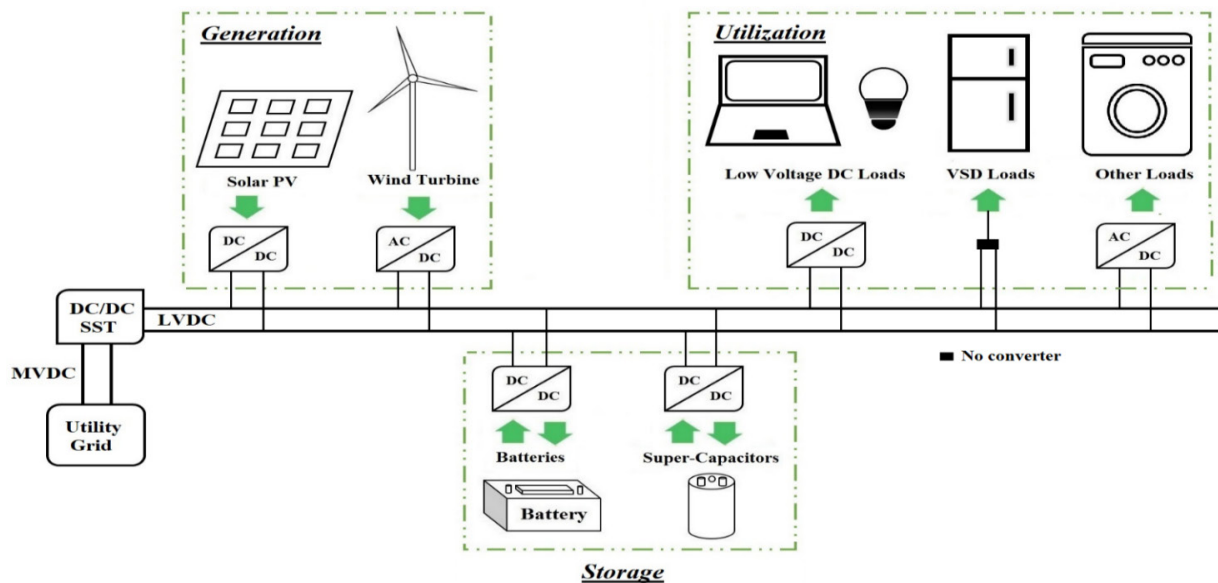
the power system in the form of HVDC transmission. And now the efficiency of DC power distribution for a variety of cases is in the research phase. If DC can win the domain for power distribution, then an all DC power system may be envisioned. This may be especially true for the future micro-grids and a DC microgrid where DC generation can come from solar PV while the energy distribution and utilization are DC as well may become a reality. The concept of DC microgrids has seen a significant amount of research efforts. References [21]–[28] are some of the recent publications that mention DC microgrids.

The current effort is directed towards the efficiency of a DC distribution grid for residential buildings. DC/DC solid state transformers (SST) serve to replace the conventional transformers in such a system, and the efficiency of these power electronic transformers is crucial for the efficiency of the overall system. This research, which may be considered as an extension of our earlier efforts [16] and [17] aims to investigate the efficiency improvement of the system gained via the use of a modular architecture for the DC/DC SST. This architecture can allow the use of a required number of modules only in the converter and lower the converter losses leading to a higher value for the converter as well as system efficiency.

1.1. Contribution and Novelty of this work

The contribution of this effort is the enhancement of a residential DC grid efficiency at reduced system loading via a modular approach for system SSTs. DC distribution systems show a decrement in efficiency as the load reduces due to the efficiency variation of constituent DC/DC power electronic SSTs. Various studies related to the efficiency of DC systems are mentioned in the literature review subsection of this paper; none of the research efforts that we are aware of attempts to improve the residential DC distribution system efficiency via modular approach for the DC/DC SSTs. The next section discusses some of the applications of DC power distribution.

2. DC Distribution – Applications

2.1. Distributed Power Systems

Traditionally, DC has been present as a medium of power transfer for the case of distributed power systems (DPS) [29]–[33]. The power system of the international space station (ISS) is an example of this. The DC microgrid concept can be considered as a modified version of the DC DPS [34].

2.2. DC Microgrids

Microgrids may be described as a concept that can bring together distributed generation, residential/commercial utilization as well as hybrid storage technologies while having appropriate control to allow both utility-grid connected or independent operation. Figure 1 shows a schematic diagram of a DC microgrid. The extension of this concept to DC power may be considered as natural as:

i. Distributed generation via solar and wind (having AC/DC/AC conversion) produce DC power.
ii. A wide variety of electronic loads in homes and offices require DC power.
iii. Storage technologies of batteries, super-capacitors and fuel cells [35], [36] are DC in nature.

The concept of DC microgrid has seen various research efforts as cited earlier, some of these will be discussed in the subsequent section.

2.3. Collector Parks for Renewable Energy

The concept of DC power collection has seen quite a number of research efforts for off-shore wind farms [37]–[42]. In these systems, power is collected from various off-shore wind turbines and then sent on-shore via DC lines. This way, the charging current and the corresponding losses that would have appeared if the line had been AC, can be avoided. The concept of DC power collection has been proposed for Solar Parks as well [43].

2.4. Miscellaneous

The concept of DC power distribution has been there for telecommunication facilities [44], [45] as well as data centers [46]–[49]. Reference [50] mentions a practical DC data center. Furthermore, with the advent of Hybrid Electric Vehicles, the need for their charging stations arises and DC bus based charging stations for the electric vehicles have been mentioned [51]–[54].

3. Literature Review / Background of the current effort

This section mentions various publications related to DC power distribution, especially its losses and efficiency. A background for the current work is also presented here. The subsequent subsection mentions the publications around/before 2012, while the subsection after it mentions the publications after 2012.

3.1. Past – Around/Before 2012

Reference [55] was an earlier effort of one of the current authors. DC power distribution paradigm was compared with a counterpart AC system with respect to system efficiency. Both Medium voltage (MV) and Low voltage (LV) portions were considered in the study. DC was found to be more efficient than AC; furthermore, a mathematical portion was presented to determine the minimum required efficiency for power electronic converters (PEC) in a DC system. However, this work did not discuss any method to increase the efficiency of PEC so as to meet the minimum requirement. The efficiency enhancement of the system via an increase of efficiency of constituent DC/DC power electronic SST is the topic of the current work.

D. Hammerstrom compares DC and AC distributions for residential buildings in [56]. The results of this work are based on a number of stages of energy conversion – conduction losses are not included. The results showed that for the case of DC power being supplied via a rectifier, DC does not have an advantage over AC. However, if a local DC generator is present, higher efficiency will be obtained for DC.

Reference [57] presents the feasibility of DC for commercial facilities – power losses (conduction losses; PEC losses have not been mentioned explicitly) have been presented for AC and DC distribution systems. The authors mention that it may be concluded that DC can lead to big advantages. Reference [58] mentions that 380V DC brings reliability and efficiency to sustainable data centers. It further mentions that for data centers 380V DC is 28% more efficient than 208V AC and 7% more efficient than 415V AC. Reference [59] was an efficiency comparison of AC and DC power distribution for data centers. A small-scale demonstration showed 7% saving in input power for a 400V DC distribution as compared to 480V AC system.

The authors of [60] present a proto-type DC distribution system for residential applications where a hybrid energy source is present. A photovoltaic – wind – fuel cell hybrid energy system was developed and natively DC loads were used. Voltage and current results for the system were presented and a life-cycle cost was also given. However, the work did not include an efficiency study of the system. Amin et. al give energy consumption and losses for some common household appliances for LVAC and

LVDC systems in [61]. They mention that total energy consumption is lowest for their 48V DC system.

Stark et. al present a loss comparison for AC and DC distribution in [62]. Both MVDC and LVDC portions have been included. The authors conclude that AC and DC distribution have the same merit if loads are half AC and half DC. The authors mention that the DC-DC converter is the defining device in the possible application of DC distribution. They also mention that one problem with these converters is a reduction of efficiency when operated below rating. This is precisely the issue which the current research effort may be able to address; via the use of modular approach for the main DC-DC power electronic SST, the SST, as well as the system efficiency, may be kept from lowering too much when the system load is reduced.

3.2. Present – Post-2012

The authors of [63] determine the savings in energy via the use of DC distribution for U.S. net-metered residential buildings. They present comparative results of energy savings for AC and DC building power distribution systems. They assume that distribution losses for both cases are comparable and all appliances are assumed to be DC-internal. They conclude that direct-DC could give significant energy savings in the U.S. houses with net-metered photovoltaic systems. This work was detailed enough to include not only the PEC losses occurring in the system but also the variation in PEC efficiency by variation of load. At partial loading, the PEC efficiency reduces. As mentioned earlier, the current effort is an attempt to avoid the lowering of SST PEC as well as system efficiency due to the reduction in system load.

The authors of [64] give comparative results of system efficiency for DC and AC power. Their analysis for office building indicates an efficiency gain of 14.9% when supplied by DC power as compared to AC power. The authors of [65] give an efficiency comparison of AC and DC power distribution for a commercial building. They conclude that for applications with AC and DC power sources with self-consumption and frequent battery usage, DC distribution may improve the system efficiency as compared to AC but the improvement is only 1.3%.

Fregosi et. al discuss a DC microgrid in [66]. They mention that an equivalent AC system was installed next to the DC system for efficiency comparison and the latter was found to use photovoltaic energy 8% more effectively. The authors of [67] give an efficiency comparison of AC and DC microgrid for the residential load. They assume that the residential load comprises equal proportions of AC and DC loads. For their comparative study, the authors conclude that the overall efficiency of the AC system was higher than that of DC; however, the addition of PV generation led to a reduction in AC system efficiency and an increase in efficiency of DC system.

A comparative analysis of system efficiency for AC and DC residential power distribution paradigms [16] was an effort of two of the current authors; related to DC distribution efficiency. Comparative system efficiency results can be presented for a variety of parameters of the systems such as voltage levels etc.; in [16] the comparative results of system efficiency were presented

for variation in PEC efficiency values. The first part of the work showed that AC had better merit than DC. In the second part, we included the concept of Variable Speed Drive (VSD) based air-conditioning leading to a large increase in DC power demand of residential loads. DC distribution showed a slightly higher performance than AC for this case. A schematic diagram of the DC distribution with VSD loads is shown here in Figure 2.

The highest power loss in this DC system was found to be occurring in the primary DC/DC SST. The current research effort aims to reduce the losses occurring in this converter by using a modular approach based architecture for it.
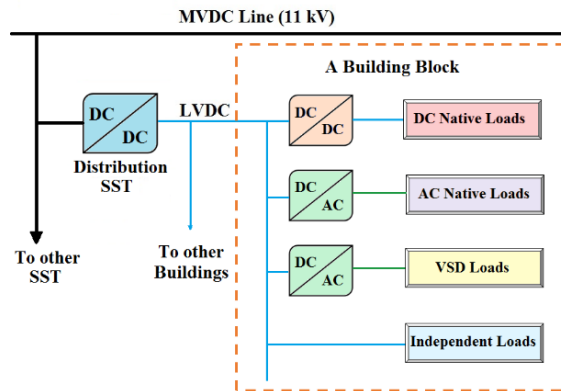
Reference [15] gives a detailed efficiency comparison of DC and AC power distribution system. This work is aimed at commercial buildings with the concept of zero net energy (ZNE) in mind; where solar energy generation, as well as battery storage, are present on-site. The loads are assumed to be natively DC. Results are presented for two scenarios: small office having 48V DC system and medium office having 380V DC system. In the conclusion section, the authors mention that their research has found that the baseline efficiency savings are 9.9% and 11.9% for small and medium office buildings respectively; while the best-case scenario gives savings of 17.9% and 18.5%. They further mention that their study confirms that DC distribution is best suited for (commercial) buildings having the large solar capacity and large battery as well as a high voltage distribution backbone. The presence of battery is apparently a crucial factor for this research. The authors mention at one place that one of their results shows that a battery-less ZNE building barely benefits from DC distribution.

Reference [17] was another of earlier effort of two current authors related to DC distribution efficiency analysis. We performed the analysis for U.S. residential loads while taking into account load variation and the corresponding converter efficiency variation for the DC/DC converter transformer. A complete day was divided into three portions: Night (00:00 – 06:00), Day-1 (06:00 – 15:00) and Day-2 (15:00 – 00:00) and residential load was divided into these portions with Night, Day-1 and Day-2 demanding lowest to highest power respectively. An efficiency curve was chosen for the DC/DC converter transformer and its efficiency values were 86%, 89% and 90% for Night, Day-1 and Day-2; correspondingly, the system efficiency values were

82.85%, 85.64% and 85.73% for the three modes respectively. We went on to perform an analysis for the division of losses occurring in the system. The results are presented here in Figure 3.

The results depicted an increase in the DC/DC transformer losses (as a percentage of total losses occurring in the corresponding time) with the reduction of system load. We mentioned in the future work that, "A future work may be to investigate the feasibility of a DC power distribution system which comprises a modular architecture based DC/DC converter transformer." In the current effort, we move towards DC distribution system efficiency enhancement via modular arrangement for its DC/DC SST. Prior to our work, [13] and [68] present an efficiency improvement of DC networks. Reference [68] improves the efficiency via optimal operation of multiple distributed generators in the system, while [13] contributes via proposing an enhanced tool for optimizing the analytical design of DC network that focuses on energy savings and improving efficiency. Our original contribution is the enhancing system efficiency of a DC grid at reduced loading via the use of modular SSTs. This modular approach and the system modeling will be presented in the subsequent section of this paper.
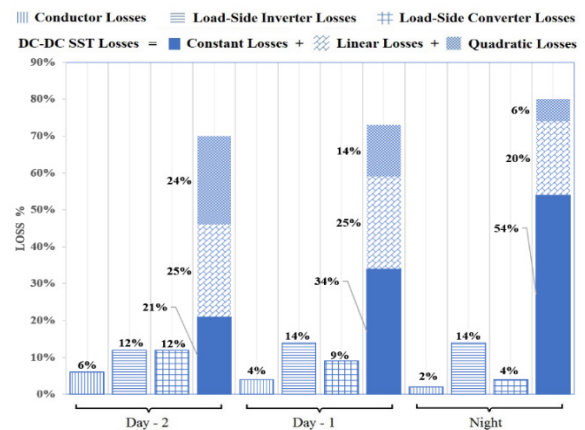
4. Modeling of the DC Distribution Grid and Loads

4.1. Residential Load Modelling

The residential load modeling of a typical US building was presented by us in [16] using data from Energy Information Administration and Department of Energy; the model is discussed here briefly with Table 1 giving residential energy splits divided into A, D and I categories. These stand for AC native, DC native and Independent loads (such as Electric Iron which may be energized by AC or DC power provided suitable voltage level is supplied).

Based on data of Table 1 and taking space heating and cooling as DC power demanding loads, a single building the block may be built. As in [16], we assume 21 buildings to bunched together and connected to a single distribution DC/DC SST. However, in contrast to [16] which used 230V as the in-building distribution voltage for DC (as well as AC systems to get a fair comparison), here we choose 380V as the distribution voltage which may draw
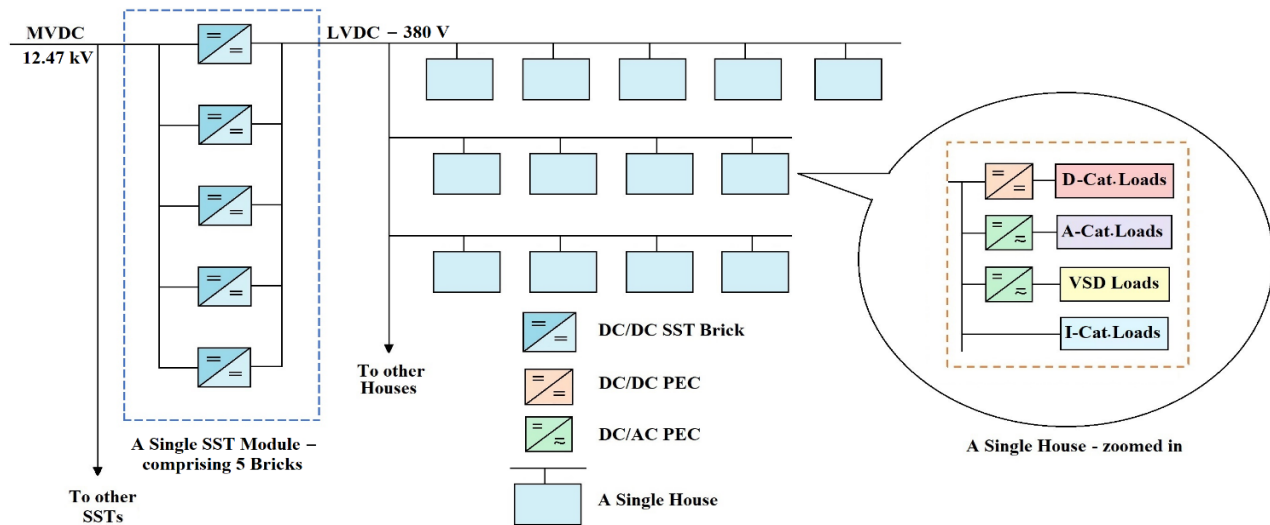
Figure 4: Schematic diagram of DC distribution grid with modular SST

Table 1. Residential energy splits, categories & energy demand [16]

Category	Energy Used (Quad. Btu)	Energy %	A D I	Energy kWh	A cat. kWh	D cat. kWh	I cat. kWh
Space Heating	0.42	8.78	I	2.64	---	---	2.64
Water Heating	0.48	9.93	I	2.98	---	---	2.98
Space Cooling	1.02	21.24	A	6.37	6.37	---	---
Lighting	0.53	11.04	D	3.31	---	3.31	---
Refri-geration	0.45	9.45	A	2.84	2.84	---	---
Electronics	0.33	6.86	D	2.06	---	2.06	---
Wet Cleaning	0.33	6.80	A	2.04	2.04	---	---
Cooking	0.11	2.36	I	0.71	---	---	0.71
Computers	0.19	3.95	D	1.19	---	1.19	---
Other	0.94	19.69		5.91	1.97	1.97	1.97
Total	4.795			30.03	13.22	8.52	8.29
Power demand in kW					0.55	0.36	0.35

4.2. Modular Architecture of DC/DC SST

Here the SST has been used with a modular architecture. Figure 4 shows a schematic diagram of a portion of the DC distribution system with modular architecture for the SST.

In [17] we used a single graph for modeling efficiency variation of a DC/DC PEC being used as SST. In this research, we have used five graphs from various sources to model SST efficiency variation. These graphs, presented in Figure 5, will be used one at a time. Two of these [69], [73] are from practical market available

DC/DC converters while the remaining three are from research literature [70]–[72]. MATLAB/Simulink has been used for the simulations in this study where a single simulation experiment determines the total input power P_{in_total} demanded by the system. The system efficiency η_{sys} is then determined by

$$\eta_{system} = \frac{P_{o_{total}}}{P_{in_{total}}} \qquad (3)$$

where P_{o_total} refers to the total power demand of the residential loads. For a single simulation experiment, a single graph out of the five reference curves given in Figure 5, is used to model the efficiency versus load variation characteristic of the modules called bricks of the SST.

5. Simulation runs and efficiency result

In contrast to [17] where we made everyday usage assumptions to include load variation and only three steps (Night, Day-1 and Day-2 as mentioned earlier) were used; over here we take ten steps of load variation – 10% to 100% of its value, as given in Table 2.

5.1. Basic Load Division Algorithm

To start with, the algorithm for load division upon bricks is kept simple. The idea is to keep one brick for taking up all the load variation while keeping all the others (if being used) at their maximum efficiency. Following are the steps for the basic algorithm.

a. Load one brick only if total load is less than its Maximum efficiency point (MEP).
b. If load exceeds this value, then keep one brick at its maximum efficiency point, and put all the remaining load on second brick
c. If load exceeds MEP of the second brick, keep the second brick at its MEP and put all extra load on third brick and so on.
d. If the MEP of all the converter bricks has been attained, beyond this, divide all the extra load equally among the bricks.

292
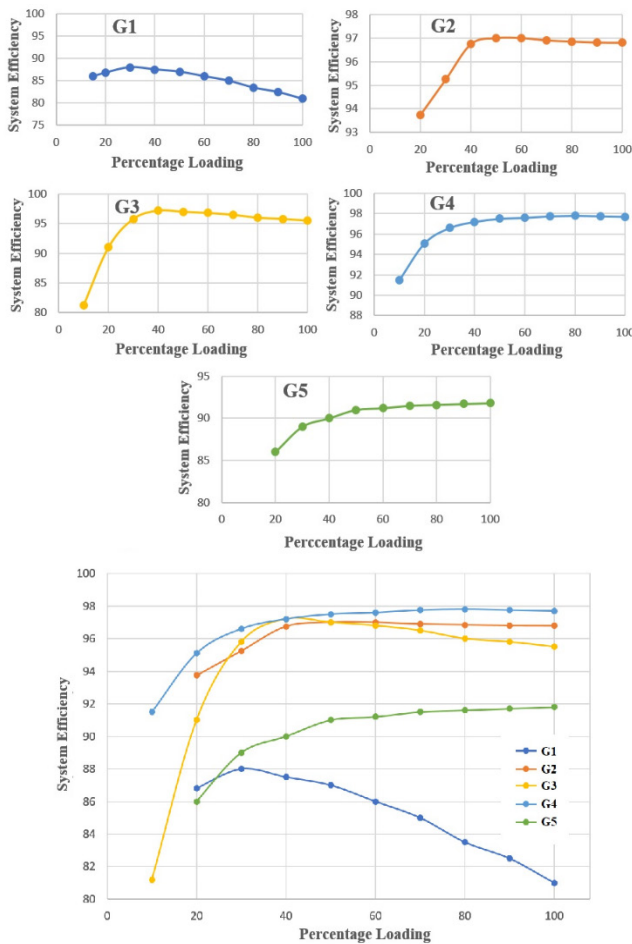
Figure 5:
1) G1- PowerStax F351 Full Brick 350W DC/DC PEC [69]
2) G2- Boost Mode Conversion efficiency curve – with V1 = 23V [70]
3) G3 – Conventional Dual Half-Bridge DC/DC converter [71]
4) G4 – CLLC DC/DC converter [72]
5) G5 - Cosel DBS700B28 - 700W DC/DC converter [73]
6) G1 – G5 combined

As an example, consider G1 graph. Its maximum efficiency point is at 29% loading, so for a single brick, this point occurs at 2.9 kW (since a single brick is rated at 10kW, while the five-brick modular SST is rated at 50kW). Now, as long as the SST load is less than 2.9kW, only one brick will operate. When the load exceeds 2.9kW, the first brick is fixed at its MEP load i.e. 2.9kW, and the next brick makes up for the remaining load. When the second brick load exceeds 2.9kW (i.e. total load exceeds 2 x 2.9kW), second brick is also fixed at the MEP of 2.9kW, and the remaining load (i.e. total load – 2.9kW – 2.9kW) is shifted on the third brick, and so on. Once all bricks are at maximum efficiency point (i.e. total load > 5 x 2.9kW), all bricks are loaded equally after this. The results of system simulation are presented in Tabular as well as graphical form in Table 2 and Figure 6. For comparison, the results of system efficiency with a non-modular SST have also been presented.

As evident from Table 2 and Figure 6, the modular SST based system shows a huge efficiency gain at a very low loading of the system. However, the single SST based system shows higher performance beyond 30% loading up to 60% load, and this may

be attributed to the crude load division algorithm at this stage. After 60% loading, both systems show the same performance.

Table 2. Results of system efficiency – G1 – basic load division algorithm - modular and non-modular SST

Load %	Load P_{o_total} (kW)	Modular SST		Non-Modular SST	
		Input Power P_{in_total} (kW)	System Efficiency %	Input Power P_{in_total} (kW)	System Efficiency %
10	26.28	32.58	80.7	33.81	77.7
20	52.55	66.72	78.8	67.02	78.4
30	78.83	99.36	79.3	99.67	79.1
40	105.1	132.04	79.6	131.71	79.8
50	131.38	164.75	79.75	163.32	80.4
60	157.66	197.06	80	197.06	80
70	183.94	230.72	79.7	230.72	79.7
80	210.21	264.4	79.5	264.4	79.5
90	236.49	298.74	79.2	298.74	79.2
100	262.8	335.52	78.3	335.52	78.3

Figure 6: System efficiency comparison for modular & non-modular SST cases – for G1 reference graph
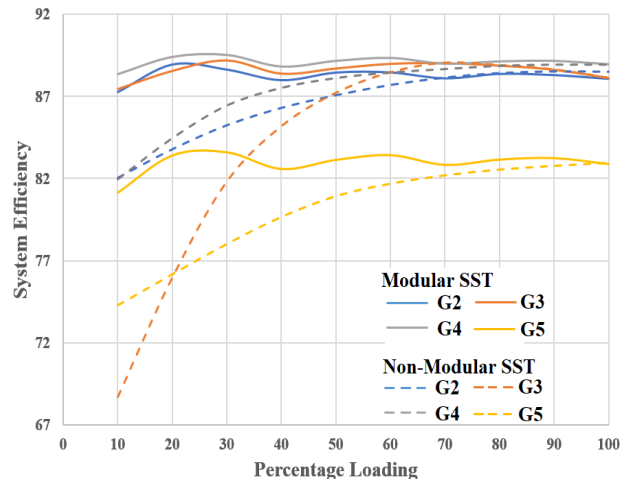
Figure 7: System efficiency comparison for modular & non-modular sst cases – for G2 – G5 reference graphs

Figure 7 shows the system performance comparisons for modular and non-modular SSTs based upon each of G2 – G5 reference curves. On the whole, the system shows superior performance with a modular SST. One reason for the improvement in system efficiency may be the reduction in the constant losses of the SST when it is used in the modular architecture. Assume that *x*% power of the SST rating is being

wasted in the constant losses for the non-modular SST case. But when the SST is shifted to modular architecture; say only one brick is working at that time – then the percentage constant losses are reduced to *x/5* (assuming same PEC efficiency characteristics and for a five brick SST). Hence at reduced load, the operation of only a limited number of bricks saves on losses and enhances overall SST and system efficiency. However, as the load increases, more and more bricks are turned on and eventually with all bricks on, the modular SST gives the same constant losses as a non-modular SST. In the next subsection, we present an improved load division technique to further enhance the system performance.

5.2. Modified Load Division Algorithm

The basic load division algorithm is modified to give better performance for system efficiency. The modified algorithm may be described as –

a. Load one brick only if the total load is less than its MEP.
b. If load exceeds this value, keep one brick at its MEP, and put remaining load in excess load bucket.
c. If load exceeds MEP of the second brick, keep the second brick at its MEP and put an extra load in excess load bucket.
d. If the MEP of all the converter bricks has been attained, beyond this, divide all the extra load equally between the bricks.

In this load division when the maximum number of bricks are at their MEP, the excess load bucket is handled in two ways

1. The remaining load in excess load bucket is divided among all MEP bricks equally.
2. If possible, the remaining load in excess load bucket is loaded on next brick which was not in use previously.

When these two ways of handling the excess load bucket are simulated for the system in parallel. Their resultant modular efficiency of the converters are compared and the one which is maximum is chosen. As an example, consider G1 graph again with MEP of 2.9 kW. As long as the SST load is less than 2.9kW, only one brick will operate. When the load exceeds 2.9kW, the first brick is fixed at its MEP load i.e. 2.9kW, and either the next brick makes up for the remaining load or remaining load is loaded on the first brick which was at MEP. The more efficient way of the two is selected. When the second brick load exceeds 2.9kW (i.e. total load exceeds 2 x 2.9kW), second brick is also fixed at the MEP of 2.9kW, and the remaining load (i.e. total load – 2.9kW – 2.9kW) is either shifted on the third brick or divided equally between MEP bricks and efficient way is selected with if condition in the algorithm. Using the same method other bricks are loaded Once all bricks are at maximum efficiency point (i.e. total load > 5 x 2.9kW), all bricks are loaded equally after this.

The results of system simulation for G1 are presented in Tabular as well as graphical form in Table 3 and Figure 8. Furthermore, the bulk of results for G2 - G5 are presented as before in a combined graph form in Figure 9. The non-modular SST based system efficiency results in Figure 9 are the same as

those in Figure 7 and have been presented for the purpose of comparison.

Table 3. Results of system efficiency – G1 – modified load division algorithm - modular and non-mo dular SST

Load %	Load P_{o_total} (kW)	Modular *SST*		Non-Modular *SST*	
		Input Power P_{in_total} (kW)	System Efficiency %	Input Power P_{in_total} (kW)	System Efficiency %
10	26.28	32.58	80.7	33.81	77.7
20	52.55	65.2	80.6	67.02	78.4
30	78.83	97.87	80.55	99.67	79.1
40	105.1	130.59	80.5	131.71	79.8
50	131.38	163.3	80.4	163.32	80.4
60	157.66	197.06	80	197.06	80
70	183.94	230.72	79.7	230.72	79.7
80	210.21	264.4	79.5	264.4	79.5
90	236.49	298.74	79.2	298.74	79.2
100	262.8	335.52	78.3	335.52	78.3

Figure 8: System efficiency comparison for modular & non-modular SST cases – for G1 reference graph
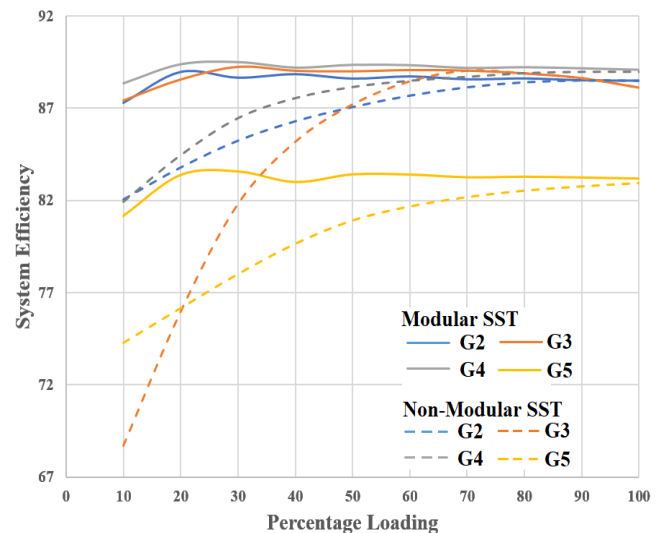
Figure 9: System efficiency comparison for modular & non-modular SST cases – for G2 – G5 reference graphs.

As can be observed by the mutual comparison of Figure 8 and Figure 6, the advanced load division technique keeps the Modular SST based system efficiency higher than the non-modular case for a longer range of system loading. Furthermore, Table 4 presents the comparative results of system efficiency when used with modular SST with basic and advanced load division methods for G2-G5. Although the advantage is slight, still the advanced technique gives better results at times. In any case, the modular SST based system proves to show higher system efficiency as compared to the non-modular SST based system.

It may be deemed that it is difficult to compare the results of system efficiency using five efficiency curves for the DC/DC SST among which two are market available converters while other three come from the research literature. Although published, the research literature used may have been using different criteria for their work. However, the aim of this paper is to present the idea of efficiency enhancement via modular architecture for the SSTs. Mutual comparison of the results is not the main concern – hence the use of disparate efficiency characteristics for the DC/DC SST should not be a problem

6. Future Work

This work may be extended to a detailed optimization technique for load division upon individual bricks of the modular SST with a goal to achieve the maximum overall system efficiency. Here we briefly discuss one such concept based upon Lagrange Multiplier.

The goal is to find such a load division of individual converter bricks which will lead to the lowest input power for the modular converter while supplying the desired load power. For this optimization, it may be more suitable to use a P_{in} versus P_o graph rather than a P_o vs efficiency graph. Such a graph for the G1 reference curve (presented in Figure 5) is presented here in Figure 10.

Assume that the general expression for this graph may be given as a quadratic equation as

$$P_{in} = \alpha + \beta P_o + \gamma P_o^2 \tag{4}$$

where α, β and γ may be derived via curve fitting for the graph, P_o is the output power of a single brick and P_{in} is its input

power. Then for the modular SST comprising 'n' bricks, it may be written that

$$\sum_{i=1}^{n} P_{in} = P_{in_{total}} = \sum_{i=1}^{n} (\alpha_i + \beta_i P_{oi} + \gamma_i P_{oi}^2) \tag{5}$$
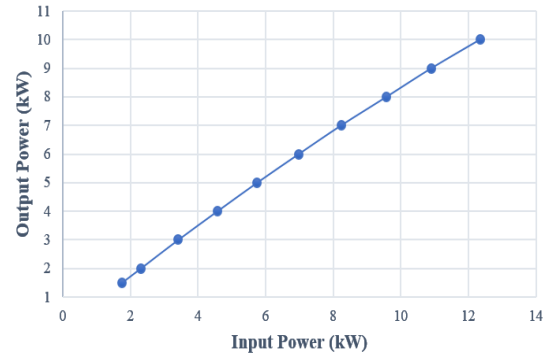
Figure 10: Input power versus output power graph – for G1 reference gap

The $P_{in\text{-}total}$ becomes the objective function whose minimum value needs to be determined. Also, if P_D is the total demand power at any given time from the modular SST, it can be written that

$$\sum_{i=1}^{n} P_{oi} = P_D \tag{6}$$

Now, by augmenting the objective function with the constraint, a Lagrange multiplier may be formed as

$$L = P_{in_{total}} + \lambda(P_D - \sum_{i=1}^{n} P_{oi}) \tag{7}$$

Such an analysis may be probed in order to derive the optimum value for P_{oi} which will give the least $P_{in\text{-}total}$ while supplying the required P_D, hence giving the highest value for

$$\eta_{SST} = \frac{\sum_{i=1}^{n} P_{oi}}{P_{in\_total}} \tag{8}$$

where η_{SST} is the net efficiency of the modular SST.

Besides this, the areas of efficiency analysis and efficiency enhancement of DC distribution systems may be applied to a DC grid composed of the modern Zero Net Energy (ZNE) buildings whereby the DC/DC SSTs will be bi-directional allowing forward as well as the reverse flow of power. ZNE is a new trend for the modern home and DC distribution may facilitate the implementation of this concept.

Table 4. Power demands of A, D & I categories during three-day portions

% Load	Non-Modular SST System Efficiency				Modular SST – Advanced Division System Efficiency				Modular SST – Basic Division System Efficiency			
	G2	G3	G4	G5	G2	G3	G4	G5	G2	G3	G4	G5
10	82.06%	68.7%	81.93%	74.26%	87.28%	87.43%	88.35%	81.14%	87.28%	87.43%	88.35%	81.14%
20	83.79%	75.97%	84.45%	76.15%	88.95%	88.56%	89.39%	83.39%	88.95%	88.56%	89.39%	83.39%
30	85.24%	81.81%	86.45%	78%	88.64%	89.24%	89.51%	83.58%	88.64%	89.21%	89.51%	83.58%
40	86.29%	85.18%	87.53%	79.65%	88.83%	89.03%	89.21%	83%	88%	88.39%	88.81%	82.57%
50	87.06%	87.21%	88.13%	80.92%	88.59%	89%	89.36%	83.42%	88.46%	88.71%	89.15%	83.12%
60	87.68%	88.46%	88.47%	81.68%	88.71%	89.07%	89.34%	83.41%	88.47%	89%	89.33%	83.41%
70	88.13%	89.04%	88.68%	82.19%	88.55%	89.04%	89.19%	83.26%	88.11%	89.04%	88.98%	82.82%
80	88.4%	88.89%	88.88%	82.54%	88.6%	88.89%	89.23%	83.29%	88.38%	88.89%	89.12%	83.14%
90	88.5%	88.63%	88.95%	82.77%	88.5%	88.63%	89.17%	83.25%	88.31%	88.63%	89.15%	83.23%
100	88.48%	88.12%	88.95%	82.95%	88.48%	88.12%	89.09%	83.19%	88.08%	88.12%	88.95%	82.87%

7. Conclusion

The use of modular architecture can help against the lowering of power/energy efficiency of a power electronic DC/DC SST. Hence the overall system efficiency of a DC grid may be improved via modular SST approach. In the current paper, we have demonstrated this via DC grid simulations for five different efficiency characteristics of the DC/DC SST, two of the graphs were derived from practical market available DC/DC converters while the other three came from research publications. We have two techniques for load division among SST bricks and although, there was not a much significant difference in system performance between the proposed basic and advanced load division methods; nevertheless, the results gave a good performance against a non-modular SST based system. An in-depth optimization of the load division technique may be a future direction for this work.

DC power distribution, given up a long time ago, is now witnessing a revival and the time may come when we are living in DC-based homes, or at least the modern homes with advanced integration of renewable-energy/ZNE concept have shifted to DC power. To this end, however, a lot of research and development may be required which may include a variety of areas, one of which is the efficiency of the system. And the factor of efficiency may be deemed do-or-die for the system, after all, it was this factor that once led to the ousting of DC in the early days of the power system. Again, this factor is important for the re-introduction of DC power in the system in the form of HVDC transmission lines. The ongoing push for higher efficiencies may even be seen in our homes via the acceptance of LEDs for lighting and inverter based (VSD based) air-conditioning technologies.

8. Funding

References

[1] M. B. Yurtseven, E. Erkin, E. Acuner, S. Mete, and S. Onaygil, "An experimental investigation of energy saving potentials for room type variable-speed air conditioners in public offices: A case study from Istanbul," Energy Build., vol. 68, pp. 165–171, 2014.

[2] B. K. Bose, "Power Electronics and Motor Drives Recent Progress and Perspective," IEEE Trans. Ind. Electron., vol. 56, no. 2, pp. 581–588, Feb. 2009.

[3] F. G. Montoya, A. Pena-Garcia, A. Juaidi, and F. Manzano-Agugliaro, "Indoor lighting techniques: An overview of evolution and new trends for energy saving," Energy Build., vol. 140, pp. 50–60, 2017.

[4] B.-L. Ahn, C.-Y. Jang, S.-B. Leigh, S. Yoo, and H. Jeong, "Effect of LED lighting on the cooling and heating loads in office buildings," Appl. Energy, vol. 113, pp. 1484–1489, 2014.

[5] Z. Chen, J. M. Guerrero, and F. Blaabjerg, "A Review of the State of the Art of Power Electronics for Wind Turbines," IEEE Trans. Power Electron., vol. 24, no. 8, pp. 1859–1875, Aug. 2009.

[6] F. Blaabjerg, M. Liserre, and K. Ma, "Power Electronics Converters for Wind Turbine Systems," IEEE Trans. Ind. Appl., vol. 48, no. 2, pp. 708–719, Mar. 2012.

[7] L. Zhao, J. Lu, X. Cui, L. Xie, Y. Ju, and K. He, "The Altitude Effect and Correction of Audible Noise for HVDC Transmission Lines," IEEE Trans. Power Deliv., vol. 32, no. 4, pp. 1954–1963, Aug. 2017.

[8] J. M. Johnson and A. Yadav, "Complete protection scheme for fault detection, classification and location estimation in HVDC transmission lines using support vector machines," IET Sci. Meas. Technol., vol. 11, no. 3, pp. 279–287, 2017.

[9] A. Kalair, N. Abas, and N. Khan, "Comparative study of HVAC and HVDC transmission systems," Renew. Sustain. Energy Rev., vol. 59, pp. 1653–1675, 2016.

[10] "HV Direct Current Transmission System (HVDC)." [Online]. Available: https://www.energy.siemens.com/hq/en/power-transmission/hvdc/. Last accessed date: 31 July, 2017

[11] "HVDC." [Online]. Available: http://new.abb.com/systems/hvdc. Last accessed date: 31 July, 2017

[12] B. Glasgo, I. L. Azevedo, and C. Hendrickson, "How much electricity can we save by using direct current circuits in homes? Understanding the potential for electricity savings and assessing the feasibility of a transition towards DC powered buildings," Appl. Energy, vol. 180, pp. 66–75, 2016.

[13] M. Fantauzzi, D. Lauria, F. Mottola, and A. Scalfati, "Sizing energy storage systems in DC networks: A general methodology based upon power losses minimization," Appl. Energy, vol. 187, pp. 862–872, 2017.

[14] G.-H. Gwon, C.-H. Kim, Y.-S. Oh, C.-H. Noh, T.-H. Jung, and J. Han, "Mitigation of voltage unbalance by using static load transfer switch in bipolar low voltage DC distribution system," Int. J. Electr. Power Energy Syst., vol. 90, pp. 158–167, 2017.

[15] D. L. Gerber, V. Vossos, W. Feng, C. Marnay, B. Nordman, and R. Brown, "A simulation-based efficiency comparison of AC and DC power distribution networks in commercial buildings," Appl. Energy, p. , 2017.

[16] F. Dastgeer and H. E. Gelani, "A Comparative analysis of system efficiency for AC and DC residential power distribution paradigms," Energy Build., vol. 138, pp. 648–654, 2017.

[17] H. E. Gelani and F. Dastgeer, "Efficiency Analyses of a DC Residential Power Distribution System for the Modern Home," Adv. Electr. Comput. Eng., vol. 15, no. 1, pp. 135–142, 2015.

[18] T. Dragicevic, J. C. Vasquez, J. M. Guerrero, and D. Skrlec, "Advanced LVDC Electrical Power Architectures and Microgrids: A step toward a new generation of power distribution networks.," IEEE Electrif. Mag., vol. 2, no. 1, pp. 54–65, Mar. 2014.

[19] M. H. Ryu, H. S. Kim, J. W. Baek, H. G. Kim, and J. H. Jung, "Effective Test Bed of 380-V DC Distribution System Using Isolated Power Converters," IEEE Trans. Ind. Electron., vol. 62, no. 7, pp. 4525–4536, Jul. 2015.

[20] B. Nordman and K. Christensen, "DC Local Power Distribution: Technology, Deployment, and Pathways to Success," IEEE Electrif. Mag., vol. 4, no. 2, pp. 29–36, Jun. 2016.

[21] A. A. Mohamed, A. T. Elsayed, T. A. Youssef, and O. A. Mohammed, "Hierarchical control for DC microgrid clusters with high penetration of distributed energy resources," Electr. Power Syst. Res., vol. 148, pp. 210–219, 2017.

[22] Q. Xu et al., "A Decentralized Dynamic Power Sharing Strategy for Hybrid Energy Storage System in Autonomous DC Microgrid," IEEE Trans. Ind. Electron., vol. 64, no. 7, pp. 5930–5941, Jul. 2017.

[23] N. Yang, B. Nahid-Mobarakeh, F. Gao, D. Paire, A. Miraoui, and W. Liu, "Modeling and stability analysis of multi-time scale DC microgrid," Electr. Power Syst. Res., vol. 140, pp. 906–916, 2016.

[24] C. N. Papadimitriou, E. I. Zountouridou, and N. D. Hatziargyriou, "Review of hierarchical control in DC microgrids," Electr. Power Syst. Res., vol. 122, pp. 159–167, 2015.

[25] Q. Yang, L. Jiang, H. Zhao, and H. Zeng, "Autonomous Voltage Regulation and Current Sharing in Islanded Multi-inverter DC Microgrid," IEEE Trans. Smart Grid, vol. 3053, no. c, pp. 1–1, 2017.

[26] W. J. Ma, J. Wang, X. Lu, and V. Gupta, "Optimal Operation Mode Selection for a DC Microgrid," IEEE Trans. Smart Grid, vol. 7, no. 6, pp. 2624–2632, Nov. 2016.

[27] C. Yin, H. Wu, F. Locment, and M. Sechilariu, "Energy management of {DC} microgrid based on photovoltaic combined with diesel generator and supercapacitor," Energy Convers. Manag., vol. 132, pp. 14–27, 2017.

[28] M. I. Ghiasi, M. A. Golkar, and A. Hajizadeh, "Lyapunov Based-Distributed Fuzzy-Sliding Mode Control for Building Integrated-DC Microgrid With Plug-In Electric Vehicle," IEEE Access, vol. 5, pp. 7746–7752, 2017.

[29] C. M. Wildrick, F. C. Lee, B. H. Cho, and B. Choi, "A method of defining the load impedance specification for a stable distributed power system," in Power Electronics Specialists Conference, 1993. PESC '93 Record., 24th Annual IEEE, 1993, pp. 826–832.

[30] J. Liu, X. Feng, F. C. Lee, and D. Borojevich, "Stability margin monitoring for DC distributed power systems via perturbation approaches," IEEE Trans. Power Electron., vol. 18, no. 6, pp. 1254–1261, Nov. 2003.

[31] X. Feng, C. Liu, Z. Ye, F. C. Lee, and D. Borojevic, "Monitoring the stability of DC distributed power systems," in Industrial Electronics Society, 1999. IECON '99 Proceedings. The 25th Annual Conference of the IEEE, 1999, vol. 1, pp. 367–372 vol.1.

[32] X. Feng and F. C. Lee, "On-line measurement on stability margin of DC distributed power system," in APEC 2000. Fifteenth Annual IEEE Applied Power Electronics Conference and Exposition (Cat. No.00CH37058), 2000, vol. 2, pp. 1190–1196 vol.2.

[33] X. Feng, K. Xing, F. C. Lee, and D. Borojevic, "Individual load impedance specification for a stable DC distributed power system," APEC '99. Fourteenth Annu. Appl. Power Electron. Conf. Expo. 1999 Conf. Proc. (Cat. No.99CH36285), vol. 2, pp. 923–929 vol.2, 1999.

[34] X. Wang et al., "Decentralized Impedance Specifications for Small-Signal Stability of DC Distributed Power Systems," IEEE J. Emerg. Sel. Top. Power Electron., vol. PP, no. 99, p. 1, 2017.

[35] N. A. Ahmed, A. K. Al-Othman, and M. R. AlRashidi, "Development of an efficient utility interactive combined wind/photovoltaic/fuel cell power system with MPPT and DC bus voltage regulation," Electr. Power Syst. Res., vol. 81, no. 5, pp. 1096–1106, 2011.

[36] C. N. Papadimitriou and N. A. Vovos, "Integration of a hybrid fuel cell-battery system to a distribution grid," Electr. Power Syst. Res., vol. 81, no. 7, pp. 1299–1307, 2011.

[37] R. Barrera-Cardenas and M. Molinas, "Comparative Study of Wind Turbine Power Converters Based on Medium-Frequency AC-Link for Offshore DC-Grids," IEEE J. Emerg. Sel. Top. Power Electron., vol. 3, no. 2, pp. 525–541, Jun. 2015.

[38] F. Deng and Z. Chen, "Operation and Control of a DC-Grid Offshore Wind Farm Under DC Transmission System Faults," IEEE Trans. Power Deliv., vol. 28, no. 3, pp. 1356–1363, Jul. 2013.

[39] S. Chuangpishit, A. Tabesh, Z. Moradi-Shahrbabak, and M. Saeedifard, "Topology Design for Collector Systems of Offshore Wind Farms With Pure DC Power Systems," IEEE Trans. Ind. Electron., vol. 61, no. 1, pp. 320–328, Jan. 2014.

[40] J. Robinson, D. Jovcic, and G. Joos, "Analysis and Design of an Offshore Wind Farm Using a MV DC Grid," IEEE Trans. Power Deliv., vol. 25, no. 4, pp. 2164–2173, Oct. 2010.

[41] C. Meyer, M. Höing, A. Peterson, and R. W. De Doncker, "Control and design of DC grids for offshore wind farms," IEEE Trans. Ind. Appl., vol. 43, no. 6, pp. 1475–1482, 2007.

[42] F. Deng and Z. Chen, "Design of protective inductors for HVDC transmission line within DC grid offshore wind farms," IEEE Trans. Power Deliv., vol. 28, no. 1, pp. 75–83, 2013.

[43] H. A. B. Siddique and R. W. De Doncker, "Evaluation of DC Collector-Grid Configurations for Large Photovoltaic Parks," IEEE Trans. Power Deliv., vol. PP, no. 99, p. 1, 2017.

[44] M. Schweizer-Berberich and H. Willmes, "Concept for a 48 V DC Power Supply System with Lithium Ion Batteries for Telecom Applications," in INTELEC 05 - Twenty-Seventh International Telecommunications Conference, 2005, pp. 31–36.

[45] H. J. Chiu, T. F. Pan, C. J. Yao, and Y. K. Lo, "Automatic EMI Measurement and Filter Design System for Telecom Power Supplies," IEEE Trans. Instrum. Meas., vol. 56, no. 6, pp. 2254–2261, Dec. 2007.

[46] E. Taylor, M. Korytowski, and G. Reed, "Voltage transient propagation in AC and DC datacenter distribution architectures," in 2012 IEEE Energy Conversion Congress and Exposition (ECCE), 2012, pp. 1998–2004.

[47] K. Tan, X. Song, C. Peng, P. Liu, and A. Q. Huang, "Hierarchical protection architecture for 380V DC data center application," in 2016 IEEE Energy Conversion Congress and Exposition (ECCE), 2016, pp. 1–8.

[48] B. R. Shrestha, T. M. Hansen, and R. Tonkoski, "Reliability analysis of 380V DC distribution in data centers," in 2016 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT), 2016, pp. 1–5.

[49] C. Wang and P. Jain, "A quantitative comparison and evaluation of 48V DC and 380V DC distribution systems for datacenters," in 2014 IEEE 36th International Telecommunications Energy Conference (INTELEC), 2014, pp. 1–7.

[50] "World's most powerful DC data center online." [Online]. Available: http://www.abb.com/cawp/seitp202/187b2f29acaea090c1257a0e0029fb 1a.aspx. Last accessed date: 31 July, 2017

[51] S. Rivera, B. Wu, S. Kouro, V. Yaramasu, and J. Wang, "Electric Vehicle Charging Station Using a Neutral Point Clamped Converter with Bipolar DC Bus," IEEE Trans. Ind. Electron., vol. 62, no. 4, pp. 1999–2009, 2015.

[52] G. F. Reed, B. M. Grainger, A. R. Sparacino, R. J. Kerestes, and M. J. Korytowski, "Advancements in medium voltage DC architecture development with applications for powering electric vehicle charging stations," Energytech, 2012 IEEE, pp. 1–8, 2012.

[53] C. Capasso and O. Veneri, "Experimental study of a DC charging station for full electric and plug in hybrid vehicles," Appl. Energy, vol. 152, pp. 131–142, 2015.

[54] S. Jung, H. Lee, C. S. Song, J. H. Han, W. K. Han, and G. Jang, "Optimal Operation Plan of the Online Electric Vehicle System Through Establishment of a DC Distribution System," IEEE Trans. Power Electron., vol. 28, no. 12, pp. 5878–5889, Dec. 2013.

[55] F. Dastgeer and A. Kalam, "Efficiency comparison of DC and AC distribution systems for distributed generation," in 2009 Australasian Universities Power Engineering Conference, 2009, pp. 1–5.

[56] D. J. Hammerstrom, "AC Versus DC Distribution Systems - Did We Get it Right?," in 2007 IEEE Power Engineering Society General Meeting, 2007, pp. 1–5.

[57] A. Sannino, G. Postiglione, and M. H. J. Bollen, "Feasibility of a DC network for commercial facilities," IEEE Trans. Ind. Appl., vol. 39, no. 5, pp. 1499–1507, 2003.

[58] G. AlLee and W. Tschudi, "Edison Redux: 380 Vdc Brings Reliability and Efficiency to Sustainable Data Centers," IEEE Power Energy Mag., vol. 10, no. 6, pp. 50–59, Nov. 2012.

[59] A. Pratt, P. Kumar, and T. V Aldridge, "Evaluation of 400V DC distribution in telco and data centers to improve energy efficiency," in INTELEC 07 - 29th International Telecommunications Energy Conference, 2007, pp. 32–39.

[60] E. Cetin, A. Yilanci, H. K. Ozturk, M. Colak, I. Kasikci, and S. Iplikci, "A micro-DC power distribution system for a residential application energized by photovoltaic-wind/fuel cell hybrid energy systems," Energy Build., vol. 42, no. 8, pp. 1344–1352, 2010.

[61] M. Amin, Y. Arafat, S. Lundberg, and S. Mangold, "Low voltage DC distribution system compared with 230 V AC," 2011 IEEE Electr. Power Energy Conf. EPEC 2011, pp. 340–345, 2011.

[62] M. Starke, L. M. Tolbert, and B. Ozpineci, "AC vs. DC distribution: A loss comparison," in 2008 IEEE/PES Transmission and Distribution Conference and Exposition, 2008, pp. 1–7.

[63] V. Vossos, K. Garbesi, and H. Shen, "Energy savings from direct-DC in U.S. residential buildings," Energy Build., vol. 68, no. PARTA, pp. 223–231, 2014.

[64] Z. Liu and M. Li, "Research on Energy Efficiency of DC Distribution System," AASRI Procedia, vol. 7, pp. 68–74, 2014.

[65] J. Brenguier, M. Vallet, and F. VAILLANT, "Efficiency gap between AC and DC electrical power distribution system," in 2016 IEEE/IAS 52nd Industrial and Commercial Power Systems Technical Conference (ICPS), 2016, pp. 1–6.

[66] D. Fregosi et al., "A comparative study of DC and AC microgrids in commercial buildings across different climates and operating profiles," in 2015 IEEE First International Conference on DC Microgrids (ICDCM), 2015, pp. 159–164.

[67] H. R. Atia, A. Shakya, P. Tandukar, U. Tamrakar, T. M. Hansen, and R. Tonkoski, "Efficiency analysis of AC coupled and DC coupled microgrids considering load profile variations," IEEE Int. Conf. Electro Inf. Technol., vol. 2016–Augus, pp. 695–699, 2016.

[68] S. H. Lee and J.-W. Park, "Optimal operation of multiple DGs in DC distribution system to improve system efficiency," in 2016 IEEE/IAS 52nd Industrial and Commercial Power Systems Technical Conference (ICPS), 2016, pp. 1–9.

[69] "Powerstax F351 - Full Brick 350W DC/DC converter." [Online]. Available:http://www.powerstax.com/wpcontent/uploads/2017/06/F211 _351_501.pdf. Last accessed date: 22 July, 2017.

[70] D. Sha, X. Wang, and D. Chen, "High Efficiency Current-Fed Dual Active Bridge DC-DC Converter with ZVS Achievement Throughout Full Range of Load Using Optimized Switching Patterns," IEEE Trans. Power Electron., vol. PP, no. 99, p. 1, 2017.

[71] H. Fan and H. Li, "High-Frequency Transformer Isolated Bidirectional DC-DC Converter Modules With High Efficiency Over Wide Load Range for 20 kVA Solid-State Transformer," IEEE Trans. Power Electron., vol. 26, no. 12, pp. 3599–3608, Dec. 2011.

[72] H. S. Kim, M. H. Ryu, J. W. Baek, and J. H. Jung, "High-Efficiency Isolated Bidirectional AC-DC Converter for a DC Distribution System," IEEE Trans. Power Electron., vol. 28, no. 4, pp. 1642–1654, Apr. 2013.

[73] "Cosel DBS700B28 - 700W DC/DC converter." [Online]. Available: https://en.cosel.co.jp/product/powersupply/DBS/DBS700B/DBS700B28 /. Last accessed date: 22 July, 2017.

The Model Development of an Effective Triggering System of Production Kanban Size towards Just-In-Time (JIT) Production

Mohd Norzaimi Che Ani[*,1], Shahrul Kamaruddin[2], Ishak Abdul Azid[3]

[1]*Manufacturing Section, Universiti Kuala Lumpur – Malaysian Spanish Institute, 09000, Kedah, Malaysia*

[2]*Mechanical Department, Universiti Teknologi Petronas, 32610, Perak, Malaysia*

[3]*Mechanical Section, Universiti Kuala Lumpur–Malaysian Spanish Institute, 09000, Kedah, Malaysia*

A R T I C L E   I N F O

A B S T R A C T

The contents of this article consist of an extension original work presented at the 4th International Conference on Control, Automation and Robotics (ICCAR 2018) and aims to develop the systematic model of an effective triggering system of production Kanban size towards Just-In-Time (JIT) production system. The developed model was introduced based on the philosophy of continuous improvement program by selection of case study industry in automotive manufacturing floor and it has been carried out through the development of an effective material Kanban size model using triggering system towards achieving a JIT production system. This developed model was analysed and verified by implementing through three phases which are the investigation of the Kanban size, analysis the production idling time and analysis of the time management for material replenishment process. The implementation was carried out to ensure the main objectives of the developed model in minimizing the waste of waiting through effective Kanban size for production system will be achieved by focusing on the communication between material incoming warehouse and production floor. The significant results were obtained during the verification process where the dual material feeders were identified for every process in the production floor to ensure the effectiveness of the production system. Throughout the implementation of the developed model, the production losses time caused by miscommunication between the production floor and warehouse had been identified and it was successfully minimized and the waiting time was reduced by 36.64%.

1. Introduction

In the current revolution of the production system, the crucial issues of efficiency and effectiveness of manufacturing flow are the main challenges in meeting the demands as expected by the customer. Jasti and Kodali [1] have stated in their study, in encountering this challenges, the main requirement in fulfilling customer expectations is industrial practitioners required to accelerate their corrective action by performing continuous improvement program to survive and increase the competitiveness. One of the effective approaches practiced by automotive industry is Lean Manufacturing (LM) concept which is pioneered by Toyota Production System (TPS). The ultimate goals of LM concept are eliminated the wastes of operation in between process

to process in manufacturing plant and ensure the smoothness of the production process to ensure the process flow are streamlined and efficient [2]. The LM concept was industrialized through the techniques of process improvement program by introduced a variety number of tools and proper methods to ensure the continuous improvements culture is adopted among the employees. TPS is targeted to eliminating or minimizing all type of identified wastes and smoothing the production flow by utilizing the LM philosophy known as Just-in-Time (JIT) and Jidoka or also known as production leveling [3]. JIT production system developed by TPS is an efficient concept with the main aim to minimize the production lead times within operation of production floor as well as quick response times in solving any arising issues between process to process, or communication between material supply chain system [4]. The main purpose of JIT system is to increase the smoothness of production flow and minimizing the overall

---

* Mohd Norzaimi Che Ani & Email: mnorzaimi@unikl.edu.my

costs of manufacturing operation covering minimization of the reject rate between process to process, optimizing the resources of production floor including space allocation for required processes, movement of the direct and indirect workers, standardization of the raw material inventory and synchronization of the internal and external transportation [5].

As introduced by TPS, LM concept focusing on the seven types of production wastes known as transportation, inventory, motion, waiting, over-production, over-processing, and defects. In the group of seven types of wastes, the most critical waste in influencing the efficiency and effectiveness of production flow is managing the inventory. Inventory means the management of raw materials from the warehouse into production floor or semi-finished products between process to process in production floor chain system, normally called as Work-in-Progress (WIP). Ineffective coordination and controlling of the WIP in the production floor causing the number of inventory will keep increasing and queues of the raw material prior transfer into production floor causing the waste of waiting in between process-to-process. As introduced by TPS in solving this issue, the LM concept has been introduced through implementation of the Kanban system. Kanban is a Japanese word was introduced in the scheduling system of raw material management for LM concept and JIT production system and it works as a management and controlling the inventory system either between production floor and raw material warehouse or management the internal supply chain which is controlling the WIP between process to process in production floor [6, 7]. As current practice, most of the automotive industries was implementing the Kanban system and the main benefits was obtained is ability to standardize an optimum limit or quantity of the inventory or WIP and preventing tasks of overloading during performing production processes, or also known as minimizing the 3M (Muda, Mura, Muri) [8]. LM concept well accepted and recognized by most of the industrial practitioners in the worldwide based on fruitful accomplishment in TPS especially controlling the inventory through one piece flow activity especially in automotive industries. However, the implementation of the one piece flow activity in di different types of industries is impossible as implemented by TPS because of several factors such as distance and location between suppliers and customers, facilities of the transportation system, a designed system of production flow and the availability of the facilities inside production floor. In tackling this issues, an effective WIP system through standardization of Kanban size was introduced with the main purpose to ensure effective management of holding inventory and minimize the waiting time in the production floor [9, 10].

In the current scenario, as practiced by the majority of the industrial practitioners in improving the efficiency and effectiveness of production system through implementation the Kanban system, but unstandardized numbers of inventory due to multiple products in production floor is required and it was caused miscommunication between production floor and incoming material warehouse due to inefficient inventory management system. Even though Kanban system widely applied in automotive industries but the main issue facing by automotive industry is to every single unit of finished-good vehicle needs thousands number of raw materials or semi-finished product and dealing with the various level of suppliers [11]. Hence, a direction of this research

paper, the implementation of the Kanban system using an effective number of feeders by standardizing Kanban size for in-house production floor has been developed and introduced. In ensuring the effectiveness of the developed Kanban system, the "triggering system" of management WIP was initiated to ensure production floor meeting the principles of JIT production system.

The focus of this paper is directed towards various literature that discussed the implementation of Kanban system in manufacturing system especially the inventory interaction between the incoming material warehouse and production floor. The first objective is to identify the critical success factor of achieving the optimum efficiency and effectiveness of production system by exploring the mechanism of the production flow. For a second objective, even though the Kanban system has been successful in manufacturing industries and widely recognized, but as the issues in effective communication between an incoming material warehouse and production floor are usually overlooked, the effective triggering system production Kanban size is introduced. In this paper, Section 2 describes the concept of production system to identify the critical area of influencing production efficiency and effectiveness, Section 3 presents on the development of the Effective Triggering System Production Kanban Size model, Section 4 presents the results and discussion of model verification in selected case study, and Section 5 concludes the overall results of the developed model.

2. The Concept of Production System

In this section, the current concept of the production system in various manufacturing industries of the production system will be discussed and explored. Since the production system is the main subject in manufacturing industries, the production efficiency and effectiveness must be closely monitored and performed to ensure the organizations meet the customer expectation and become competitive with other competitors. As indicated in the era of global competitiveness, the manufacturing organizations are enduring enormous pressure from their customers and competitors [12]. All these factors have given means to integrate the effective system of production flow with the complete production process (starting from the suppliers to the delivery to the customer) [13].

Expanding the inductive research approach, the attention has been focused on the flow of the production system according to the general factory set-up from various industrial sectors. That means, practically, the production system can be segregated into three sections; incoming raw materials, production processes and outgoing finished good product. In general interpretation, the production system can be summarized as input, process, and output. The input consists of incoming raw material, known as downstream area, which requires multiple numbers of tasks to ensure supplying the right parts at the right time into the process. The process in the production floor involves multiple numbers of sub-processes of semi-finished product and the process is continued until product completion. The output consists of inspection, packaging and shipping the good product to the customer based on the order in outgoing section.

2.1. Key Elements of Production System

As the production system can be generally divided into three sections, the key elements of the production system involving all

these sections. For input, the management of the raw material begins with the preparation of the raw materials either semi-finished product from sub-contractor or the child part from suppliers. A verification process of the raw materials by OEM normally goes through acceptance sampling analysis to ensure that raw materials meet the specification [14]. Acceptance sampling involves the process of inspection and testing of the product. A sample of goods is taken randomly from the lot or batch and a decision is made whether to accept the raw materials from suppliers based on the results of acceptance sampling and it will determine whether a batch of the raw material of goods should be accepted or rejected [15]. Once the quality of raw material is confirmed, it will be stored in warehouse storage area for the inventory purpose. The raw material will then be retrieved from the storage area, and the assigned worker will perform the preparation activity to ensure the specification quality and quantity according to the order from production, and this whole process is known as an upstream process. The raw materials will be temporarily parked at designed area, normally called kitting area or Work-in-Progress (WIP) area, with the purpose of the replenishment process. At this stage, the raw materials are now ready for use in production processes.

The next element is process, also known as midstream process, which means the raw materials will be processed from raw materials or semi-finished products. The production process is concerned with transforming a range of inputs into outputs as required according to the customer [16]. In the production processes, normally the Statistical Process Control (SPC) will be applied to ensure the finished product meets the specification. SPC is a statistical technique used for measuring, analyzing and controlling process variation in order to prevent defects or non-conformance in products. SPC involves inspecting a random sample of the output from a process and it will decide whether the process is producing products with characteristics that fall within a predetermined range [17]. Abnormality results from the SPC process will lead to the process improvement activity in order to solve the issue. After production processes are completed, and then the finished product will be transferred to the next section which is the output.

In the output stage, also known as a downstream process, the product will involve with the specific package which are packaging and shipping [18]. The finished product will be packed according to the customer expectation and at the same time, the assigned worker will perform the visual inspection to ensure that the finished product is free from defects. Some of the issues might arise at this stage such as cosmetic issue due to handling process and lead time. After completing the packaging process, then the finished product will be transferred into the transportation system for the shipping process. The production system is considered completed once it reaches the stage of shipping.

2.2. Issues pertaining to the elements in Production System

In the current global scenario, the competitiveness in the manufacturing industries keeps increasing as the customers ask for the minimization of the operation costs and increased quality of the product at the same time [19]. In order to ensure meeting this situation, the efficiency and effectiveness of production system play main roles in improving the production system and at the same time minimizing the cost of operation. To achieve these objectives, the whole chain of the production system, upstream, midstream and downstream, should the improved. This is also highlighted by Lean Manufacturing (LM) approach, where the process improvement should be studied from suppliers to customers [20]. Since the upstream section is the predecessor process for the production process, it should be prioritized to achieve optimum efficiency of the production system. Even though the upstream process is typically seen to play a minor role in the production system, it affects the lead time, part shortage, part mixing or product failure, which leads to inefficient production. Thus, the upstream process must achieve high efficiency, reliability and traceability to ensure the parts transferred into the production process meet the expectation.

In achieving high efficiency of communication for the upstream process, the synchronization interaction between warehouse and production must be effective in term of planning daily production required, quality requirement and time management of raw material replenishment process. Usually in achieving effective synchronize communication, the process of internal supply chain between warehouse management and production floor will be connected using inventory control system [21]. The chain of the production process was started with warehousing because it is the first section in the manufacturing process which is responsible controls physical of the receiving material and it performs the material verification tasks during the process of receiving from suppliers or vendors, incoming material sampling analysis and segregating the raw material before transferring into the production floor. The management of the warehouse must know the physical shape of the raw material and also know all information of the time and place of supplying raw material into production floor. The process of monitoring and verification of the raw material inventory is about the measurement of the critical parameters and quality inspections for achieving an effective management of the inventory [22]. Inefficient inventory management in the upstream process will be caused interruption of the whole production flow and it can lead to high stoppage time, increasing the production lead times, congested queues of transferring raw materials, smoothness of product flow and increasing costly operation setups [23].

3. The Development of Effective Triggering System of Production Kanban Size Model

In order to ensure achieving the optimum efficiency and effectiveness of production system, an effective triggering system for Kanban management system was developed and introduced in this research-based philosophy of continuous improvement program through three phases. The first phase required analysis of Kanban size then follows by the second phase consists of the investigation of the production idling time using queuing analysis and third phase the material preparation 'triggering system' was introduced. The framework development of an effective triggering system for Kanban management system is illustrated in Figure 1.

Based on the framework illustrated in Figure 1, in the first phase, the analysis of the production Kanban size was properly carried out in ensuring achieve the optimization of the material replenishment cycle through minimizing the production losses time during raw activity of raw material replenishment. During

executing this phase, the actual size of production lot size is required to be identified to ensure determination of the suitable number of production lot size achieving the optimum solution. The number of the production lot size considered same from day-to-day production operation based on long-term production planning but the quantity of the raw material or semi-finished products from suppliers would be different due to the production requirement as daily planned by production planner. In the determination of an effective production Kanban quantity, the reliable information on the individual process cycle time from production floor and raw material preparation per unit of the finished-good product were needed to ensure the accuracy of calculating the effective Kanban size as formulated in Equation (1).
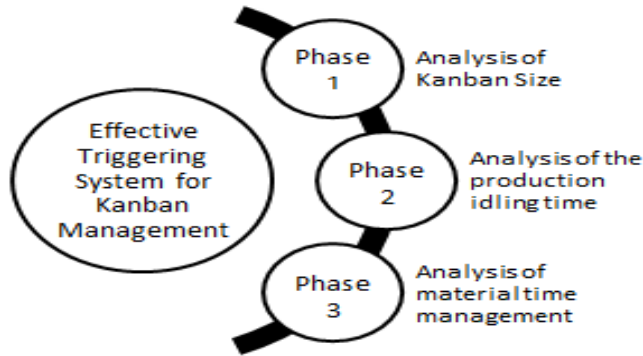
Figure 1. The framework development of effective triggering system for Kanban management system

$$\text{Effective Kanban size} = \frac{\text{Material Prepration Cycle Time}}{\text{Production Cycle Time}} \quad (1)$$

The Kanban size was identified to ensure planning of the upstream process in proper manner. Once the Kanban size was determined, the analysis of every motion performed by warehouse personnel for the upstream process will be analysed. The worker will be performed the assigned tasks to complete the material preparation process and time will be recorded. Since the warehouse activities and production floor normally set-up based on segregated building, the traveling time from the warehouse into the production floor must be considered in motion time study analysis of material preparation activities. This situation created the waste of transportation and the production process will be interrupted during the material replenishment process.

Once the Kanban size has been determined in Phase 1, then in the following phase which is Phase 2 the application of the queuing system theory will be used to analyse the production idling time for each process in production floor with the purpose to identify the wastes of production waiting time. The queuing theory model consisted of mathematical study based on the application of the probability concept to identify the production waiting time, or idling time-based on the queuing behavior of raw material management [24]. The application queuing theory model was considered in this research, because in the queuing analysis theory, the queuing behavior will be counted as raw material queues waiting for prior transfer into production floor is based on consideration of the production processing time and cycle of raw material replenishment from warehouse into production floor, and both of the parameters are then will be converted into hourly service rate, known as µ and hourly arrival rate, known as ʎ. The

determination for both µ and ʎ were analysed based on Equations (2) and (3).

$$\text{Service rate, } \mu = \frac{1 \text{ hour}}{\text{mean service cycle time}} \quad (2)$$

$$\text{Arrival rate, } \Lambda = \frac{1 \text{ hour}}{\text{mean arrial rate}} \quad (3)$$

In queuing system theory, a customize model of queuing behavior is constructed to determine the queue lengths (number of inventory) and the waiting time (production idling time) that can be predicted based on systematic analysis of queuing behavior either single queuing behavior model or multiple queuing behavior models. As practiced by most of the industries, normally production floor where the process-to-process in connected and receiving the raw material from the warehouse, the material preparation activity in warehouse considered as a supplier for production floor. Based on this condition the single queuing is more suitable for the further analysis. Single queueing theory model normally practices by using Kendall's notation in the form A/S/C where A describes the time arrivals to the queue, S the size of queues and C the number of servers or counters at the node [25]. Even though the single theory model required a single queues length and single server, but multiple situations of queuing behavior are possible to adopt. Based on available theories of algorithms for basic single theory model, several conditions have been identified such as basic single server, constant or undefined service time, finite queuing line and finite calling population. The selection of an application for each condition depending on the activities conducted in queuing behavior. Shortle *et al.* [26] were segregated the group of basic queuing theory as summarized in Table I.

Table 1: Type of queuing system

Theory of Algorithms	Queuing System	Queuing Behavior
Single server	Basic single server	Determined cycle time
Single server	Undefined or constants service time	Undefined or fixed cycle time
Single server	Finite queuing line	Finite input
Single server	Finite calling population	Finite server
Multiple server	Multiple servers waiting line	Multiple similar servers

Once a single model of the queuing system was determined, then detailed analysis using probability theory in queuing analysis system will be carried out. In queuing theory, generally five main parameters will be analysed which are the probability the server is idle (Po), the average of number of customer in the queuing system (L), the average number of customers in the waiting line (Lq), the average time a customer spends in the total queuing line (W) and the average time customer spends waiting in the queue to be served (Wq). As practiced in the production system, smoothness interaction between warehouse and production floor is required, the definition L and W was customized to suit with a production system. L was defined based on the number of production

processes waiting to be completed for material replenishment activity in production floor, while W was defined as an average time the production process waiting for completion of the material replenishment activity. Subsequently, the determination for both of the L and W will be based on the identified service rate, μ, and arrival rate, Ʌ as mentioned earlier. Since this research focusing on the material replenishment process versus the production system, the most significant queuing behavior system was finite calling population as indicated in Table I. It was used in this research because the main constraints during material replenishment activity is the quantity of material feeder where is considered as terminal or server and the production processes considered as customer for the warehouse during replenishment activities. Furthermore, in ensuring the reliability of the determination of both L and W, the number of entire processes in production floor also considered. The determination for both L and W for finite calling population queuing system was formulated as Equation (4) until Equation (8).

$$Po = \frac{1}{[\sum_0^N \frac{N!}{(N-n)!}(\frac{Ʌ}{\mu})^n]} \qquad (4)$$

$$Lq = N - [(\frac{Ʌ+\mu}{Ʌ})(1 - Po)] \qquad (5)$$

$$L = Lq + (1 - Po) \qquad (6)$$

$$Wq = \frac{Lq}{(N-L)Ʌ} \qquad (7)$$

$$W = Wq + \frac{1}{\mu} \qquad (8)$$

Successful implementation and completion of Phase 2, the 'triggering system' for effective management of WIP will be developed in the third phase. In this phase, time management analysis of inventory planning was carried out in determination the cycle time of WIP 'triggering system' in ensuring sustainability of the developed production Kanban system towards achieving optimum efficiency and effectiveness of production system. Hence, in this phase, the raw material replenishment time from warehouse and cycle was determined based on the daily production scheduling and planned operation hours to ensure the effective preparation time of warehouse activity. The identification of the cycle for material replenishment activity will be determined based on planned working time per days such as a number of working shifts and the daily targeted output as per planned by production planning. Once the material replenishment cycle and time was identified, then the traffic management for movement of raw material activities in production floor could be identified to ensure smoothness of material replenishment process. This is important because the process of transferring the raw materials from the warehouse into production floor required high repetitive movements and traveling due segregated warehouse and production floor from different buildings. In this phase, the proper management of internal transportation system is required by implementing visual management system such as color coding and alarm system.

4. Results and Discussion of Model Development

The selected case study industry for this research is an automotive manufacturing industry which operating the process of assembling the car. The production system was designed using

single production assembly layout with connected from process-to-process and the entire production floor and it was divided into five sections which are a trim section, chassis section, final assembly section, accessories assembly section, and final inspection section. The raw material warehouse was segregated from the different building which is next to the production floor building and the material preparation and acceptance sampling process were performed in the warehouse building and it will be transferred into the production floor based daily production planning. Each section in the production floor and warehouse management was managed by the different groups of management and it needs an efficient material preparation activity during establishment the smoothness of production system.

Since the processes of production floor in the selected case study was set-up based on five sections, one of the section will be selected as initial improvement area to be conducted to verify the developed model in this research. Even though five sections is required in assembling the car, but only three sections were considered and possible conducting the improvement activity of verification process, which are Trim section, Chassis section, and Final Assembly section. These three sections were identified as potential sections in influencing unsmooth production process due to interaction with the multiple types and quantity of raw materials from the warehouse. The other two sections which are Accessories Assembly and Final Inspection were not involved with replenishment of raw material because only conducting the process of checking and verification of finished-good product and it was considered as a stand-alone process. Based on the several criteria, Trim section was selected and in implementing the developed model as a case study for this research. The Trim section was selected according to two criteria which are this section considered as a predecessor for overall production chain system because it was set-up as first process and the failure or inefficient process of Trim section would be interrupted the entire production flow. Furthermore, which is the second criterion in the selection of Trim section because a number of the assembly sub-processes was the highest compared to the other two sections.

Table 2: Result of motion time study for material preparation activity in warehouse

Activity	Cycle time (Minutes)
1. Unboxing	5.00
2. Quality Inspection	13.00
3. Segregation	7.00
4. Material arrangement	5.00
5. Transferring process	10.00
Total Cycle Time :	40.00

Currently, as practice by selected case study industry, the Trim section required nine sub-processes, while Chassis and Final Assembly only required six sub-processes for each section. Means, 21 assembly processes are required to assemble a finished-good product. A higher number of sub-processes in production floor meant it required a higher number of raw materials. Based on this supported condition, the Trim section was finalized as a selected

section to implement the developed framework of effective Kanban system.

In Phase 1, motion time study of material preparation activities in the warehouse has been observed using motion time study analysis which are every step of activities was measured and the results show the required cycle time in of material preparation averagely at 40 minutes to complete a set of raw material for a single unit of finished-good product. The cycle time of material preparation was defined based on detailed analysis as shown in Table 2. This information would be used in this phase for the determination of the effective Kanban size.

Once the process cycle time of material preparation in warehouse activity was determined, then the cycle time of production processes was analysed using motion time study analysis in production floor. The motion time study was captured and compiled based on detailed activities for every process was analysed to determine the actual process cycle time with the purpose to calculate the effective Kanban size. The selected production layout was identified consists of 25 processes from five sections to complete a single unit of finished-good product. All related activities and detailed motions practiced by the workers for every process was captured using a stopwatch during the workers performing their tasks and the captured results of the motion time study for the entire processes was converted into average as shown in Table III. As presented in Table III, from the entire 25 processes, the sub-process sequence number 14 was identified as the bottleneck process because it is required 47.25 minutes. This cycle time was considered as the production cycle time because it was identified the highest cycle time compared to others processes. As mentioned earlier, from entire 25 processes, only 21 processes were considered in determination of the effective Kanban size because that processes involved with material replenishment and the other 4 processes (process 22 until sub-process 25) were verification, checking and testing process without involvement of raw material replenishment process and the average process cycle time of these 21 processes is 37.63 minutes. That means, averagely every 37.63 minutes production processes required to replenish with a new set of raw materials from warehouse regularly based on the planned production output.

Table 3: Result of motion time study for selected production layout

Process	1	2	3	4	5
Time (Min)	40.51	22.50	45.20	42.38	32.34
Process	6	7	8	9	10
Time (Min)	36.42	40.16	45.25	42.37	42.20
Process No.	11	12	13	14	15
Time (Min)	47.20	45.00	47.20	47.25	45.56
Process No.	16	17	18	19	20
Time (Min)	25.52	35.00	26.20	25.22	35.00
Process No.	21	22	23	24	25
Time (Min)	47.00	10.00	10.00	8.00	8.00

Successful identification of the production cycle time and warehouse cycle time, the Kanban size was calculated by using Equation (1) as mentioned in the previous section. As derived by Equation (1), the optimum Kanban size was then identified at 0.85 units or can be roundup into 1.0 unit. The results of calculated Kanban size represents material replenishment with minimum Kanban size of 1 unit was required in production floor in order to complete one unit of finished-good product in every process for current production floor as planned output by production floor. The internal transportation activity between warehouse and production floor required raw material trolley as a material handler, as in material replenishment activity, and it known as a material feeder. Based on identified information of Kanban size, then the queuing system theory was conducted in the following phase to analyse the production waiting time during material replenishment processes.

Since the production floor only required a single unit of the material feeder to operating the production processes, the comparison in term of production waiting time was made by increasing number of the material feeder into two units, known as dual feeders. The two units of material feeders were identified based on an assumption of the change-over process which is one of the material feeder was considered as work-in-progress (WIP) during performing an assembly process by production personnel and the other one feeder was parked as standby area for replenishment purpose. Means, each process of production floor was occupied with two material feeders. During the operation of the production system, the emptied material feeder would be transferred into the warehouse for performing the material preparation activity. In the material preparation activity, some processes are required such as unboxing, quality inspection, material segregation; material arrangement and transferring into production floor are required to ensure the requirement was met. At the completion of the first phase, Phase 2 was performed using queuing analysis theory to analyse the waste of production waiting time.

The queuing system analysis was conducted in the selected section of case study analysis based on the determined service rate and arrival rate as identified from process cycle time of production floor and material preparation in the warehouse. The verification process in the determination of the optimum number of the material feeders was simulated based on two conditions based on calculated effective Kanban size from the previous section. Based on the results of Kanban size is 1 unit per feeder, and then the simulation was decided based on two conditions which are a single feeder and double or dual feeders. The results were obtained from this simulation was then analysed by comparison on the average waiting time or idling time of production process due to material replenishment process. In the analysis of queuing theory system, the arrival rate known as ʎ according to the capacity of the hourly arrival rate of raw materials from the warehouse into production floor. As identified from previous phase, the warehouse only can supply the raw materials from warehouse into production floor at every 40 minutes, and then it was converted into hourly service rate of material replenishment known as µ at 1.5 by using Equation (2). Then, the arrival of raw material from the warehouse was converted into hourly arrival rate known as ʎ by using Equation (3) resulted as 1.69 based on one hour divided by production average production cycle time which is 37.63 minutes as presented in Table

3 earlier. Then, the optimum number of the material feeder was determined to minimize the wastes of production waiting time or idling time by comparing the idling time between a single feeder and dual feeders as simulated. Based on the results of both conditions, the arrival rate could be different from single and dual feeders because increasing number of material feeder during material replenishment process, simultaneously decreasing the cycle of material replenishment process. This situation required to further analyse of cycle time for material replenishment process and converted into service rate, μ and the results show increasing from 1.5 into 3.0 as using Equation (1). This situation occurred because dual feeders required multiplying by two and the service rate, μ as summarized in Table 4.

Table 4: Result of queuing analysis

Feeder	Service rate, μ	Arrival rate, ʎ	No. of processes, N
Single	1.5	1.69	21
Dual	3.0	1.69	21

The results presented in Table IV then were then further investigated to analyse the wastes of production waiting time by application of Equation (4) until Equation (8) as mentioned in the previous section. Both of the feeders systems were then analysed in details to ensure the justification of the most effective system for material preparation frequency with the purpose of waste elimination and minimizing the production idling for achieving JIT production system and smoothing the production flow. The results of the comparative analysis between both feeders systems were calculated as presented in Table 5.

Table 5: Result of feeder comparison

Parameter	Single feeder	Dual feeders
L	19.11	18.22
W	6.66	4.22

Based on the result obtained from Table V, according to number processes in production floor consisted of 21 processes, the average number of process waiting to be replenished by the material handler is 19.11 processes by using a single feeder. That means that once the raw material was emptied in any process in production floor, the subsequent process needs to wait the next cycle of replenishment process for required to wait completion of replenishment process for another 19.11 processes to fulfill its request. Translating into cycle time, it could be interpreted as a average individual process in production floor waiting to be replenished by a new set of raw material is 6.66 hours per cycle of raw material replenishment process. Simulation of single feeder impacting the production losses time or stoppage time due to the waste of waiting is 6.66 hours per day. Comparing with the simulation results using dual feeders of material replenishment, the average number of production waiting time to be replenished is 18.22 processes, while the average production waiting time to be replenished is 4.22 hours. Both of the simulation results were then compared in term of production lost time or stoppage time. Since dual feeders is only 4.22 hours per day instead of 6.66 hours per day by a single feeder and it was chosen as effective Kanban size

based on elimination 36.64% of the lost time or production stoppage. The results of queuing analysis was presented the dual feeders is an optimum number of material feeders to meet the production requirement because of reducing the waiting time at 36.66% in production floor and simultaneously increasing the production capacity. This situation improved the production efficiency in term of capacity planning because closing the gap between actual production outputs versus planned production output.

Once an effective Kanban size and material preparation frequency was determined in the first phase, and number of optimum material feeders was identified in the second phase, then in third phase, the process of standardization was carried out in the production floor by using visual management concept to ensure achieving and sustaining the optimum production efficiency and effectiveness towards JIT system. As resulted from comparison simulation results from single and dual feeders, a number of material feeders was identified as dual feeders because it was recognized as the most effective. Accordingly, the comparison of the number of production output for both situations; single and dual feeders also were analysed. Application of the single feeder, production floor required planning for the stoppage time is every 6.66 hours due the material replenishment activity and the planned production time for this case study industry is 8 hours per day, means the effective production time only 1.34 hours or 80.40 minutes per day which is during material replenishment process, the production was idling. This situation will be impacted on the actual output of production process where the production system can only achieve for 2 units per day based on the effective available time after deducting the lost time divided by bottleneck process as presented in Table III. Comparing by the implementation of the dual feeders for material replenishment process, the effective available production time will be maximized through parallel performing the job tasks during material replenishment process. Thus, the simulation results was recommended the dual feeders must be implementing during material replenishment process in production floor which is one feeder was considered "in-use" and another feeder considered "standby". The implementation of the dual feeders for material replenishment process in production floor was eliminated the production stoppage time due to waste of production waiting for raw material to be replenished. The effective available production time was increased as planned 3.78 hours or equivalent to 226.80 minutes after deducting 4.22 hours of lost time or stoppage time from planned production time 8 hours. The production output was drastically increased by 150% which is from 2 units into 5 units per day based on calculation from 3.78 hours available hours divided by identified production bottleneck cycle time which is 47.25 minutes as resulted from Table III.

The determined number of optimum material feeders indirectly identified the number of production Kanban size. To ensure synchronize communication between an incoming material warehouse and production floor, the optimum the number of Kanban size must be similar with number of material feeders. Since dual feeders is an optimum solution for material replenishment process, the Kanban size also suggested as 2 units similar to the number of material feeders to ensure synchronize communication and maximizing the production efficiency. According to Kanban size of 2 units per lot and current production capacity only produced 5 units per day based on eight hours

production planning, 1,400 units of finished good products will be achieved per year with planned working days is 280 days. Based on this result, the number of the cycle for part replenishment was determined three cycles per day formulated from the comparison between daily production planning and Kanban size. Three cycles of material replenishment at production floor is mandatory for everyday activity according to dual material feeder concepts and production forecast. The three cycles meant the upstream time cycle of the process must be at every 2.67 hours according to the derivation from available time, 8 hours per day and cycle required, 3 cycles per day.

Based on the observed time, the result showed that the cycle time completion for one cycle of an upstream process according to the seven tasks was 40 minutes, as defined in Phase 1. It meant the material preparation process time per Kanban size or per dual feeders required 80 minutes per cycle upon verified by supervisor and the feeders were then transferred to the production area.

In the process of maintaining the sustainability of the improvement process for material replenishment process, the routing of the material replenishment feeders was designed for easy and quick detection of empty feeder based on designed "triggering system" from the previous phase. The suitable area in the production floor was marked with "ready-in-use" for parking the materials feeder and it was located in the dedicated area known as WIP area. The main purpose of floor marking is for easy triggering by production personnel and ensure on the readiness of the raw material to be used once required. The location of emptied feeder also was designed and located at the next to each process of the production floor and close to the pathway. The identified area in the production floor was marked with "vacant" and using red color marking to trigger the emptied feeder during the replenishing process. Once the emptied feeder triggered by the material handle, then he/she would be communicated with warehouse personnel for the preparing of raw material according to triggering production process. Since the finite frequency of material preparation cycle was applied in production floor, the scheduled time of each cycle for the material preparation was identified based on three cycles of per day as planned by production floor. All the three cycles of replenishment time were defined and fixed on 09:20 am, 12:00 am and 03:40 pm every day to maximize the production output as shown in Table 6.

Table 6: Part preparation scheduling for triggering system and comparison between single and dual feeders

Time	08:00	09:20	12:00	15:40
Replenish Cycle	0	1	2	3

Achieving efficiency and effectiveness of production system required a systematic approach of wastes elimination. In actual industrial practices, most of the industries performed their continuous improvement activities to improve production efficiency by implementing a Kanban system as benchmarked by LM. But, inefficient determine the Kanban size causing miscommunication between the production floor and warehouse which is created the waste of production waiting time. In order to solve this issue, this research was developed and implemented to ensure an effective way in the determination of the production Kanban size will be achieved through calculated effective number

of feeders and develop the triggering system of the material replenishment activity. The main focuses of this research is in an automotive production system which is automotive manufacturer required dealing with thousands number of the raw materials from multiple levels of suppliers or vendors. This condition required optimum efficiency and effectiveness of interaction between process to process and suppliers to customer to ensure synchronization of interaction are achieved. Using the developed approach of effective Kanban size, the wastes of production idling causing waiting to be replenished was drastically improved by 36.64% and indirectly optimized material replenishment process and the production output was increased from 2 units into 5 units per day.

Achieving optimum efficiency and effectiveness of production system is the main goal of any management organization in industrial practices as practices through continuous improvement program, but the lesson learned from this research is that the continuous improvement program should not just focus on inside the production floor only, but outside of production floor also must be considered as potential area of improvement activity such as upstream and downstream processes. Organizations are strived to reduce the wastes of production system inside production floor as much as possible. Using this developed triggering system of production Kanban size, the production waiting has been drastically reduced and drives the production floor towards JIT production system. Widely continuous improvement activity focusing on the interruption time of the production process is expected in the long run after implementing the continuous improvement program in all related sections of the production system. As an encouraging result in this research through the development and implementation of the improvement model, the team decides to continue performing more improvement activities project for production floor for other sections in the future research. The positive results from this developed framework in this research, the impacts from obtained results also change the mindset of the industrial practitioners in avoiding "fire-fighting method" of tackling the production issues.

5. Conclusion

In this research article, the continuous improvement activity in the manufacturing industry concentrating on inventory management has been conducted by introducing and implementing the model of effective Kanban size. The developed model was verified and implemented in the selected case study and the obtained results were analysed through utilizing the number of optimum material feeders according to calculated effective Kanban size and analysis of the production idling time using queuing system theory. Two conditions of material replenishment activity were simulated and analysed which are single and dual material feeders in order to achieve JIT production system. The behavior of the production system through communication between warehouse and production floor was investigated in order to ensure the accuracy and reliability of the improvement process. The ultimate objectives are to ensure the implementation of the Kanban system using an effective number of material feeders for production floor achieving optimum efficiency and effectiveness. To ensure sustainability of the developed model the WIP "triggering system" was introduced based on time management and visual management. Successful implementation of the developed framework of an effective

Kanban size based on an effective quantity of feeders and triggering system leads production system eliminating the wastes of waiting by 36.64%. In overall, the results of this research was meeting the research objectives which are effective communication between the incoming material warehouse and production floor was developed through designing the effective Kanban size and the systematic triggering system was introduced based on effective scheduling of material replenishment process.

Acknowledgment

References

[1] N. V. K. Jasti and R. Kodali, "Lean production: literature review and trends," *International Journal of Production Research,* vol. 53, pp. 867-885, 2015.

[2] S. Bhasin and P. Burcher, "Lean viewed as a philosophy," *Journal of manufacturing technology management,* vol. 17, pp. 56-72, 2006.

[3] J. K. Liker and J. M. Morgan, "The Toyota way in services: the case of lean product development," *The Academy of Management Perspectives,* vol. 20, pp. 5-20, 2006.

[4] K. Amasaka, ""New JIT": A new management technology principle at Toyota," *International Journal of Production Economics,* vol. 80, pp. 135-144, 2002.

[5] R. R. Fullerton and C. S. McWatters, "The role of performance measures and incentive systems in relation to the degree of JIT implementation," *Accounting, Organizations and Society,* vol. 27, pp. 711-735, 2002.

[6] W. J. Hopp and M. L. Spearman, "To pull or not to pull: what is the question?," *Manufacturing & service operations management,* vol. 6, pp. 133-148, 2004.

[7] B.-R. Lea and H. Min, "Selection of management accounting systems in Just-In-Time and Theory of Constraints-based manufacturing," *International Journal of Production Research,* vol. 41, pp. 2879-2910, 2003.

[8] E. Lander and J. K. Liker, "The Toyota Production System and art: making highly customized and creative products the Toyota way," *International Journal of Production Research,* vol. 45, pp. 3681-3698, 2007.

[9] B. Rouwenhorst, B. Reuter, V. Stockrahm, G.-J. van Houtum, R. Mantel, and W. Zijm, "Warehouse design and control: Framework and literature review," *European Journal of Operational Research,* vol. 122, pp. 515-533, 2000.

[10] Y. Monden, *Toyota production system: an integrated approach to just-in-time*: Productivity Press, 2011.

[11] Z.-L. Chen, "Integrated production and distribution operations," in *Handbook of quantitative supply chain analysis*, ed: Springer, 2004, pp. 711-745.

[12] H. W.-c. Yeung, "Regional development and the competitive dynamics of global production networks: an East Asian perspective," *Regional Studies,* vol. 43, pp. 325-351, 2009.

[13] A. Gunasekaran, C. Patel, and R. E. McGaughey, "A framework for supply chain performance measurement," *International journal of production economics,* vol. 87, pp. 333-347, 2004.

[14] C.-W. Wu and W. L. Pearn, "A variables sampling plan based on C pmk for product acceptance determination," *European Journal of Operational Research,* vol. 184, pp. 549-560, 2008.

[15] K. Dumičić, V. Bahovec, and N. K. Živadinović, "Studying an OC curve of an acceptance sampling plan: a statistical quality control tool," in *7th WSEAS International Conference on Mathematics and Computers in Business and Economics (MCBE'06)*, 2006.

[16] J. C. Aurich, C. Fuchs, and C. Wagenknecht, "Life cycle oriented design of technical Product-Service Systems," *Journal of Cleaner Production,* vol. 14, pp. 1480-1494, 2006.

[17] C.-W. Wu, W. Pearn, and S. Kotz, "An overview of theory and practice on process capability indices for quality assurance," *International journal of production economics,* vol. 117, pp. 338-359, 2009.

[18] M. Rother and J. Shook, *Learning to see: value stream mapping to add value and eliminate muda*: Lean Enterprise Institute, 2003.

[19] J. Knezevic and V. Narayan, "Business performance and maintenance: How are safety, quality, reliability, productivity and maintenance related?," *Journal of Quality in Maintenance Engineering,* vol. 18, pp. 183-195, 2012.

[20] G. A. Zsidisin and L. M. Ellram, "An agency theory investigation of supply risk m anagement," *Journal of supply chain management,* vol. 39, pp. 15-27, 2003.

[21] A. Sinclair and M. Monge, "Concept facility based on single-use systems, Part 2," *BioProcess Int,* vol. 3, 2005.

[22] A. Neely, M. Gregory, and K. Platts, "Performance measurement system design: A literature review and research agenda," *International journal of operations & production management,* vol. 25, pp. 1228-1263, 2005.

[23] M. Rosvall and O. Christensen, "A model for cost effective improvement of a raw material management in the inventory: A Case study," ed, 2012.

[24] R. Nelson, *Probability, stochastic processes, and queueing theory: the mathematics of computer performance modeling*: Springer Science & Business Media, 2013.

[25] O. Handel and A. Borrmann, "Service bottlenecks in pedestrian dynamics," *Transportmetrica A: Transport Science,* pp. 1-14, 2017.

[26] J. F. Shortle, J. M. Thompson, D. Gross, and C. M. Harris, *Fundamentals of queueing theory* vol. 399: John Wiley & Sons, 2018.

A Model of EM Fields from Static Data for Moving Conducting Cylinder

Esmail Mohamed Mohamed Abuhdima[*,1], Robert Prewitt Penno[2]

[1]*Computer Science, Physics and Engineering, Benedict College, Columbia, SC 29204, USA*

[2]*Electrical and Computer Engineering, University of Dayton, Dayton, OH 45469, USA*

A R T I C L E I N F O	A B S T R A C T
	The effect of rotation and translation of an electromagnetic (EM) scatterer upon the scattered field is studied for the case of TM polarization. The Franklin and Lorenz transformations are used to analyze the phase and the magnitude of the backscattered field during the movement of the scatterer. Also, a new model is presented that simulates rotation and translation using static data. This stationary data is generated using FEKO. These simulation results of the exact analytical solution are compared with the results of the proposed model using static data. Finally, a comparison between the phase and magnitude of the backscattered field of a scatterer undergoing both translation and rotation for the two different modes (TM and TE) is presented.

1. Introduction

This work is an extension of previous work [1], which studied the behavior of the scattered field from a conducting cylinder in rotation and translation for the case of TE-mode, i.e. the direction of the incident magnetic field is parallel to the axis of the conducting cylinder. However, the current work concentrates on the scattered field of the moving conducting cylinder in the case of TM- mode, where the direction of the incident electric field is parallel to the axis of the conducting cylinder. Furthermore, a comparison between both modes (TE and TM) is presented herein.

It is known that recent research has focused on the study of different scattering phenomena so as to better evaluate the nature of the scattering body. Historically, the Galilean transformation was used to investigated the effect of the rotation of the conducting cylinder by many researchers [2,3,4]. It was seen from measured results how the behavior of scattered field of the smooth metallic cylinder is affected by rotation. This behavior showed that the incident frequency signal is affected by the rotation of the smooth metallic cylinder. According to the measured results, it is useful to have an analytical expression to explain this phenomenon [5]. The previous work found that the pattern of the scattered field of the rotating cylinder is different in comparison with stationary one especially in the case of TE mode [6,7,8]. In this extensional work,

the behavior of the scattered field is investigated, when the polarization of the incident wave is TM polarized form. The expression of the scattered electric field in rotation and translation is found using the Franklin transformation and Lorentz transformation. The new model is used to simulate rotation and translation from stationary static data. Finally, the simulation results of this new model and the previously proposed model [1] are compared.
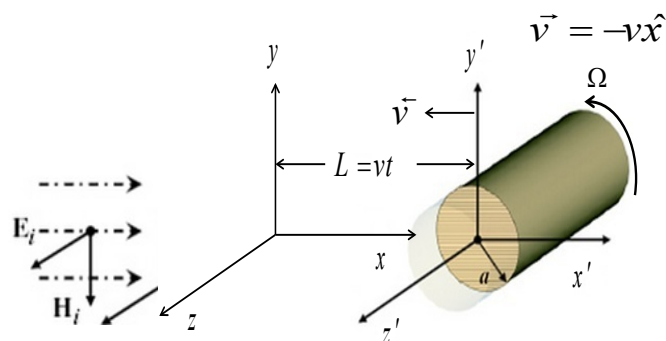
Figure 1: TM -mode of the conducting cylinder in rotation and translation

Figure 1 shows the description of the problem and the incident electric field is represented by [9]

*Corresponding Author: Esmail Mohamed Mohamed Abuhdima, Faculty of Computer Science, Physics and Engineering, Benedict College, Columbia, SC, USA. Email: esmail.abuhdima@benedict.edu

$$E_z^i = e^{-jkx} = \sum_{-\infty}^{+\infty} j^{-n} J_n(kr) e^{jn\varphi}, \qquad (1)$$

where $k = \omega/c = 2\pi f/c$ and c is speed of light.

In terms of the Franklin transformation, the scattered electrical field from a rotating conducting cylinder is given by [10]

$$E_z^{sc} = \sum_{-\infty}^{+\infty} B_n H_n^{(2)}(kr) e^{jn\varphi}, \qquad (2)$$

where, B_n is a complex scattering number. This unknown constant number is found using boundary conditions at $r = a$, where a is the radius of the conducting cylinder [11]

$$B_n = \frac{j^{-n}\left[c\gamma_n J_n'(\gamma_n a) J_n(ka) - k J_n(\gamma_n a) J_n'(Ka)\right]}{k J_n(\gamma_n a) H_n'^{(2)}(Ka) - c\gamma_n J_n'(\gamma_n a) H_n^{(2)}(ka)}, \qquad (3)$$

where, $\gamma_n^2 = k^2 N^2(1-jq) + \dfrac{n\omega\Omega}{c^2}(2N^2 - 2 - jN^2 q)$,

$N^2 = \varepsilon_r \mu_r$, $q = \dfrac{\sigma}{\varepsilon\omega}$, Ω is angular velocity, ω is angular frequency, σ is the conductivity, ε is the permittivity, and μ is the permeability. In this case, when the conductivity of the cylinder (σ) is very large, (2) goes back to the original expression of the backscattered field of perfect conducting cylinder in the stationary case as [12]

$$\lim_{\sigma \to large} E_z^{sc} = E_s^{sc} = \sum_{-\infty}^{+\infty} \frac{J_n(ka)}{H_n^{(2)}(Ka)} H_n^{(2)}(kr) e^{jn\varnothing}. \qquad (4)$$

In this step, the approximation to Lorentz transformation is applied to (2) to analyze the phase and magnitude of the backscattered field during translation [13]. The backscattered field of the conducting cylinder in rotation and translation is written as

$$E_m^{sc} = -\hat{z}\sqrt{\frac{2j}{\pi k'}} \sum_{-\infty}^{+\infty} B_n' e^{jn\pi}, \qquad (5)$$

Where,

$$B_n' = \frac{j^{-n}\left[c\gamma_n' J_n'(\gamma_n' a) J_n(k'a) - k' J_n(\gamma_n' a) J_n'(k'a)\right]}{k' J_n(\gamma_n' a) H_n'^{(2)}(k'a) - c\gamma_n' J_n'(\gamma_n' a) H_n^{(2)}(k'a)},$$

$$\gamma_n'^2 = k'^2 N^2(1-jq) + \frac{n\omega'\Omega}{c^2}(2N^2 - 2 - jN^2 q),$$

$$\omega' = \omega\left(\frac{c+v}{c-v}\right) \approx \omega\left(1 + 2\frac{v}{c}\right) \quad \text{and} \quad k' = k\left(\frac{c+v}{c-v}\right) \approx k\left(1 + 2\frac{v}{c}\right).$$

The objective of this work is to create a model that simulates translation and rotation of the static data for a complex scatterer as shown in Figure 2. The features of this improved model are written as

$$E_m^{sc} = E_s^{sc} \times A(s), \qquad (6)$$
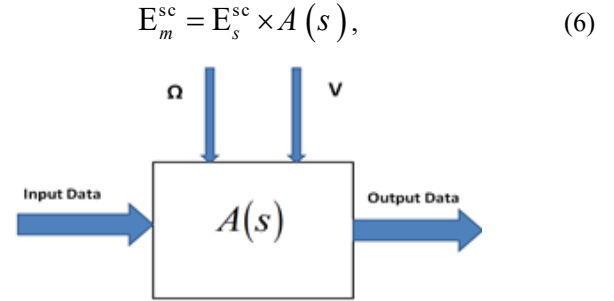
Figure 2: Proposed model

E_s^{sc} is the backscattering electric field of the perfect conducting cylinder in stationary case [12] and $A(s)$ is a transformation given by

$$A(s) = \frac{1}{\sqrt{1 + 2\left(\dfrac{v}{c}\right)}} \frac{\displaystyle\sum_{-\infty}^{+\infty} B_n' e^{jn\pi}}{\displaystyle\sum_{-\infty}^{+\infty} \frac{J_n(ka)}{H_n^{(2)}(Ka)} e^{jn\pi}}. \qquad (7)$$

The definition of the model is given by (7).

3. Comparison of TM and TE modes

It is interesting to compare results of different types of wave polarization (TM and TE). This paper studies the scattered field of the scatterer under rotation and translation when the polarization of incident wave is in TM-polarized form. The result shows that the backscattered field is changed during the translation of the scatterer with the introduction of the term, $\left(1 + 2\dfrac{v}{c}\right)$, as shown in (5). Moreover, in the case of the TM-mode, the backscattered field is not affected by the rotation of the scatterer as shown in (5). In the case of TE-mode [1], it was found that the pattern of the phase and magnitude of the backscattered field were affected by the rotation and translation of the scatterer. The effect of rotation shows up with Ω terms as shown in (4) [1], in addition to the effect of the translation as evidenced by the presence of $\left(1 + 2\dfrac{v}{c}\right)$ terms, similar to the case of the TM-mode. Moreover. the effect of rotation upon the scatterer is more evident for the TE-mode than the TM-mode. Also, the effect of translation of the scatterer upon the phase and magnitude of the backscattered field is evident in both cases (TM and TE).

4. Numerical Results

This section includes the simulation results of the new model. The conducting cylinder, which is shown in Figure 1, is modeled using FEKO. It is seen that the incident electric field is z-polarized with the incident plane wave is in the x-direction. It is assumed that

the conductivity of the moving cylinder is $\sigma = 5.76 \times 10^7 \ s/m$. The height of the conducting cylinder is $0.1\,m$ and its radius is $0.025\,m$. The simulation frequency changed from 10GHz to 14GHz. The static data, generated using FEKO, is inserted into the model as shown in Figure 2 to evaluate the effects of rotation and translation at certain different values of $\beta_a = \Omega a/c$ and v/c, where St and Mo indicate to the stationary and moving cases respectively. Figures 3 and 4 show that the results of both methods are identical and give similar behavior where the slopes of the backscattered magnitudes are equal as shown in Figure 4. Also, the phase and magnitude of the backscattered fields are not affected when the cylinder is rotating with angular velocity $\left(\beta_a = 0.02 \right)$ and translating with relative velocity $\left(v/c = 0 \right)$. However, after the relative velocity is increased to $v/c = 0.02$ and $\beta_a = 0.02$, it is seen that the phase and magnitude of the backscattered field are shifted when the direction of translating cylinder is in the positive x-axis as shown in Figures 5 and 6. It is important to see the phase and magnitude when the translation of the rotating conducting cylinder is in the opposite direction. It is assumed that the relative velocity of the rotating cylinder is $\dfrac{v}{c} = -0.02$ and $\beta_a = 0.02$. It is shown the magnitude of the backscattered field is shifted higher in frequency but it has the same slope as shown in Figure 7. The pattern of the phase is shifted lower as shown in Figure 8.
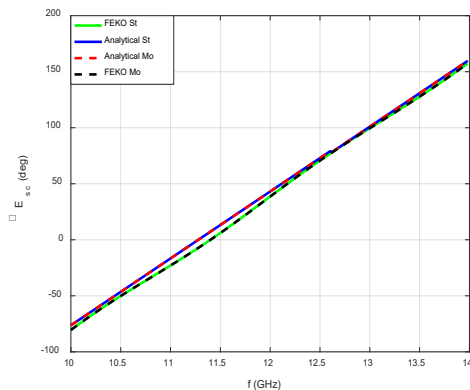
5: Magnitude of TM at $\beta_a = 0.02 \ and \ v/c = 0.02$

$\beta_a = 0.02 \ and \ v/c = 0.02$

3: Magnitude of TM at $\beta_a = 0.02 \ and \ v/c = 0$

$\beta_a = 0.02 \ and \ v/c = -0.02$
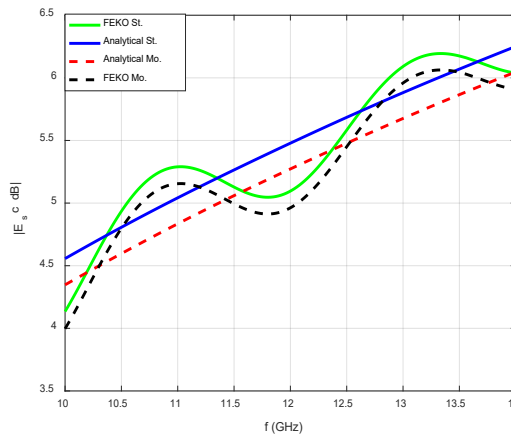
4: Phase of TM at $\beta_a = 0.02 \ and \ v/c = 0$

The phase of the backscattered fields for TM mode and TE mode is shown in Figures 6 and 10, respectively. It is noted that the effect of rotation of the conducting cylinder is clearly observed by the periodicity in the backscattered phase in the case of TE mode, the period of which increases or decreases when the speed of the rotation decreases or increases respectively as shown in Figure 11 [1]. The effect of the translation is evident in both modes
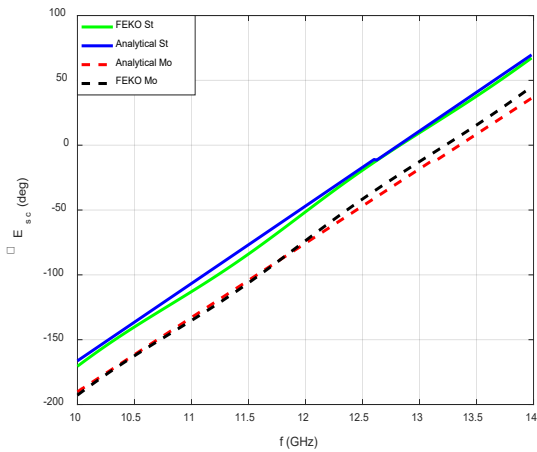
$$\beta_a = 0.02 \ and \ v/c = -0.02$$
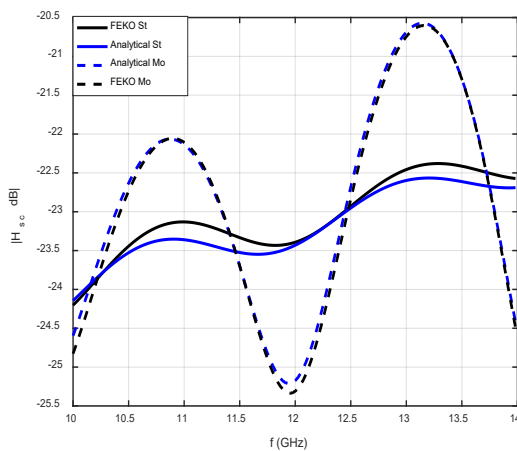
$$\beta_a = 0.02 \ and \ v/c = 0.02$$
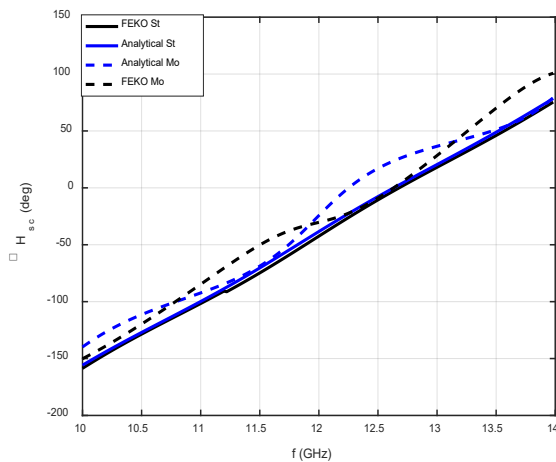
$$\beta_a = 0.02 \ and \ v/c = 0.02$$

(TM and TE) as shown in Figures 5 and 9, and is observed as a shift in the phase and magnitude of the backscattered field. The value of this shift is related to the speed of translation and the direction of the shift is according to the direction of the translation of rotating conducting cylinder. If the translation is along the positive x-axis, the phase is shifted higher as shown in Figures 6 and 10 and the magnitude is shifted lower as shown in Figures 5 and 9.

5. Conclusion

This work studies the effect of the movement of a complex scatterer upon scattering from an incident TM polarized wave. Two types of transformation (Franklin and Lorentz) are used together to analyze the behavior of the phase and magnitude of the backscattered field from the movement of the conducting cylinder. In the TM-mode case, the backscattered phase increases (shifted higher) and the magnitude of the backscattered field decreases (shifted lower) when the direction of translation is parallel to the direction of the incident wave. When the direction of translation is opposite to the direction of the incident wave, the backscattered phase decreases (shifted lower) and the backscattered magnitude increases (shifted higher). The analytical expression and simulation results show that the effect of rotation of the conducting cylinder on the backscattered field is not evident in the case of TM mode. The result of the comparison between the pattern of the backscattered field for both polarizations (TM and TE) shows that the backscattered phase is affected by the rotation of the conducting cylinder, but this effect is more evident in the TE case. The sinusoidal behavior of the backscattered phase is caused by the rotation of the conducting cylinder, whose periodicity reveals important information about the scatterer. Moreover, the magnitude of the backscattered field is affected by the translation in both polarizations. This effect appears as a shift, according to the direction of the translation of a rotating conducting cylinder. Finally, this study attempts to improve the fidelity of modeling the return of an EM scatterer. An extension of this work is planned that would evaluate the effects of the relative movement of the scatterer upon the scattered field in the case of the oblique incident wave.

References

[1] E. Abuhdima and R. Penno," A New Model for Simulation of Scattered EM fields from a Conducting cylinder in Rotation and Translation using Static Data," NAECON-OIS, 2017.

[2] D. De Zutter, "Scattering by a rotating circular cylinder with finite conductivity,"IEEE Transactions on Antenna and Propagation, vol. AP-31, No. 1, PP. 166-169, 1983.

[3] P. Hillion, "Scattering by a rotating circular conducting cylinder I, "Mathematical physics, vol. 41, 1998.

[4] P. Hillion, "Scattering by a rotating circular conducting cylinder II, "Mathematical physics, vol. 41, 1998.

[5] John K. Christensen and Michael J. Underhill,"Doppler Measurements of Smooth and Rough Surface High Frequency Scattering from Spinning Steel Cylinders," IEEE A&E Systems Magazine , 2004.

[6] E. Abuhdima and R. Penno,"Simulation of the scattered EM fields from a rotating conducting cylinder," IEEE International Radar, 2015.

[7] E. Abuhdima and R. Penno,"An Improved Model for the Phase of Backscattered Electromagnetic Fields from a Conducting Rotating Cylinder," NAECON-OIS, 2015.

[8] E. Abuhdima and R. Penno, "Simulation of Scattered EM Field of a Rotating Conducting Number Cylinder using Static Data," Journal of Energy and Power Engineering, Volume 11, Number 6, June 2017.

[9] Harrington, Time Harmonic Elecromagnetic Fields, McGraw-Hill, 1961, pg 234.

[10] Philip Franklin, "The meaning of rotation in the special theory of relativity," Proc. Nat. Acad. Sci., vol. 8, No. 9, 1922.

[11] E. Abuhdima, "Simulation of the Scattered EM field of Moving Dynamic Object using Static Data," Ph.D Thesis, University of Dayton, 2017.

[12] C. A. Balanis, Advanced Engineering Electromagnetics, John Wiley &Sons, Inc,(1989).

[13] E. Abuhdima and R. Penno,"The Effect of Rotation and Translation upon the Scattered EM Fields of a Conducting Cylinder," NAECON-OIS, 2016.

Protecting Private Data using Improved Honey Encryption and Honeywords Generation Algorithm

Thanda Win[*], Khin Su Myat Moe

Yangon Technological University, Department of Computer Engineering and Information Technology, 11010, Myanmar

A R T I C L E I N F O	A B S T R A C T
	Nowadays, many companies and organizations use various encryption techniques to secure their private data before send it through the unsecure network. However, many attackers try to store private data by using various attacks. Most of the organizations use password based encryption algorithm (PBE) to protect their private data. However, the existing PBE methods are vulnerable in brute-force attacks because of user-generated weak or repeated passwords. The problem of weak password based encryption algorithm prompted us to introduce honey encryption (HE). Honey encryption and honeywords help to minimize this vulnerability of password based encryption algorithm and they can prevent the text based messages from brute face attacks in order to make messaging communication more secure and efficient. However, the conventional HE has message space limitation in Distribution Transforming Encoder (DTE) process and storage overhead problem in honeywords generation process. Therefore, our proposed honey encryption algorithm uses discrete distribution function in DTE process to solve message space limitation instead of using cumulative distribution function. Storage overhead and typo safety problem can overcome by using our proposed honeywords generation method in honey encryption algorithm. Furthermore, we also propose hashing and salting algorithm for securing password and key expansion. In this paper, we describe the case studies as a calculation of DTE and hashing and salting algorithm. And then, we design and implement the improved honey encryption mechanisms and honeywords generation algorithm in message transmission process. Finally, we show the comparison results using mathematical model in DTE process and security analysis of our system.

1. Introduction

Most of organizations over the world want to exchange their private messages secretly over the unsecure communication channel. So, they use end to end encryption services with various data encryption algorithm for their private data. One type of encryption process called the end-to-end encryption (E2EE) can read the messages which own the secrete keys [1]. Most of the organizations use password based encryption (PBE) algorithm because user can select and remember their key easily. However, the existing PBE algorithms are weak because hackers can get the message easily if they find the keys using the various attacks. Among them, brute force attack is the vulnerable to the applications and websites using encryption algorithm because these attacks attack the password files in the database until the real key or passwords find [2]. So, many researchers try to prevent the brute force attacks by using various methods.

The two countermeasure methods to brute force attacks have been applied by many modern systems. The first method is to enhance the time complexity for attackers. When the time complexity increases, the attackers take many computation times for cracking the passwords and many systems protect the data by increasing key length pseudorandom number generator (PRNG). Therefore, many hashing algorithm such as MD5 and SHA256 are invented for improving computations times [3]. However, the brute force attack can reveal the key within the polynomial time. So, the honeywords generation methods are used in many cryptosystem for protecting the keys. The attackers can face the difficulties and can't stole the password file if the system uses honeyword generation method that stores the honeywords with the real passwords in the password files. The purpose of decoy passwords or honeywords are used for the attackers that attack the hashing password file using various attacks especially brute force attacks. So, the existing honeywords generation methods generate

*Thanda Win, Yangon , +959797102699, thanda80@gmail.com

at least 20 honeywords for one real password [4]. So, the existing method meets the storage overhead problem.

The second method is statistical coding scheme that can produce false plaintext messages [3] using ASCII code table. But, the attacker can easily know this false message is not the real message because these messages are meaningless messages. HE is an encryption algorithm that can produce a ciphertext which when decrypted by any wrong key by the brute force attacker, then it looks like meaningful fake plaintext which is not the original message. So this encryption provides security against brute force attack. Distribution transforming encoder (DTE) is the main part of HE and it performs the mapping of the message space into the seed space of binary bits string. According to the probabilities of a message in the message space, it maps the message to a seed range using DTE process in a seed space randomly. Then the resulting seed space is XORed with the key to get the ciphertext. For decryption, the ciphertext is XORed with the key and the seed is obtained. Then DTE uses the seed location to map it back to the original plaintext message. Even if the key is incorrect, the decryption process produces a honey message from the message space and thus can deceive the attackers [4]. In this existing HE algorithm, the message space can't accept more than four messages.

This paper described three contributions. Firstly, we report new DTE process for overcoming message space limitation problem in HE algorithm. And then, we design and implement the improved honey encryption system using new DTE process and apply the concept to message transmission application. These applications are based on uniformly distributed message spaces and the symmetric encryption mechanism. We also propose the new honeywords generation algorithm and hashing algorithm. Secondly, we apply two countermeasure attacks such as honey encryption and honeywords generation algorithms in message transmission process. Thirdly, we discuss comparison results of our proposed system and existing systems.

2. Related Works

For protecting the brute force attack, we firstly described two different methods such as hashing algorithm and statical code scheme using ASCII code table. However, these two methods can't fully against the brute force attacks. So the honeywords generation method and honey encryption algorithm are reported in the introduction. Although two methods can protect the brute force attack, they have weakness in their process. In this part, we describe the weakness of password hashing and ASCII code table. And then, we described the limitation problem of honeywords generation methods and honey encryption process as follows.

In the past year, D. Gross described the dangerous hashing passwords that are used in the various websites such as Evernote, Yahoo, LinkedIn and many other websites [5]. Due to the password cracking technique proposed by M. Weir, S. Aggarwal, B. de Medeiros, and B. Glodek, the hashing password can easily know by the attackers [6]. P.G. Kelley, S. Komanduri, M.L. Mazurek, R. Shay, T. Vidas, L. Bauer, N. Christin, L.F. Cranor, and J. Lopez reported most of the passwords of users are weak and easy guessed password. So, their proposed method can easily know the hashing passwords of user [7]. The attackers can get three billion guesses per second if the system using some hashing

algorithm [8]. Therefore, passwords are stored using hashing algorithm in the database are not against the various attacks.

S. Kharod, N. Sharma and A. Sharma proposed a hashing method using salting and differential masking that can provide the better security of the password with the faster time [9]. Although this method is better method for the passwords with better processing time, it can't strongly protect the brute force attack.

Above the description, hashing algorithm that uses for increasing computation time and producing false plaintext messages are not fully protect the various attacks especially brute force attack. Therefore, honeywords generation algorithm appeared in 2013 and honey encryption described in 2014.

In order to improve the safety of the hashed passwords, A. Juels and R. L. Rivest suggested a new method called honeywords generation method. They maintain the honeywords with the real password for each user's account by creating honeywords. The attacker cannot classify which password is real password if he gets inversion file of hashed password or honeywords. The auxiliary server or honey checker can classify the real password and honeyword for login process and will set off an alarm to categorize between the honeywords and real password [10]. However, this method can cause typing mistake of users during entering of password because honeywords generation method create the similar honeyword with the real password. The main weakness of this method is storage overhead problem.

Due to the development of graphical processing unit (GPU), the adversary can solve the hashing password files. So, Z. A. Genc, S. Kardas and M. S. Kiraz proposed the new method by keeping the decoy password and real password into the database. In this work, the author described the honeychecker for testing entering the false passwords to the system. Although getting the password file and coverting the hash code into the password, he can't enter the system without passing the honeychecker [11]. Therefore, this method can against the attacks of attackers, it meets the storage overhead problem.

For making less the existing problem of honeywords generation method, N. Chakraborty and S. Mondal was created the honey circular list algorithm. In this paper, the authors described circular list for storage of hashing password. When the attacker gets the password file, he cannot compute the distance of this password and he cannot login into the system. The honey circular list method can make less the storage problem of earlier existing algorithms [12]. However, this method can't almost solve the problem of honeywords producing methods.

S. R. Shinge and R. Patil presented a data encryption and decryption that uses values in ASCII code table. In this ASCII code table encryption approach, the plaintext messages and encryption keys are converting into the string using ASCII values [13] for producing ciphertexts. If an adversary stoles the key file and then he tries to decrypt the ciphertext use one key in this key files, he gets the meaningless ciphertext if he uses the false key. So, the attacker can know easily he uses the false key in this process.

A new encryption technique that can protect brute force attack called honey encryption (HE) algorithm was proposed by N. Tyagi, J. Wang, K. Wen and D. Zuo. When the attacker tries to get decrypted data by using brute force attack, that technique

can deceive the attacker and can give bogus meaningful messages. A new encoding and decoding scheme called a distribution transforming encoder (DTE) are used by HE and it has the limitation for assigning the plaintext messages [4]. The DTE in that system has limitation for placing plaintext message into the seed space.

According to the previous works, the existing system meets the storage overhead and message space limitation problems. Therefore, we propose a system to overcome message space limitation and storage overhead problem. This honeywords generation system can make less storage cost compared to the existing honeywords generation algorithm. Moreover, it can easily overcome the typo safety problem. For securing password files, we use new hashing and salting algorithm. The hashing and salting time is faster than the existing MD5 hashing algorithm and hashing algorithm using salting and differential masking . Finally our system using new honeywords generation and hashing algorithm, new DTE process can overcome message space limitation problem using DTE compared with the existing honey encryption algorithm.

3. Proposed Methodology

This part reports about the methodology of the proposed techniques such as the user authentication and login process, honeywords generation method, hashing and salting algorithm, distribution transforming encoder and honey encryption algorithm.

3.1. User Authentication and Login Process

In any login procedure, when the user is a new member, he needs to make the registration process. Otherwise, the user enters the system with his username and password. After making the registration process, the system produces the honeywords for this member using honeywords generation algorithm and the sweetwords that includes honeywords and real password are kept into the password file using our proposed hashing and salting algorithm. The registration process and login process are shown in the Figure 1(a) and 1(b). After the registration process, he can login into the system with his password.

The server detects the login password whether it has in the database or not. If the password exists in the database, the password is sent to the honeychecker by the server for categorization of honeyword and actual password. When the login password doesn't equal to the user's real password that stores in the database, administrator can get the alarm message from the honeychecker. Otherwise, the user can enter into the system successfully.

3.2. HoneyChecker

Honeychecker is an auxiliary secure server that can be used for categorisation of sweetwords including real passwords and false passwords. The honeychecker stores secret information such as the index of real passwords and the hashing passwords. The main server can communicate with the honeychecker when a user enters the system. When the false password or honeyword is entered into the system, the system administration can know immediately by raising an alarm signal from the honeychecker. In our system, the honeychecker performs two main processes. The

first process is to classify the honeywords or real password when the user logging the system. The second is to send an alarm message to the administrator if the attacker enters with the honeywords.

(a)

(b)
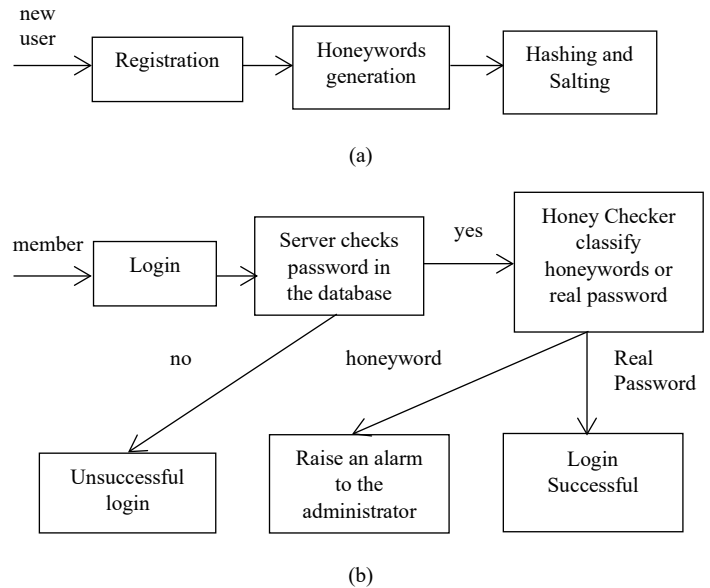
Figure 1: (a): User Registration Process (b): User Login process

3.3. Honeywords Generation Method

For every web application, passwords are notoriously weak authentication in the authentication process because users choose easily guess passwords by the attackers that apply dictionary, rainbow table, brute force attacks etc.. Therefore, as many technologies such as graphic processing unit (GPU) are developing, the attackers can get the user's password and can crack easily hashing password without consuming many times. So, the real passwords and the honeywords that are generated from our proposed honeywords generation algorithm are kept using our proposed hashing and salting algorithm for each account in the database by the system. The purpose of honeywords generation methods generates the honeywords to deceive the attackers [14]. However, the existing honeywords generation algorithms have the storage overhead problem.

Table 1. Example of password file P1 in main server

user name	honeyword index set
David	(21,35,40,55)
Suzy	(18,35,40,55)
Jane	(18,21,40,55)
Joy	(18,21,35,55)
August	(18,21,35,40)

Table 2. Example of password file P2 in honey checker

real password index	hashing password
18	763f7b09c8ac104bd
21	3e8fdfcd7f3615428b0c
35	7e8fcb6b2bc
40	365b4f15a8d445d
55	984d078ac47c2c82987d19

So, we proposed the new honeywords generation algorithm for making less the storage cost. In our proposed algorithm, we denote the member's passwords of our system as honeyword without creating and storing the individual honeywords in the database. Therefore, we create two password files P1 and P2 in the server's database and honeychecker's database. The password file P1 that includes username and honeyword index set in the main server's database. Real password index and hashing passwords including in password file P2 are kept by the system in the honeychecker's database. The table is sorted randomly according to the user registration position. The example storing of honeywords and real passwords are shown in table 1 and table 2.

The detail procedure of honeywords generation is described as follows.

Inputs:

1. username u_i and password p_i

2. After creating accounts, index list of real password and honeywords need to assign

Procedure

Step 1: User account registration and honeywords creation

a. P1 and P2 password files creation

   P1 stores username and honeyindex set h_i, P2 keeps the index of real password r_i and the hash of the corresponding user's passwords $H(p_i)$

b. The list index set of honeywords h_i and real password p_i are created like the following

   P1= {<u_1,h_i>, <u_2,h_2>, --------------,<u_i,h_i>}

   P2= {<$r_1,H(p_1)$>, <$r_2,H(p_2)$>,-------, <$r_i,H(p_i)$>}

Step 2: User passwords and fake password storage:

a. P1 including username and honeyindex are set in main server

b. P2 including index of real password and hashing passwords are set in the honey checker

Step 3: HoneyChecker Classification

   Set r_i, $H(p_i)$

   Sets correct password index r_i for the user u_i

Set: Login password k

Check: r_i, k

If r_i is equal to the login password k, login process successful.

Otherwise, the honeychecker sends an alarm message to the administrator.

3.4. Hashing Algorithm

In our proposed system, honeywords and real passwords are stored into the database using our proposed hashing algorithm for providing better security of key or password with faster time. The detail steps of our proposed hashing algorithm are described the following.

Step 1: Change the user's entering password string into binary bits.

Step 2: Binary bits string are appended with padding bits 00000000.

Step 3: Change the string of binary bit 1 into 0 and 0 into 1.

Step 4: The changing string are make the XOR operation with the binary string by concatenation of 0's and 1's.

Step 5: The string rotates from right by 4 characters.

Step 6: The resulting binary string converts to hexadecimal string and stores into the password file that locates in the database.

The example of our hashing algorithm is shown as following. In this example user chooses her password as December30 and our proposed hashing algorithm converts into the hexadecimal string using the following steps.

Step1: Convert the input expression (password: December30) into binary string.

01000100 01100101 01100011 01100101 01101101 01100010 01100101 01110010 00110011 00110000

Step 2: Before making the password hashing, binary bits string are appended with padding bits 00000000 (green color).

00000000 01000100 01100101 01100011 01100101 01101101 01100010 01100101 01110010 00110011 00110000

Step 3: Change the string of binary bit 1 into 0 and 0 into 1.

11111111 10111011 10011010 10011100 10011010 10010010 10011101 10011010 10001101 11001100 11001111

Step 4: The changing string are performed the XOR operation with the binary string by concatenation of 0's and 1's.

String1:

11111111 10111011 10011010 10011100 10011010 10010010 10011101 10011010 10001101 11001100 11001111

String 2:

00001111 00001111 00001111 00001111 00001111 00001111 00001111 00001111 00001111

After XORing operation, the resulting binary string are described as follows.

11110000 10110100 10010101 10010011 10010101 10011101 10010010 10010101 10000010 11000011 11000000

Step 5: The resulting string rotates from right by 4 characters. So, first 4 characters are rotated to the right.

00001111 00001011 01001001 01011001 00111001 01011001 11011001 00101001 01011000 00101100 00111100

Step 6: The resulting rotate string converts to hexadecimal string and stores into the password file that locates in the database. So, the resulting hexadecimal string is as

follows.

Hexadecimal String - 7B49593959D929582C3C

3.5. Distribution Transforming Encoder

Distribution transforming encoder (DTE) process is the main key of honey encryption algorithm that assigns the message space M to the binary bits string of seed space S. Depending on the number of messages, the range of seed space assigned for distribution. The existing DTE process uses cumulative distribution function to map the message space into the seed space and it meets the message space limitation problem [4]. The DTE process includes encoding and decoding. In the encoding process, the DTE apply the discrete distribution function for assigning the plaintext messages M to the seed space S. The DTE process encodes the plaintext messages M to the range of seed space S is assigned randomly. On the other hand, the decoding process is harder than encoding process. In the decoding process, the seed space S decodes into the message space M using DTE process. One of our main contributions is to construct the DTE process using discrete distribution function to overcome message space limitation problem.
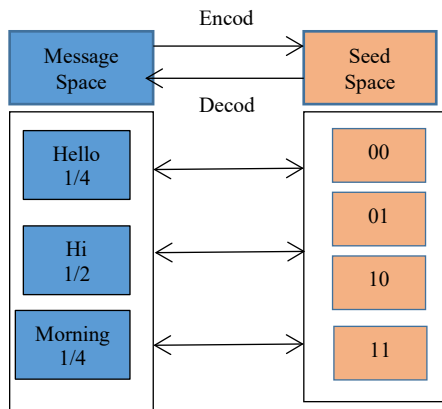
Figure 2: Basic Example of DTE Process

Figure 2 illustrates a basic example of DTE process. In this figure, message space consists of different elements in this case M={ hi, hello, morning}. Depending on the number of messages, the probability of each element using discrete distribution function is generated.

3.6. Honey Encryption Algorithm

Password-based encryption (PBE) and hashing techniques cannot fully against the brute force attacks since most of the users that use password as encryption key select the weak and repeatedly user-generated passwords. Honey encryption (HE) algorithm is one of the deceive algorithm to attackers who use brute force attack. The purpose of HE algorithm is to lure the attackers. HE generates the honey message through the brute force attack and deceives the attackers in a way that they cannot classify which messages are true messages. The core creation of existing HE is the distribution transforming encoder (DTE) that uses cumulative distribution function. Constructing the better DTE process can provide to get better HE algorithm [4].

The contribution of this paper consists of three main categories: honeywords generation, improved hashing algorithm and distribution transforming encoder. In our honeywords

generation algorithm, storage overhead problem can be reduced. Our improved hashing algorithm is faster than existing MD5 hashing algorithm. Our improved DTE process that uses discrete distribution function can improve the message space limitation. Figure 3(a) and 3(b) show the basic encryption and decryption process of honey encryption algorithm.

(a)

(b)
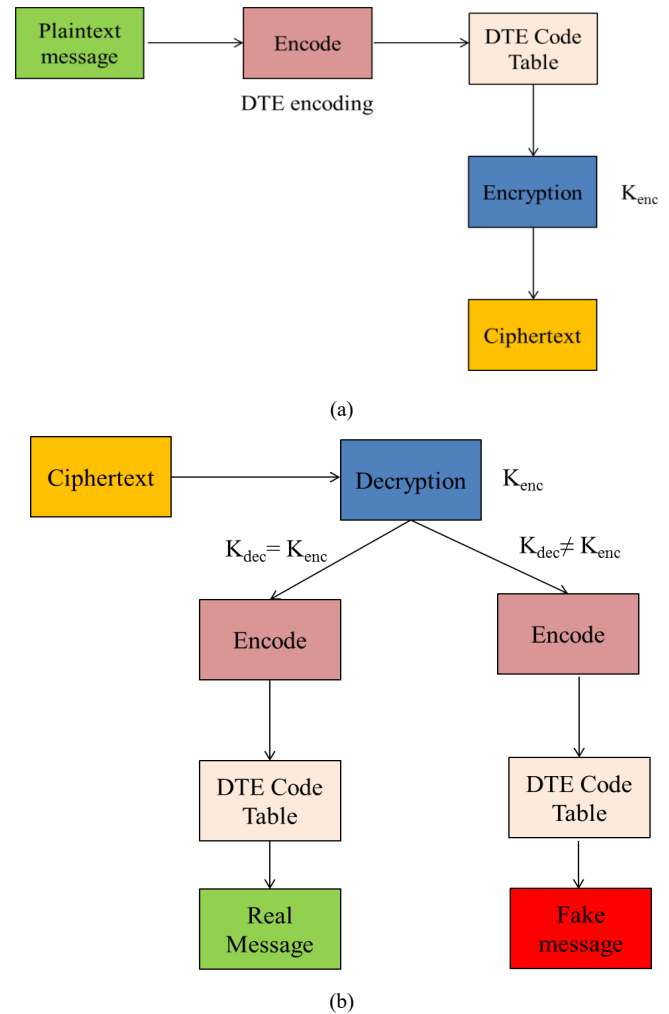
Figure 3 (a) Basic Encryption Process of HE Algorithm (b) Basic Decryption Process of HE Algorithm

The encryption process includes the plaintext message M encoding process using encoding code table and encrypting with encryption keys K_{Enc} to produce the ciphertext. The code table is made by using DTE process that use discrete distribution function. The sender sends resulting ciphertext to the receiver via the internet.

In the decryption process, the receiver decrypts the ciphertext with K_{Dec}. The decryption result decodes with the code table that equals the code table using in encryption process. If encryption key K_{Enc} and decryption key K_{Dec} are the same, the receiver can gets a *real* message that the sender sends. Otherwise, the receiver will get a bogus message that look like real message.

4. Proposed System Implementation

This section discussed about the flowchart of our proposed system, system design, example of the proposed system and mathematical model of the proposed system.

4.1. Flowchart of Proposed System

Figure 4 describes the detail flowchart of the proposed system.
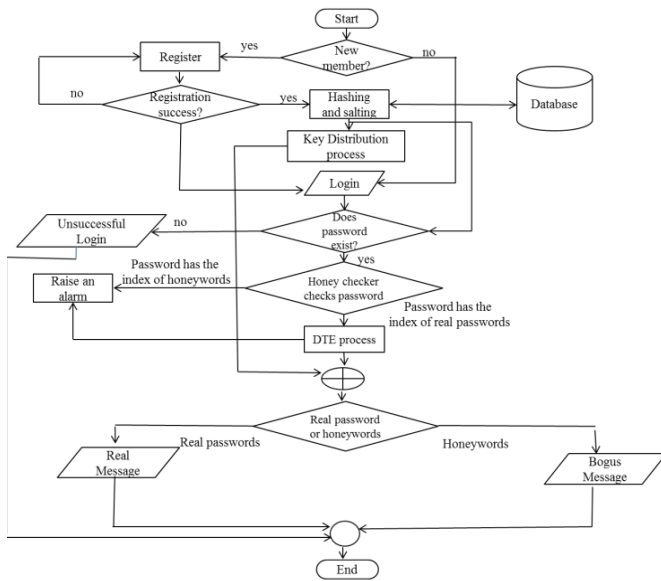
Figure 4: Flowchart of Proposed System

If a new user wants to enter this system, he needs to make the registration process at least one time. He doesn't need to make the registration process to enter the system if he is a member of the system. After making the registration process, he can login into the system with his username and password. If he enters into the system with his username and password, the system checks his password exists or not in the database. In this stage, he can get three conditions. Firstly, he can get unsuccessful login message from the system if his password doesn't have in the database. At that time, he needs to check his username and password if he is a member of this system. Otherwise, he needs to make registration process. His password will be sent by the server sends his password to honey checker for categorization of honeyword or sugarword if his password exists in the database. Secondly, if his password is real password, he will perform the DTE process and gets the real message from the sender's send. Otherwise, he will perform the DTE process and he will get the bogus message.

4.2. System Design of the Proposed System

The process of our proposed system includes DTE process and key or password distribution process. DTE process makes encoding and decoding of original plaintext message. DTE that uses discrete distribution function for encoding message space M also called plaintext message such as $5000, $10000, $750, $900, $200 into the seed space S. The range of seed space denotes as binary bits. The number of plaintext message space depends on the directly proportional to the range of seed space. For placing the number of messages m in plaintext message M to seed space S_m, DTE requires discrete distribution function in order to overcome message limitation.

In keys or passwords distribution, honeywords generation and key expansion of this password file of honeywords and real

passwords are considered. In the honeywords generation process, we generate the honeywords by denoting the other user's real passwords and stores as honey indexes for overcoming weakness of existing honeywords generation methods such as storage overhead problem and typo safety problem. By using the other's sweetword as decoy password, the adversary becomes complicated which one is correct password or honeywords. In this case, the password and honeywords are stored in the password file using our proposed hashing algorithm to provide better security. And then, the resulting password and honeywords are randomly placed into the seed space S_k using DTE process. After placing password file into the seed space S_k, the key expansion result S_k is XORed with the message distribution result S_m of DTE. Finally, the ciphertext message is successfully created. During the decryption process, the ciphertext is XORed with the secret key to get the seed space S_m. And then we decode this seed space S_m to get original plaintext message.
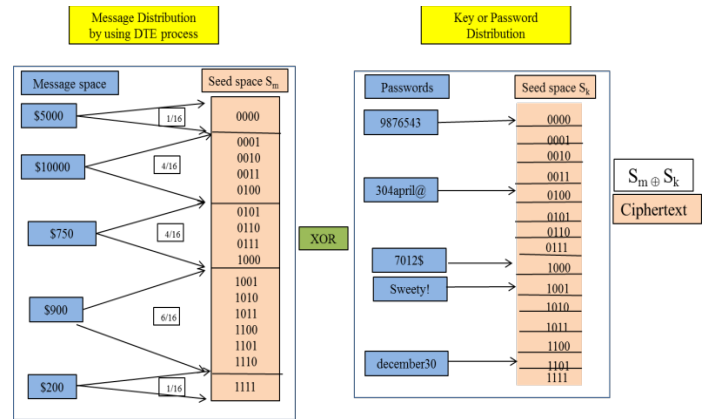
Figure 5: Proposed System Design

4.3. Example of Our Proposed System

Figure 5 describes the detail example design process of our proposed system. In this example, different messages such as M = {$5000, $10000, $750, $900, $200} are denoted as message space. In this case, message space M has five elements. According to the number of messages, probabilities are assigned to each element. By calculating, the seed space range of message M is $2^{5-1} = 2^4 = 16$ and seed space is 4-bits binary strings. Depending on the results of probabilities values, different messages elements are mapped into the range of seed space by the system. For assigning the message space, the messages are ordered randomly. With this randomly ordering, the messages get different probabilities and the range of seed space will be different. The secrete key or passwords are K = {9876543, 30april@, 7012$, Sweety!, december30}.

Using discrete distribution function, probability of the messages is {1/16, 4/16, 6/16, 4/16, 1/16}. So, the seed space S_m of the message "$5000" is {0000} and "$10000 is {0001, 0010, 0011, 0100}. And then, seed space S_m of $750 is {0101, 0110, 0111, 1000, 1001, 1010} and $900 is {1011, 1100, 1101, 1110}. Finally, the seed space S_m of $200 is {1111}. According key distribution process, the seed space S_k of 9876543 is {0000} and

april30@ is {0100}. The next 7012$ is {1000} and Sweety! is {1001}. And then, the seed space S_k of december30 is {1101}.

In the encrypting process, Bob wants to send encrypting message "$5000" to Alice under the secrete key as april30@. So, we compute seed space S_m by using DTE process and then the result S_m of $5000 is (0000).Similarly, according to the key or password distribution, seed space S_k of password april30@ is (0100). And then, Bob sends the resulting ciphertext (0100) by XORing the seed space S_m and S_k to Alice.

The decryption process is harder than encryption process because it includes message recovery process. In this process, Alice tries to decrypt ciphertext C by using password april30@ of Bob. Firstly, Alice computes seed space of key distribution S_k and this S_k (0100) is XORed with the ciphertext (0100) to get seed space of message distribution S_m (0000). And then, he decodes S_m by using DTE process to get original plaintext message "$5000".

On the other hand, attacker John wants to get Bob message. So he attacks password file by using brute force attack. John gets password file but he doesn't classify which is Bob password. So, he randomly chooses password 9876543 and tries to decrypt Bob message. John computes seed space of 9876543 to get key distribution S_k (0000). And then he makes XOR process of ciphertext C and seed space S_k. The resulting seed space of message distribution S_m is (0100). And then, he decodes S_m by using DTE process to get Bob message. Finally, John gets "$10000" after decoding process of seed space of message distribution. In this process, he doesn't know which message is true message or decoy message. This encryption algorithm can make confusion for attackers and it can strongly protect brute force attack.

4.4. Design Methodology

The methodology proposed in this research work is based on Honey encryption algorithm using honeyword Generation Method. The above mentioned work is divided into various modules as follows:

- Registration: At the time of registration, entering password, the system converts password into hash values and store into database. Along with hash values the real password's hash is also store at specific random position. After entering encryption key for encryption and decryption, system will generate honeywords for deceiving the attackers.
- Honeyword Generation: Honeywords are generated from the real password and in case any hacker tries to hack into the account by guessing the password the main user is sent alerts in form of a mail or some message so he knows that somebody is trying to login using his or her account.
- Login: If a user wants to login into the system, the user will type username and password. If the password matched with the hash password, the user is authenticated.
- Message Encryption and Decryption: User can send encryption message via the communication channel using encryption and decryption key.

- Generating honey message or fake message: If the encryption key is not match to the decryption key, the system will generate the honey message or fake message for counterattacking the attacker.
- Admin Management: Admin has the privileges to control over the mechanism. Admin has the authorization to maintain the user accounts and messages management.
- Hacker: If the hacker tries to break the system and enters any honeyword then the alert is given to the Actual user. When the attacker gets the encryption key file and he tries to decrypt the ciphertext using the honeywords, he will get the honey messages.

4.5. Mathematical Model of Our Proposed System

The core process of HE uses DTE for assigning the message space M into the seed space S_m using discrete distribution function in our proposed system. The discrete distribution function is a branch of probability function. The values z in the sample space Ω of discrete probability f(z) theory must have between 0 and 1 and the total values of f(z) for all values z in the sample space Ω must equal to one. This discrete probability theory satisfies the following two properties [15].

1. $f(z)\varepsilon$ [0,1] for all $x \varepsilon \Omega$
2. $\sum f(z) = 1$

- DTE using Discrete Distribution Function

We denote S be seed space or sample space and the value of seed space is [0,1]. According to the discrete distribution function, we calculate to get one for the total value of probability P(z) for all message in seed space S. Let n be total number of messages, x be number of '0" observed, M be the message space that contains number of messages and m be message. The following examples prove that DTE using discrete distribution function satisfies the properties of discrete probability theory. In this example we use number of five messages M={m1,m2,m3,m4,m5}. Firstly, we calculate seed space S.

Number of messages n= 5

For Distribution transforming encoder,

Seed space $S_m = 2^{n-1} = 2^{5-1} = 2^{5-1} = 2^4 = 16$

S={0000,0001,0010,0011,0100,0101,0110,0111, 1000,1001, 1010, 1011, 1100, 1101,1110,1111}

By using Discrete Distribution Function,

$m_1 = 1/16$
$m_2 = 4/16$
$m_3 = 6/16$
$m_4 = 4/16$
$m_5 = 1/16$

M	m1	m2	m3	m4	m5	Total
P(z)	1/16	4/16	6/16	4/16	1/16	1

- DTE using Cumulative Distribution Function

According to the cumulative distribution function, we should get one for the total value of probability P(z) over all messages in seed space S to satisfy the properties of probability theory [10]. The following example describes that DTE using cumultative distribution function cannot satisfy the probability theory because the total values of probability P(z) of message in seed space S doesn't get one [15]. In this example we use five messages M= {m1,m2,m3,m4,m5} and calculate the probability P(z).

Number of message n= 5

For Distribution transforming encoder,

Seed space $S_m = 2^{n-1} = 2^{5-1} = 2^4 = 16$

By using cumultative distribution function

m1= 1/16 [(1,1)]

m2= 2/16 [(1,2),(2,1)]

m3= 2/16 [(1,3),(3,1)]

m4= 3/16 [(1,4),(4,1),(2,2)]

m5= 2/16 [(1,5),(5,1)]

M	m1	m2	m3	m4	m5	Total
P(z)	1/16	2/16	2/16	3/16	2/16	10/16

5. Results and Discussion

Many security areas such as web page and application using internet service used PBE for protection of their data from various attacks. However, PBE can't solve problem of the various security attacks especially brute force attack because most of the passwords are not much stronger. Our proposed honey encryption algorithm and honeywords generation methods can against brute force attack and many other attacks by deceiving the attacks. Moreover, our proposed method overcome the weakness of existing honey encryption and honeywords generations algorithm. The detail analysis and comparison results are shown in the following.

5.1 Implementation and Experimental Results

In this section we implement our proposed methodology in web application by developing a simulated messaging application and the language used is python programming language.

To evaluate the performance of our proposed honey encryption algorithm, we need to consider the false messages. The testing of the performance was conducted to distinguish between the correct key and incorrect keys. In this testing, we encode the message using the correct key. Then, we decrypt the message using the various incorrect keys. For all the testing, our proposed system generates the fake messages that are different from the original messages. The Figure 6 shows the original plaintext that was transmitted.

An attacker trying several keys is presented with the version of Figure 7 and Figure 8. There is no way the adversary can tell

the original message from the fake message even if he acquires the original message.

Figure 6: Original Plaintext Message

Figure 7: Intercepted Message acquired by Attacker using incorrect key

Figure 8: Intercepted Message acquired by Attacker using incorrect key

5.2 Comparative Analysis of Our Proposed System

Honey encryption (HE) algorithm is an interesting challenges algorithm and it plays an important role in various areas such as

business groups, banking system and computer users. Moreover, HE can protect from large number of attacks. Here it examines the comparative analysis of our proposed system and existing systems are described as follows.

1) Brute Force Attack

The web application is attacked by the attackers using the various attacks. Among them, the most dangerous attack for websites is brute force attack. The purpose of this attack is to steal the passwords and encryption key for obtaining the encrypted data. The attacker attacks the password file using the brute force attack and it finds the possible keys and password until the correct one is found. However, the attackers can't get the encryption key because of our proposed method. Although the password file in the database has been attacked by brute force attack, the attacker cannot classify which is real key or password if he gets password file. Moreover, attackers can take many times comparing into the existing security system because our system uses improved hashing and salting algorithm.

2) *DoS Attack*

The machines or network resources can be stopped and cann't be given any services to the users by the Denial of service (DoS) attack. Our proposed system is intended to reduce the DoS vulnerabilities. When a person tries to login into the system, the system automatically checks login of the user. If the attacker tries to login with his attacking password file, the system raises alarm for user and attack detected messages. And then, the system can block that particular attacker.

3) *Dictionary Attack*

Dictionary attack is one of the types of password guessing attacks and an attacker tries to login by cycling through combination of common words. Generally, this attack succeeds because most of the users choose weak or repeated passwords. However, our system can protect the passwords from the dictionary attack. When the attacker tries to login with randomly chooses an account with guessed password from password file, it can be two probabilities which are successful or fail. First is taking the password from the password file is correct and the next is getting the honeywords. If the attacker login with honeywords, the system raises an alarm for the user account and can detect this attack. Otherwise, if he enters with the real password easily, he can get the real message easily. But he cannot classify this message is real or fake messages.

4) *DTE Process*

The existing honey encryption algorithm and our proposed system denotes DTE as the main process. In this HE process, the message space that includes all plaintext messages is mapped into the seed space using DTE process. The existing DTE process uses cumulative distribution function for mapping the message into the seed space and it meets the message limitation problem. In our proposed system, DTE uses discrete distribution function and it can overcome the message limitation problem.

Table 3. Comparison Results of Existing Methods and Our Proposed Method

Techniques	Brute Force Attack	DoS Attack	Dictionary Attack	DTE Process
Password - based Encryption	Weak	Weak	Weak	-
Honey Encryption	Strong	Strong	Strong	Message Space Limitation
Our Proposed System	Strong	Strong	Strong	Overcome Message Space Limitation Problem

5.3 Comparison Results of Existing Honeywords Generation Methods and Our Proposed Honeywords Generation Method

In this section, we describe the comparison of the honeywords generation methods including our models in terms of flatness, typo safety and storage overhead.

1) *Flatness*

Our proposed system kept the list of s sweetwords that includes the real passwords and honeywords in the passwords files. There are two types of flatness such as perfectly-flat and approximately-flat for computing the security levels of honeywords generation methods. Depending on the types of flatness, we can decide which honeywords generation method is the better method for securing of our applications. Flatness computes that the attackers get how many times the real passwords. Let consider the attacker attacking the probability is w and this attacking probability compares the results of real password probability $1/s$. When $w >= 1/s$, the honeywords generation is approximately-flat and an adversary can guess the real password at random position. Otherwise, it said to be perfectly-flat if $w < s$ [5].

Our proposed system is perfectly-flat and the attacker can't guess the real password position because we choose the honeyword from the other user's passwords. Our system has more advantages in security when the system includes more members.

2) *Typo Safety*

This is the important problem of existing honeywords generation algorithm. By creating the similar honeywords using honeywords generation algorithm, most of the users become typing mistake problem [5]. However, our proposed system can mostly reduce this typo safety problem comparing with the existing methods. Due to the using of other user's passwords as honeywords, the user can't miss with other user's passwords.

3) *Storage Overhead*

In the existing honeywords generation process, the algorithm generates at least twenty honeywords for deceiving the attackers. So, the attackers conflict and can't easily distinguish which one is real password or false password. But,

the system stores at least 21 passwords for one user in the database and the database becomes the storage overhead problem. In our proposed system, we denote s is sugarword that includes real password and honeywords. So, the honeywords become (s-1) because the user has one real password. Among the existing system, paired distance protocol (PDP) method is the best reducing storage overhead problem and it stores only random string of honeywords without storing the whole honeywords [7]. However, our proposed method can mostly reduce the storage overhead than the PDP algorithm because our system uses the indexes of member's password as honeyindexes.

Table 4. Comparison Results of Existing Honeywords Generation Methods and Our Proposed Method

No	Methods	Flatness	Typo Safety	Storage Overhead
1	CTD	1/k	Low	(s-1)*n
2	Password Model	1/k	High	(s-1)*n
3	Take a Tail	1/k	High	(s-1)*n
4	PDP	1/k	High	(1+RS)*n
5	Our Method	1/k	High	(1*n)

CTD- Chaffing by Tweaking

PDP- Paired Distance Protocol

s- Number of sweetwords in password file

n- Number of users

RS- Random String

5.4 Time Comparison Results of MD5 and Our Proposed Hashing Algorithm

Our proposed honey encryption process are designed by python programming language. This python programming language is flexible for all real world system and easy for the programmers to study. In this paper, we proposed three main contributions: improved DTE process using discrete distribution function, new honeywords generation method and the last is improved hashing algorithm for providing the better security of our proposed system with less time complexity comparing to the existing MD5 hashing algorithm. From the experimental results, the MD5 hashing algorithm takes more time than our proposed algorithm for converting the passwords into the hash code because our proposed system use simple binary operation.

Table5. Time Comparison Results of Existing MD5 Hashing Algorithm and Our Proposed Algorithm

		Time Complexity (Milliseconds)			
Length of Passwords (in character)		6	8	10	12
Methods	MD5	1740	2854	3260	3681
	Our Proposed Algorithm	31	46	62	78

In the following table, we uses the user's passwords such as "123456, hello123, student321, December3040" for the comparing the time results of other proposed hashing algorithm and existing MD5 hashing algorithm.

6. Conclusions

Most of the organizations try to against the brute force attack for protecting their private data. Honeywords generation algorithm can make the confusion for the attackers by storing the real password with the false passwords or honeywords in the password files. However, honeywords generation algorithm can meet the storage overhead problem if many of the users use this application. Honey encryption algorithm can protect the brute force attack by producing honey messages if the attackers use the honeywords. The existing honey encryption algorithm has the message space limitation problem and it can't send more than four messages by using cumulative distribution function. Our proposed honeywords generation algorithm and honey encryption algorithm can solve the storage overhead problem and message space limitation problem. Moreover, we proposed new hashing algorithm for securing our system and for getting the faster processing time compared to the existing MD5 hashing system.

References

[1] J.H. Saltzer, D. P. Reed and D.D. Clark, "Review for Paper: End to End Arguments in System Design" Polytechnic University, USA, 2006.

[2] K. S. M. Moe, T. Win, "Enhanced Honey Encryption Algorithm for Increasing Message Space against Brute Force Attack" in 15th Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI) Conference, Thailand, Chaing Rai, 2018.

[3] A. Juels and T. Ristenpart, "Honey encryption: security beyond the brute-force bound," in Advances in Cryptology EUROCRYPT 2014, Springer, Berlin, Germany, 2014, vol. 8441 of Lecture Notes in Computer Science, pp. 293–310.

[4] N.Tyagi, J. Wang, K. Wen and D. Zuo, Honey Encryption Application, Computer and network Security, Springer, 2015.

[5] D. Gross, "50 million compromised in Evernote hack" CNN, 4 March 2013.

[6] M. Weir, S. Aggarwal, B. de Medeiros, and B. Glodek, "Password cracking using probabilistic context-free grammars," In IEEE Symposium on Security and Privacy (SP), pp-162-175, 2009.

[7] P.G. Kelley, S. Komanduri, M.L. Mazurek, R. Shay, T. Vidas, L. Bauer, N. Christin, L.F. Cranor, and J. Lopez, "Guess again (and again and again): Measuring password strength by simulating password- cracking algorithms," In IEEE Symposium on Security and Privacy (SP), pp- 523-537, 2012.

[8] M. Bakker and R. V.D .Jagt, "GPU-based password cracking. Technical report," Univ. of Amsterdam, 2010.

[9] S. Kharod, N. Sharma and A. Sharma, "An Improved Hashing Based Password Security Scheme Using Salting and Differential Masking, in 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Direction)," India, 2018.

[10] A. Juels and R. L. Rivest, "Honeywords Making Password Cracking Detectable" in ACM SIGSAC Conference on Computer & Communications Security, ser. CCS '13. New York, NY, USA: ACM, 2013, pp. 145–160.

[11] Z. A. Genc, S.Kardas and M.S. Kiraz, "Examination of a New Defense Mechanism: Honeywords," in 11th WISTP International Conference on Information Security Theory and Practice, Heraklion, Crete, Greece, 2017.

[12] N. Chakraborty and S. Mondal, "A New Optimized Honeyword Generation Approach for Enhancing Security and Usability," arXiv preprint arXiv: 1509.06094v1, 2015.

[13] S. R. Shinge and R. Patil, "An Encryption Algorithm based on ASCII Value of Data" International Journal of Computer Science and Information Technologies, vol. 5(6), 2014, pp-7232-7234.

[14] A. Juels and R. L. Revist, "Honeywords: Making Password Cracking Detectable" In MIT CSAIL, 2013.

[15] B. Addanki, Computer Physicist, "What is the difference between a probability density function and cumulative distribution function," 2015. http:// www.quora.com.

The Value of Integrating MSRP Protocol in E-learning Platforms of Universities

Khalifa Sylla[1,*], Samuel Ouya[2], Masamba Seck[3], Gervais Mendy[2]

[1] *Virtual University of Senegal, Applied Mathematics and Computer, 15126 Dakar-Fann, Senegal*

[2] *Polytechnic High School of Dakar, Computer Science Department, 5085 Dakar-Fann, Senegal*

[3]*Alioune Diop University of Bambey, Computer Science Department, 16 617 Dakar Fann, Senegal*

A R T I C L E I N F O	A B S T R A C T
	The proposed solution is intended for virtual universities wishing to improve their distance learning platforms. These enhancements include the instructional content model, how to deliver courses to students, and how to evaluate students' knowledge. With this solution that improves the functionality of Moodle, virtual classes now include new features such as the transfer of heavy files, control of remote screens, video, audio, chat without degradation of the quality of service. The solution will be particularly useful for language teachers who have difficulty in assessing the speaking skills of their students. Indeed, the platform with its audio / video features, integrates a tool to assess the knowledge and improve the speaking skills of students. Teachers of technical subjects such as programming or computer networks will also be able to use this solution to improve the way they deliver courses, including making it easier for students to do practical work. *On the technical side, the solution is based on the MSRP-RELAY, MSRP and Websocket protocols. It is therefore compatible with smartphones, tablets and computers. Tests were conducted with teachers of languages and STEM subjects to evaluate their students in direct synchronous encounters without necessarily going through multiple choice questions.*

1. Introduction

This paper is an extension of work originally presented in conference International Conference of Advanced

Communications Technology (ICACT) 2018 in Korea [1]. Currently, in many countries, particularly in Africa, higher education faces several constraints, including the lack of facilities to accommodate new graduates in traditional universities and teaching materials for practical work.

With the democratization of the Internet and the emergence of new ICT services, we can say that technology is at the service of education. Many countries are establishing national interconnection of public buildings by fiber optics. This promotes high bandwidth internet access, the ability to do cloud computing, and so on. Classical universities have taken advantage of these services and now offer Open and Distance Training [2]. For a purely virtual university, it is therefore a challenge to offer solid platforms with advanced features that allow students to carry out

practical work while guaranteeing quality in audio-video communications during these practical work, especially for students. students in STEM [3].

Indeed, open and distance learning was once essential for professionals seeking skills development. Today, with digital universities, a young graduate from the bachelor's degree can start his university course with distance education. This explains why most Senegalese universities are increasingly interested in this educational innovation in their training offerings.

It is therefore very important for these young people to receive quality training, especially those who study STEM. Among other features of the training platforms used by universities to meet this need include communication tools (chat, audio, video), online self-assessment tools, [4] and so on. The Moodle platform [5] is undoubtedly one of the most used today by universities.

Moodle's very popular BigBlueButton [6] plugin allows you to create virtual classes using the WEBRTC [7] protocol. It includes all the tools you need for a virtual classroom (audio, video, screen sharing using flash [8]), but there are some shortcomings that

deserve to be raised. The possibility of entering mathematical formulas and the impossibility of taking remote control of computers seriously handicap the dispensing of certain courses and, by extension, the quality of the students' training.

Given the requirements of distance learning today, we believe that an e-learning platform should no longer be limited to the traditional features we already know. Universities would benefit from including features such as:

- Taking into account the transfer of files with a very large size, regardless of their type;

- The possibility for a teacher or a student to share his screen with the other actors during a distance course ;

 - The ability to remotely control student terminals ;

- The possibility of sharing software with students.

Distance learning also faces serious problems when it comes to assessing students' knowledge of subjects such as modern languages. It is therefore important to think about solving this problem too.

To achieve this, we use the power of the Kamailio server, the MSRP protocol [9] and the Websocket [10]. The purpose of the work is to contribute to the improvement of e-learning platforms and thus to promote better learning for students.

The organization of the article is as follows :

Section II is dedicated to the presentation of the main protocols used namely: MSRP, MSRP-RELAY, Websocket. Section III will describe how our solution works. We thus find the proposed architecture, the presentation of Kamailio and Moodle. In section IV, the results obtained from the tests will be presented. Finally, Section V concluded and gives an opening on future work.

2.   Definition of MSRP, MSRP-Relay and Websocket

2.1. MSRP

The MSRP protocol is used for the exchange of instant messages and any other binary content other than ASCII (MIME)[11]. It is connection-oriented and is particularly useful for exchanging very large files or remote desktop sharing. Since MSRP[12] can not perform a session setup on its own, it is usually paired with the SIP protocol[13]. The SIP protocol is responsible for establishing and releasing the multimedia session.

MSRP defines two request types, or methods. SEND requests are used to deliver a complete message or a chunk (a portion of a complete message), while REPORT requests report on the status of a previously sent message, or a range of bytes inside a message. [8]

2.2. MSRP Relay

 The MSRP protocol has NAT traversal problems. To solve this problem, the MRSP Relay is used in many implementations. [14] This is the case of our proposed solution with Kamailio. The operating principle is very simple. Customers who want to send a message can go through a relay or not. When a relay is used, it is the latter who sends the message to the customer or another relay (Figure 2).
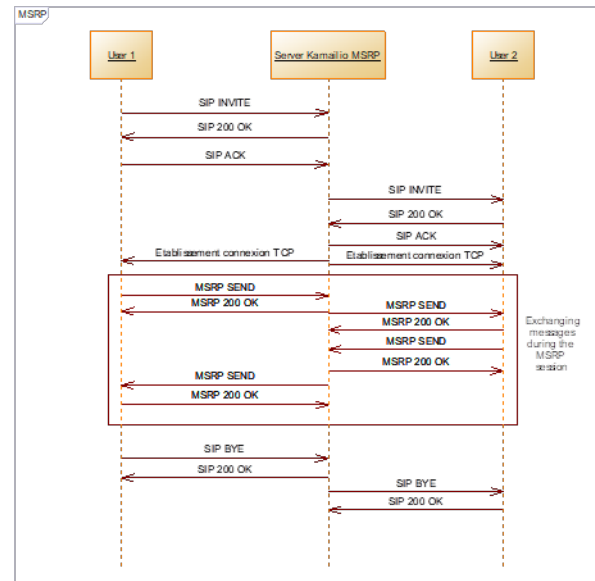
Figure 1: messaging session with MSRP
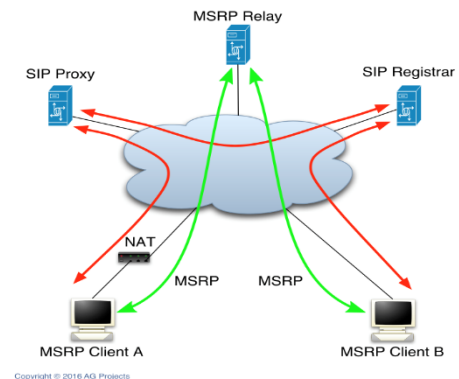
Figure 2: session message with MSRP Relay

2.3. Websocket

The WebSocket protocol enables two-way real-time communication between clients and servers. With this API you can send messages to a server and receive its responses in an events based mode without having to go to the server to get an answer [15]. The protocol has two parts : a handshake and the data transfer. Once the client and server have both sent their handshakes, and if the handshake was successful, then the data transfer part starts. This is a two-way communication channel where each side can, independently from the other, send data at will. The client can request that the server use a specific subprotocol (such as the WebSocket MSRP subprotocol) by including the "Sec-WebSocket-Protocol" field in its handshake [16].

2.4. Websocket protocol as a transport for MSRP

As the name suggests, websocket msrp subprotocol is a subset of the WebSocket protocol. He specializes in the transport of MSRP messages. These exchanges take place between a Websocket client and a node that will be called MSRP relay. It is not superfluous to wonder about the information contained in these messages. These are sometimes MSRP requests, sometimes answers. All this is

done using a Websocket connection [17] In the websocket msrp subprotocol, we find the terms clients and servers. The MSRP WebSocket Client is the entity with the ability to open multiple outbound connections to other MSRP-Relay. These other MSRP-Relay are called MSRP Websocket server. So, an MSRP WebSocket Server is simply an MSRP-Relay responsible for listening to incoming connections. These connections come from Websockets clients. Communication is obviously done with the websocket msrp subprotocol. It should first be that certain rules are respected. During the handshake procedure, both entities agree on the use of the websocket msrp subprotocol for communication. To do this, the MSRP Websocket has the obligation to put in a header called sec-websocket, the keyword 'msrp' in his handshake. Naturally, the server also in responding does the same operation in its header.

3. Architecture of our solution

3.1. Solution Architecture

The solution we proposed works as shown in Figure 3 below. The basis of architecture is the well-known moodle platform. It is to this platform that we added new features via a plugin that has been integrated. This plugin uses a free IP telephony server called Kamailio [17] whose IP address will be set in the configuration file of the module and materialized on the same figure by:

• Presentation Lib: this is the set of libraries of the plugin interacting with the MSRP API [18].

• UI presentation: represents the views of our plugin
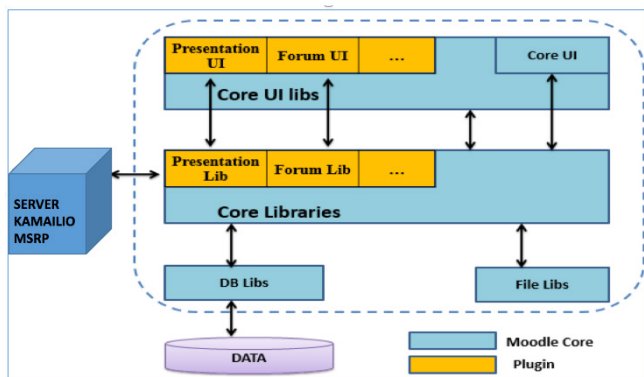
Figure 3: Architecture de la solution

3.2. How the solution works

The Websocket, MSRP and MSRP Relay protocols are the main ones involved in our platform. The solution is based on the interconnection between the kamailio server and the Moodle platform in order to push the limits of Moodlle, especially in terms of heavy file sharing or remote learning of a student's machine in order to show him good practices. The MSRP relay is mainly used to solve NAT traversal problems. Thanks to the loading and activation of the modules, routing rules have been defined especially in case of MSRP transaction. Figure 4 describes message exchanges between two users during an MSRP session. During this session, features such as screen sharing or sending

large files can be used by both students and the teacher. Concretely, the user 1 sends an instant message to the user 2. The WebSocket Kamailio MSRP server receives the message and sends it to the MSRP relay which finally conveys the message to the user 2.
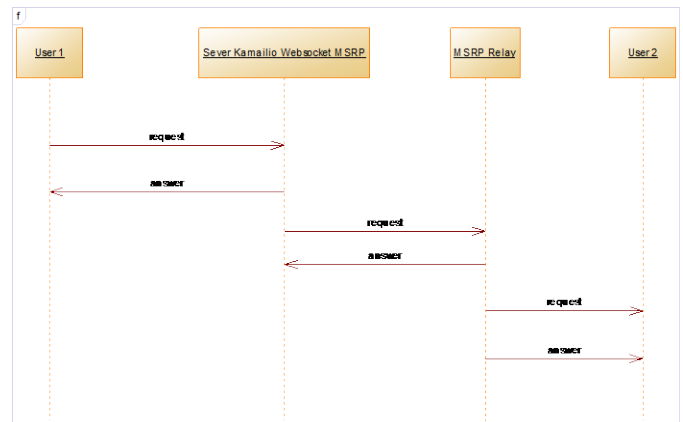
Figure 4: typical MSRP session 1

Figure 5: configuration Server loading modules

Figure 6 : MSRP transaction routing configuration

3.3. Description of Moodle

Moodle is an open source online course management system used by many organizations, universities and even businesses to offer online learning and the online component of blended learning. [19]. It is one of the most used open source course management systems in the world. The acronym MOODLE stands for "Modular Object Oriented Dynamic Learning Environment". It was in 1990 that Marin Dougiamas started designing the platform in Australia. It is an online platform under the GNU Public License (GPL). If the solution is adopted by several universities, it is mainly because it offers a set of integrated tools initially designed

on the basis of the pedagogical model called socioconstructivism [21]. Its user interface is user friendly and the solution works equally well on desktops and mobile devices. From the point of view of administration, Moodle is very ibleexible. For example, many options are available for user authentication and course registration because of the many plug-ins available. These provide organizations with many approaches to managing how they integrate Moodle into their other systems. Figure 2 below shows the architecture of Moodle.
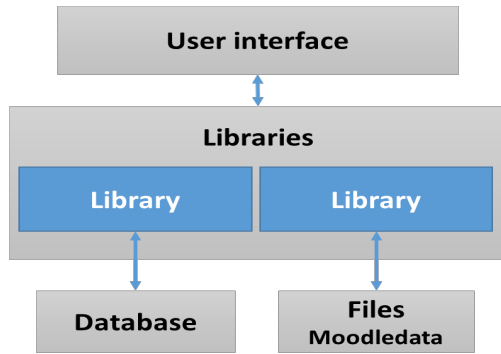
Figure 7 : Architecture de Moodle

Table 1: 1 moodle basic files

Files or Folders	Description of files or folders
version.php	It contains the following information: The declared version of the module; the minimum required version of Moodle for the plugin to work and cron settings (regular task).
index.html	Allows you to reach the list of instances of pedagogical objects in a class.
mod.html	Allows Moodle to reach the object editing screen.
icon.gif	the icon that appears in the class to symbolize the activity.
lib.php	Library functions where Moodle will look for various callbacks that provide the "standard" module information.
view.php	Allows you to make all the screens of the module.
Backup	Folder containing backup and restore information.

3.4. Moodle plugin development

A plugin or plug-in is a package that complements a host software to bring new features.

It can be positioned in a pedagogical sequence of a course. The following table shows some basic files of a moodle plugin.

3.5. Description of Kamailio

Kamailio is a very powerful open source SIP server, capable of handling thousands of calls per second without degradation of the quality of service. It can be used to create real-time VoIP communication platforms. It offers multiple features, including support for WebSocket for WebRTC, IPv4 and IPv6, TCP, UDP, TLS for secure communication, IMS extensions, and even an MSRP relay module discussed in this article. The MSRP relay module used in this work reuses the Kamailio core module to provide MSRP routing capabilities. A not insignificant advantage of this module is that it can access all other extensions of the Kamailio server. These are authentication, accounting, security and protection against denial of service attacks, etc. Kamailio has the ability to handle both SIP traffic and SIP traffic received on the same port. This is how Kamailio is used as a standalone MSRP relay where you can have both SIP and MSRP processing instances. You can also configure the module so that some ports will be for SIP only and others for MSRP.

4. Results

To prove the relevance of our solution, we deployed and tested it with 22 students and obtained the following results. Especially in this extended version, we were able to give access to a paid software to create architectures with several licenses. Thus, thanks to a single license, students can connect remotely and use the software to do their practical work, practice without the help of the teacher, etc.

4.1. Screen sharing

Figure 8 shows the invitation for screen sharing with the three options accept, reject, and occupy.
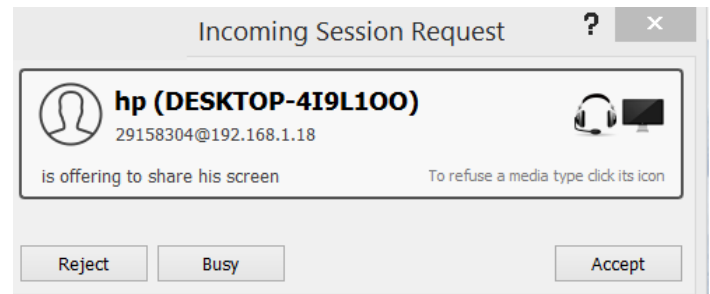
Figure 8: incoming session request

Figure 9 shows the sharing and remote control of a computer between trainer and learners. The teacher asks to take control of the learner's machine in order to show him how to handle.

Figure 9: Sharing and remote control of a computer

Figure 10 shows an illustration of the sharing and remote control of a workstation between actors in a digital university. This feature makes possible to pass audio video streams and even chat.
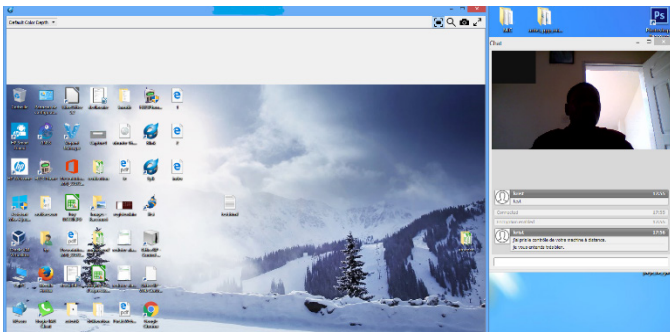
Figure 10 : remote sharing and control integrating audio and video

The importance of practical work in teaching in STEM is no longer demonstrated. An environment rich in practical work ensures better education for students. Relatively expensive software required for student training may not be accessible to them remotely. Thanks to our platform, students can access virtual machines containing these programs. Figure 11 shows how students can access these virtual machines through our platform. Thus, the student will be able to practice freely by choosing the virtual machine containing the software which he needs.
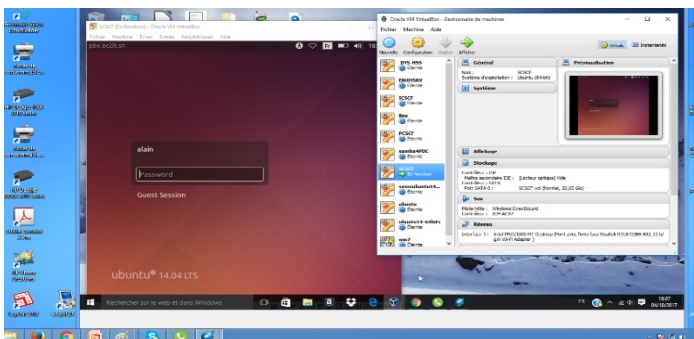
Figure 11: Remote access to a virtualization server for hands-on labs

Figure 12 shows a specific work environment for students learning computer programming. Thanks to a collaborative development environment from a web browser, students can work on the same project together. It may be to develop a project together using the eclipse development environment for example.
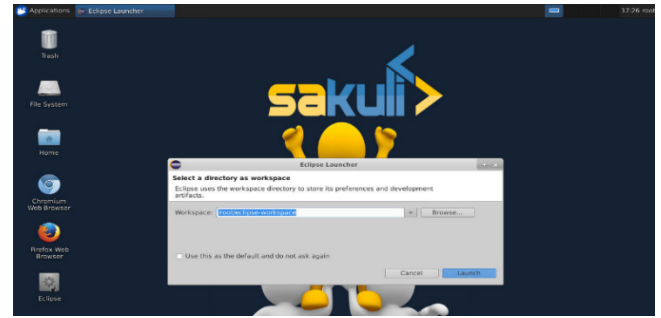
Figure 12: sharing development environment

In a classical classroom training, the teacher sometimes assists a student in difficulty. He can therefore show him step by step how to use this or that feature. This practice should also be possible in distance learning. This is what we demonstrate with our platform (Figure 13). A student's machine was remote-controlled by the teacher. The latter shows his student some features specific to a scientific computing software.
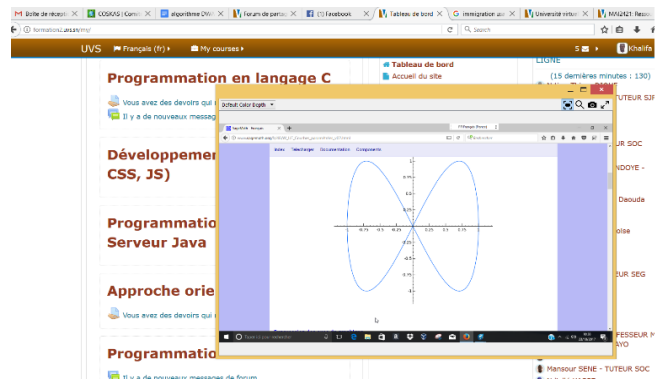
Figure 13: remote control of a learner's machine

Some programs are essential for learners as part of their training (development of architectures, mathematical simulation software, etc.). Sometimes, these tools are not free and not all students can acquire the license. To help the students, we provide them with a virtual machine on which we install the tools and buy the license. Thus, with a single license, during a session all students in the class will be able to use the software.

Figure 14 shows the E-DRAW software for example, used by computer science students to create diagrams, architectures, etc. Each student can therefore connect remotely, use the software to draw his architectures and save his work.
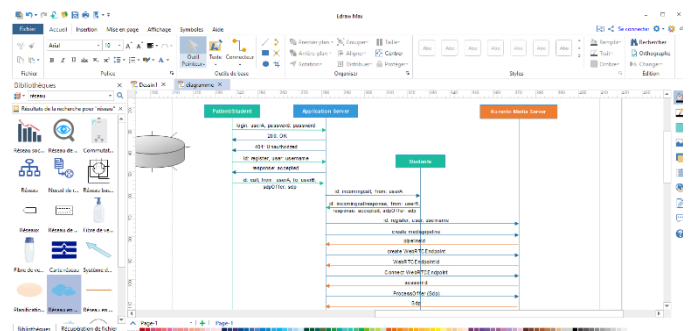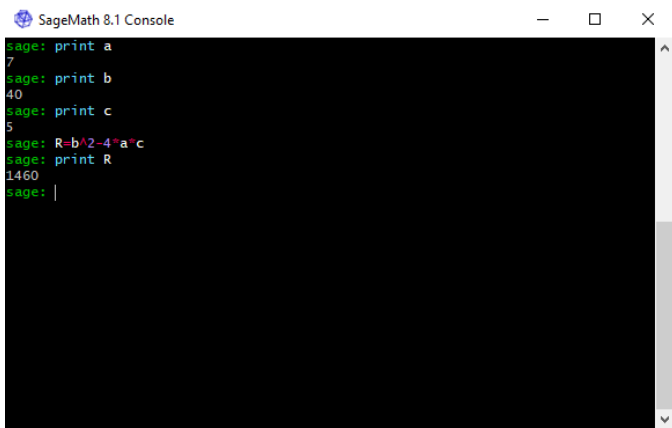
Figure 14: remote use of edraw

We also note that students at the beginning of their studies have difficulty installing certain software because of the limited resources of their computers. So we made SageMath software available to students to allow them to do their simulations remotely without worrying about the performance of their computer.

Figure 15: remote use of SageMath

4.2. Transfer of large files

This is about sending very heavy files simply by using drag and drop. It should be specified that the one to whom the file is intended has the ability to accept or reject the heavy file sent. If it chooses to accept the file, the download will start as you can see in Figure 16.

Figure 16 file receive 1

5. Conclusion and perspective

In this work, we proposed the integration of innovative tools in distance learning devices, based on the protocols WebSocket, MSRP, and the Kamailio media server, to allow the different actors of virtual universities to improve the quality of practical work in online labs. Students from the Virtual University of Senegal were able to use this solution effectively to collaborate in real time on practical work as if they were physically on the Moodle platform. It should be noted that the students of the Virtual University of Senegal are already used to the use of Moodle.It also allows the various actors of the training device to exchange large files thanks to the flexibility of the MSRP protocolThe platform has also been experimented with math students for online tutorials. The result was very satisfying. Two major features in this solution have been very useful for language teachers. This is audio communication and large file transfer in real time. It was helpful for them to truly appreciate the reading and pronunciation skills of their students.

The websocket msrp subprotocol also had a real impact in the implementation of the solution. Indeed, it made it easy to integrate our solution as a Moodle plugin. There was no question of changing the habits of students who are already accustomed to the Moodle GUI.

In our experiments, we used twenty-four students, which allowed us to easily manage the quality of service to users. We plan to use the Computing Cloud and a load-balancing device to manage the scalability, allowing the MOOCs to use our solution.

Acknowledgment

References

[1] K. Sylla, M. Seck, S. Ouya and G. Mendy, "Impact of MSRP protocol integration in e-learning platforms of universities," 2018 20th International Conference on Advanced Communication Technology (ICACT), Chuncheon-si Gangwon-do, Korea (South), 2018, pp. 1-1.

[2] A. Sharma, "A proposed e-learning system facilitating recommendation using content tagging and student learning styles," 2017 5th National Conference on E-Learning & E Learning Technologies (ELELTECH), Hyderabad, India, 2017, pp. 1-6.

[3] S.B. Nite ; M. Margaret ; R.M. Capraro ; J. Morgan; C.A. Peterson, "Science, technology, engineering and mathematics (STEM) education: A longitudinal examination of secondary school intervention,"Frontiers in Education Conference (FIE), 2014 IEEE , vol., no., pp.1,7, 22-25 Oct. 2014.

[4] J. P. López, A. Cerezo, J. M. Menéndez and J. P. Ballesteros, "Usage of mobile devices as collaborative tools for education and preparation of official exams," 2015 International Symposium on Consumer Electronics (ISCE), Madrid, 2015, pp. 1-2.

[5] C.C. Aydin ; G. Tirkes, "Open source learning management systems in elearning and Moodle," Education Engineering (EDUCON), 2010 IEEE , vol., no., pp.593,600, 14-16 April 2010

[6] S. Ouya, A. B. Mbacke, G. Mendy, P. W. Diouf and K. Sy, "Social network integration to e-learning environment," 2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA), Marrakech, 2015, pp. 1-4.

[7] S. Ouya, K. Sylla, P. M. D. Faye, M. Y. Sow and C. Lishou, "Impact of integrating WebRTC in universities' e-learning platforms," 2015 5th World Congress on Information and Communication Technologies (WICT), Marrakech, 2015, pp. 13-17.

[8] M. I. Santally, R. K. Sungkur, N. Sooltangos, P. Mounier and R. Mira-Bahadoor, "Flash to HTML5 adaptations: A case-study for UpToTen.com," 2016 International Conference on Advances in Computing and Communication Engineering (ICACCE), Durban, South Africa, 2016, pp. 375- 380.

[9] B. Campbell ; R. Mahy and C. Jennings, "The Message Session Relay Protocol (MSRP)", RFC 4975, DOI 10.17487/RFC4975, September 2007, <https://www.rfc-editor.org/info/rfc4975>.

[10] V. N. Gridin, G. D. Dmitrevich, V. I. Anisimov and S. A. Vasiliev, "Methods of constructing service-oriented computer — Aided circuit design systems based on WebSocket Protocol and Oracle database," 2017 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), St. Petersburg, 2017, pp. 409-412.

[11] Ned. FREED, RFC2045 multipurpose Internet mail extensions (MIME) part one: format of Internet message bodies. http://ds. internic. net/rfc/rfc2045. txt, 1996.

[12] C. Jennings ; R. Mahy and R. Adam. "Relay Extensions for the Message Sessions Relay Protocol (MSRP)." (2007).

[13] H. Yang, H. Lin, J. Li and W. Lei, "The Design and Implementation of an Enhanced SIP ALG to Support MSRP Sessions," 2009 Ninth International Conference on Hybrid Intelligent Systems, Shenyang, 2009, pp. 289-292.

[14] C. Jennings, R. Mahy and A. Roach. "Relay Extensions for the Message Sessions Relay Protocol (MSRP)." (2007).

[15] I. Fette and A. Melnikov, "The WebSocket Protocol", RFC 6455, DOI 10.17487/RFC6455, December 2011, <https://www.rfc-editor.org/info/rfc6455>.

[16] P. Dunkley ; G. Llewellyn ; V. Pascual ; G. Salgueiro and R. Ravindranath, "The WebSocket Protocol as a Transport for the Message Session Relay Protocol (MSRP)", RFC 7977, DOI 10.17487/RFC7977, September 2016, <https://www.rfc-editor.org/info/rfc7977>.

[17] VOZNAK, Miroslav and L. MACURA, Kamailio syntax generator and configuration file parser. In: Proceedings of the 15th WSEAS International Conference on Computers, WSEAS, Wisconsin, Estados Unidos. 2011. p. 308-312.

[18] H. M. Rissanen, T. Mecklin and M. Opsenica, "Design and Implementation of a RESTful IMS API," 2010 6th International Conference on Wireless and Mobile Communications, Valencia, 2010, pp. 86-91.

[19] G. Sykamiotis, A. Charitopoulos, M. Rangoussi and D. Koulouriotis, "Extraction and presentation of access and usage data from an e-learning platform (moodle): Design and development of a software application," 2017 IEEE Global Engineering Education Conference (EDUCON), Athens, 2017, pp. 1020-1026.

[20] M. Ivanova, G. Grosseck and C. Holotescu, "Researching data privacy models in eLearning," 2015 International Conference on Information Technology Based Higher Education and Training (ITHET), Lisbon, 2015, pp. 1-6.

Q-Learning versus SVM Study for Green Context-Aware Multimodal ITS Stations

Adel Mounir Said[1], Emad Abd-Elrahman[*,1], Hossam Afifi[2]

[1]*National Telecommunication Institute , Cairo, Zip Code 11768, Egypt.*

[2] *Telecom SudParis, IMT Saclay, Zip Code 91120, France.*

A R T I C L E I N F O

A B S T R A C T

Intelligent Transportation Systems (ITS) applications can take big advantage of Context Awareness approaches. Parameters such as user mobility, passengers comfort reaction and pollution emission levels (CO_2) can enrich such applications during the decision making phase. Moreover, the expanding in ITS services offers great opportunities for travelers to find the best route to reach their destinations with the lowest or fair costs. It can offer a selecting methodology for optimal route that adapted with some processing parameters like CO_2 level, ticket cost, waiting or connection times and the overall traveling time plus the comfortability reaction for each means of transportation) in real time environment using Machine Learning (ML) tools like Q-Learning or SVM: Support Vector Machines. This paper aims at conducting a comparison study for green ITS routes (i.e. the lowest CO_2 levels). The study compares between Q-Learning and SVM techniques for identifying different variety of routes between two stops as ranked routes from best to lowest based on some traces gathered from some known transportation traces. Reinforcement Q-Learning is applied to validate the first phase in our approach to recommend the best means and SVM is used to validate the prediction phase about the best route among different routes built based on three means of transportation (metro, train and bus).

1 Introduction

The ITS network is considered as the most important element of concern to the countries in the modern smart cities era. It is being the arteries that pass through continues chain of business activities, social, cultural and tourism that promote the march of the national economy for its role in providing services to the sectors of production and other services, and is reflected from providing employment opportunities. Therefore, the new ITS construction projects occupy the first initiative in sustainable development programs and projects to achieve higher rates of growth and development through the provision of passenger traffic between all stations or cities of the proposed new development. The governments directions are focused on planning and implementations of future stations that help in the preparations and developments of the countries through intelligent decision mechanisms.

Today, the need for an ITS solution becomes crucial to provide a smart solution for managing the traffic on country's roads network. Significant outcomes expected of using ITS are to provide a smooth, smart, and sustainable transport means.

Moreover, distributing the passengers over all the available transportation means (like buses, cars, trains, planes, and ships) can lead to some kind of balance between them and help in solving the problems of traffic jams. This aim can be achieved if there will be an implementation of intelligent systems having clear information about some states statistics like current traffic means available, roads status, passengers per day and pollution's emitted per trip. Those implemented systems can predict the long term situation of city or state transportation system behavior and recommend the future planning for new smart cities. Another important thing is the world climate changes due

*Corresponding Author: Emad Abd-Elrahman, Address: 5 Mahmoud El Miligui Street, 6th district - Nasr City, Cairo - Egypt, Phone: 0020-100720268 & Email: emad.abd_elrahman @ telecom-sudparis.eu

to transportation activities. Those changes are serious and can be affected by the huge carbon emission levels from different means of transportation and hence the dangerous consequences on the humanity life. This could be an incentive to optimize and reduce the CO_2 emission either from public or private transportation means that will lead the concept of Green-ITS.

Context Awareness (CA) approach can be used to enrich the ITS services as it can adapt the control system dynamics. This shall grasp the potential advantages, which influence the control strategies taken to manage the transportation system especially, in multimodal traffic management. The multimodal traffic station has different crossings of different transportation means like metro, train, bus, and etc.

This paper is an extension of the work originally presented in (ICECTA) conference [1].

Through the work, we improved the previous implementations about advanced ITS traffic flows management [2]. The aim of the previous paper was to choose among different transportation means the best one. We considered the recommendation of the best transport mean from the available three means of transportation in multimodal station with crossing of metro, bus and train. This recommendation is done based on some user and system parameters at any time (t) for any passenger decides to reach a specific destination using the previous means. We used a reward based Q-learning approach to choose the best transport means available in multimodal stations. In the second phase of this work, we propose using Support Vector Machines (SVM) for their reputation in classification accuracy or prediction speed comparing to the previous approach.

The context of SVM foundations had been developed in 1995 by Vapnik [3] and obtained high popularity in different applications due to some attractive and intelligent features, and the promising performance in empirical tests [4].

To conclude our work, the goal is to select among different routes available between two stations the best one for the destination. The selection is carried out by classifying some collected ITS traces about all the possible routes including combinations between different means of transportation at any time (t). Moreover, we considered the influence of new parameters that are missed in the previous work [2] like:

- The passenger waiting time at the starting station till the arrival of transportation mean.

- The passenger waiting time if there is a connection(s) between different or same transport means to reach specific destination.

- The measurement of CO_2 pollution from each trip.

Furthermore, the performance evaluation study in this work is done based on real traffic statistics that gathered in real time from *SNCF Transilien* [5] and *Vianavigo* [6] French sites. This data covered some working and week-end days during one month testing period from the French transportation network of Paris area.

The construction for paper is structured as follows: In Section 2, we introduce the related work done in ITS decision-makings analysis and management using Machine Learning (ML) tools. Then, Section 3 presents the Reinforcement Q-Learning model then the SVM for the context aware ITS multimodal stations. Afterward, the evaluations done for the proposed models and their results are compared with different selected use cases of working and week-ends days in Section 4. This is conducted through extensive simulations based on collected real-time statistics from the French *SNCF Transilien* site [5]. Finally, Section 5 concludes and highlights the perspectives of this work.

2 Work Background

An Intelligent Agent (IA) term in ITS domain has been defined as: the agent which receives precepts from the environment and based on that takes smart actions [7]. The IA is considered as the transportation station in our study. This agent has the capability to take decisions based on the gathered reactions from the passengers. Upon received those reactions, a context-aware learning system is used to optimize the taken decisions. Then, it proposes an accurate guides for all passengers crossing this IA (i.e. the multimodal station). Hereafter, we highlight some relevant insights that tackled the intelligent transportation prediction systems based on reported reactions.

The following literature is organized in two sections. The first one reviews some propositions based on using reinforcement and Q-learning techniques [8-14]. The second section reviews the propositions based on using SVM techniques in ITS. As it seems to our problem context, there is no related work uses SVM learning algorithm to detect the travel time of each transport mean and hence predict the best one at each time (t). As will be seen, the most of these works in this section is relating to detecting the travel time for bus and cars on highways based on SVM.

2.1 Q-Learning based Prediction Systems

Q-Learning is typically used in reward based reinforcement learning in different ITS domains. The technique of arrival time estimation proposed in [8] a prediction model. This model considered its decision based on common context for both vehicles and their drivers. The proposed solution considered an increasing in the anticipated time of arrival based on some observations in the traffic flow patterns.

Moreover, the pattern anticipates the vehicles arrival time from the history of the passed routes (i.e. the whole previous traveled routes). The main weak point of this model is the lack of consideration either in backup routes to be used as parallel routes or in emergency situations due to traffic jam. But, this proposal

is a good solution for ITS tracking systems.

The work in [9, 10] proposed the reinforcement learning model for the traffic-based control. This is for the case of classical pre-timed systems in signal control and the control agents are also independent. Through the model, the authors implemented the learning system for their control agent based on reactions' rewards as a closed-loop control agent. This agent will interact with the environment to converge with the control policy for learning and achieving an optimal mapping for the optimal required control action and the environments state.

Moreover, the authors had been considered in another work [11] the use of the reinforcement learning approach in a decentralized system. They proposed an adaptive real-time agent for traffic signal control system that ables to minimizing the vehicles total travel time. Based on accumulated reward, Q-Learning agent is used to learn the optimal mapping between the environments state in one side and the corresponding control action in the other side.

The Q-learning model for traffic control considered in [11] is based on time varying and stochastic problems in the traffic flow while in [12], authors investigated the snag of agent-based self-optimization against multiple polices. They proposed the use of distributed w-learning reinforcement learning model. The purpose behind this idea applied in self-organized traffic control systems, is the performance improvement from multiple policies deployments simultaneously.

But, the work in [13] proposed a prediction model for the flow of traffic depending on the multi-agent reinforcement learning. This model is based on multi-cross roads control for the traffic signal.

Finally, authors in [14] concentrated on the coordination between control agents in order to adapt the roads traffic signals. Therefore, they propose to use an independent control agent mode besides an integrated mode merged with reinforcement learning. This model is capable of solving the agents' communication problems. Their results indicate that there is a delay reduction for the connection travel time according to their model proposition.

2.2 Classification Systems based SVM

SVM as a supervised machine learning tool that is based on simple principles, originated from statistical learning theory [3]. The SVM simplicity is coming from the applying of simple linear methods on the data [15].

The accuracy as well as simplicity of the SVM models are the key for many contributions based on classification and regression.

In this part, we present the most research directions that used SVMs in ITS domain as follows:

The first main concern in this direction using SVMs was focusing on the time prediction for the trip. The authors in [16-18] introduced different models based on SVMs to predict the time of each travel on the highway routes. While in [19], the authors proposed their travel time prediction model using the SVM in urban transport networks.

Other contributions considered SVM for the bus travel times prediction. In [20], SVM based hybrid model was introduced for this time prediction. Their technique was based on combining between the classical SVM, the Grubbs test method and an adaptive algorithm.

In [21], a new bus travel time prediction model for multiple dynamic routes is proposed. The model used SVM with Kalman filtering technique.

The authors in [22] proposed their time prediction model for the flow of rail transit passengers in the city of Beijing using SVR to improve the rail transportation management. They used the PLSR: Partial Least Squares Regression method to resolve multi co-linearity among the dependent variables.

Last direction in our study focused on the research directions in the traffic signs detection or prediction. In [23-26], the authors developed automatic systems for either traffic sign detection or recognition using SVM. Through [26], the authors proposed the use of SVM as a recognize module to analysis the keep-clear signs to enhance the traffic management in cities.

Another work [27] was focusing on using SVM for automatic video based vehicle detection. This is for an automatic traffic surveillance system based on ITS.

In [28], the authors proposed a design for a long-haul bus. This device is used for driving safety warning based on SVM. It is based on classifying the lane departure and forward collision.

To the best of our knowledge, SVM is not used in any ITS based model to nominate the best traffic mean or route for passengers to reach specific destination at time (t) under the context of multi-modal traffic. In the following section, the proposed model will be described and the both learning algorithms as well.

3 ITS Proposed Models

3.1 Green Transport Decision System Model (GTDS)

Usually, passenger(s) arrive to the Transport Station (TS) for reaching certain destination. Therefore, they have to decide taking a Route i (Ri) which may includes one or more transport mean at time (t). Here, a model for Green Transport Decision System (GTDS) is proposed for facilitating the passengers travels. GTDS model nominates the best Ri for passenger(s) based on Q-learning. The system nominations rely on passengers reward reactions. In contrasting SVM with Q-learning, SVM is faster and has a higher accuracy. Therefore, in this paper SVM is proposed to be used in the second phase of this work. Figure 1 demonstrated the model main inputs and the collected reactions as system feedbacks.
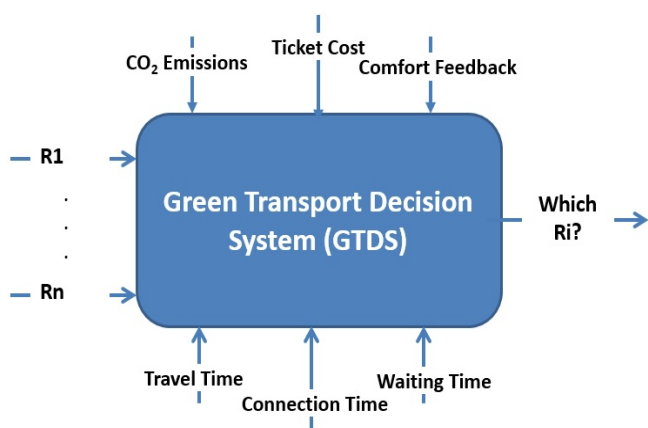
Figure 1: The main inputs and feedbacks for context-aware GTDS framework

to take a transportation mean.

Figure 2: Arrival times respect ratio for the main three transportation means, the metro comes in the first rank followed by local train and then bus as worst respect ratio to arrival on time

Stations apply the GTDS model should consider having multiple routes to the same destination. This is to provide different nominations at different situations aspects. According to that, the proposed model is applied on the departure station called: (*Gare De Paris Montparnasse*) and the arrival station called: (*Gare De Paris Nord*) as a study use case. The model is simulated by different week days activities (i.e. both weekend days and work days). The studied period consists of 25 quarters starting at 7:00 am until 1:00 pm. This period is intended to contain all the expected traffic conditions; either the rush hours or the free from traffic jam situations.

The choice of the best (*Ri*) is affected by many parameters, the most relevant are determined in the proposed model and illustrated in Fig. 1. The first factor is the passengers travel duration to reach the trend-setting. This factor will be considered taking into account the respect time for each means of transportation shown in Figure 2. The curves demonstrated in this figure have been obtained through the history of each transportation mean by collecting several observations. As indicated, the use of metro is more frequent and comfort than other transport means. The train achieves rank two and finally the bus. For this parameter which reflects respecting the arrival time, we are interested on some statistics which provide a wide sight for how the transport means follow or not the scheduled times arrival. This insight comes from the published SNCF statistics or real time scheduling plan for the transport means on the on-line site [5].

As clear in Figure 2, the best of the transport means in respecting the arrival times is the Metro. Even it may exceed its scheduled arrival times due to the rush hours. Therefore, when a passenger comes to the station and check in to see the best transportation tool to go to his destination *D*, the decision system inside the station uses the reactions from the station *D* about previous respecting arrival schedules to update the immediate reward function *r* when a passenger tries
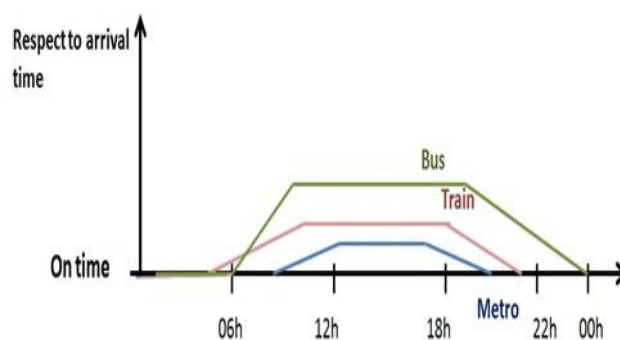
By doing so, the decision system firstly tries to optimize its future decisions biased from the previous experience along the day.

Secondly, the travel cost which could affect the system decisions. As known, the most of passengers propose the use of the cheapest transport mean.

Thirdly, if there is a waiting time before catching the transport mean.

Fourthly, in the travels of multiple transport means, the connection time is considered as an effecting parameter.

Fifthly, the comfortability coefficient for the use of *Ri* at any time (*t*). This coefficient reflects the passenger satisfaction degree of current experience about using any route *Ri*. The impact is influenced by the available free chairs in rush hours comparing to the number of passengers intend to use this route. Therefore, passengers are more interested of using the more comfortable transport means. The comfortability reaction of the passengers is shown in Figure 3. The mentioned parameters are function of time even the ticket price. In some countries, ticket cost plan changes along the week days (i.e. working days and off days). At present, the major of the world countries became concerned about the CO_2 emission because of the catastrophic effects on the world climate.

In this contribution, we are interested in adding the CO_2 parameter that affects the selection of the transport means and hence the best route. In France, the government encourages the people to use the public transport means to decrease the emission of the CO_2 and hence the air pollution. Moreover, France, during high air pollution days, encourages the use of public traffic means for free, and prevents the even/odd cars plate numbers in Paris area.
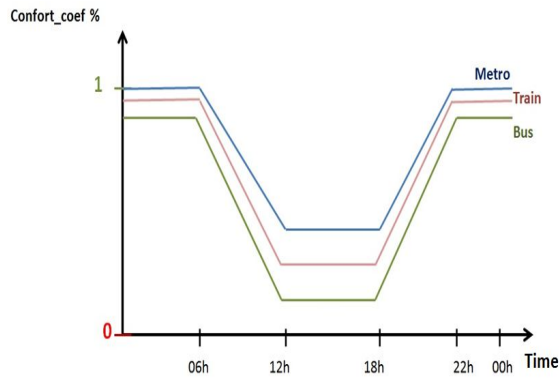
Figure 3: Passengers Comfortability Feedbacks/Reactions, comfort coefficient equal one (comfort_coef=1) represents high satisfaction from the passengers while the low level equal zero for unsatisfied trips

3.2 Reinforcement Learning Model

Using reinforcement learning, any TS can learn from its experience to guide passengers and thus, optimize its interactions for the future decisions. The mentioned experience is stated as rewards and sanctions from the prior taken decisions. This model is initialized at each TS benefiting from the collected passengers reactions. When a passenger uses Ri to reach a destination, this Ri will be rewarded or sanctioned according to these parameters: the travel time for reaching certain destination, the passengers comfortability reaction, and finally the travel cost of every Ri. Consequently, each TS will tune its travel means nominations from its gained experience.

In this paper, the TS decision problem is formulated as a Markov Decision Problem (MDP), that can be solved by the reinforcement learning algorithm. The TS will elects the best Ri for reaching a destination biased by its experience at each day hour, or another duration period based on the traffic. Next to that, the system evaluates itself by assessing the decision taken as a reward or sanction. The MDP is defined as a tuple $[i, t, a, r]$.

- i: stands for the object that is considered in this study by the route Ri.

- t: stands for the system states set that are considered in this model by the under-study time shift periods during the day.

- a: stands for the decisions set that taken by the TS. In this work, the decisions of the TS are to nominate the best Ri that can cover the passengers requirements/expectations. Only six possible Rs between the proposed TSs are considered in our use case.

- r: stands for the immediate reward that the passenger will get according to taking decision a.

A $Q_i(t, a)$ is an action-state matrix that is formed to store the gathered reward/sanction for each state

and the action pair of a route Ri at a time (t). Giving an example to that, this matrix reflects the predicted reward when takes an action (a) after using the route Ri at a time (t). The updating function of $Q_i(t, a)$ can be defined by:

$$Q_i(t, a) = (1 - \alpha) * Q_i(t, a) + (1 - \alpha) * (r_i(t) +$$

$$\gamma * max_a' Q_i(t + 1, a') \qquad (1)$$

Where:

- α: indicates the learning rate representing how fast the Q-values vary corresponding to the dynamic users reaction.

- γ: indicates the discount factor. It reflects if the immediate reward status against the future one. For the comparison, high value future reward is more valued than the immediate one.

- r_i: indicates the predicted immediate reward when choosing Ri at a time (t).

- $max_a' Q_i(t + 1, a')$: represents the maximum predicted future reward once the system reaches a state ($t + 1$) when taking any decision action a.

Furthermore, once a passenger arrives the desired destination via Ri after the scheduled arrival-time, TS will get a sanction reaction. While, in case all the Rs are delayed, bonus is counted to the Rs of the minimum counted delay time.

$$r_i(t) = \frac{mt}{B_1(t)} + (\frac{mc}{B_2}) + B_3(t) + B_4(t)$$

$$+(B_5(t) * Comfortability) + B_6(t) + B_7 \qquad (2)$$

Where:

- mt is the reference trip duration of all Ri at a time (t) that represents the minimum travel time of Ri.

- B_1 parameter indicates the travel time due to catching Ri.

- B_2 parameter indicates the ticket price of the Ri comparative to a mean value (mc) that equals to one. It is assumed that, the ticket price plan is fixed during the week-days.

- B_3 and B_4 are bonus values for the Ri of minimum waiting and connection times respectively. These parameters are comparative between all the transport means and takes a value of 1 to the lowermost times up to 0.5 to the uppermost times.

- B_5 is the passengers satisfaction reaction in percentage reflecting the comfortability of using Ri at time (t). The default value equals to 1 as shown in Figure 3.

- B_6 is an added bonus to the Rs of low delays comparatively between them and on the contrary Ri of a highest delay has no bonus.

- B_7 is an indication to the green transport means which is preferred to be used than others as it cause less pollution. This parameter is a bonus totaled to the Ri of low transport means CO_2 emission.

Passengers reactions/feedbacks about the transport means are collected in real-time as reports. Each passenger is supposed to send his reaction/feedback using a dedicated simple mobile application that will supply the Decision System Algorithm $DSA(t)$ shown in Algorithm 1.

Algorithm 1 Q-Learning Algorithm: DSA(t)

1: **Input:** For each Ri, collect the passengers' feedback at their arrival destination
2: **Input:** $L1, \cdots$, until $L6$ are 6 lists initialized by NULL
3: **Input:** Qi is initialized based on the previous knowledge of the traffic jam for every day and is rewarded based on the current time (t) by time interval (τ)
4: **Input:** Ni is the available Rs at the current time (t)
5: **Output:** The best route at time ($t + \tau$)
6: **if** $Ni \neq \emptyset$ **then**
7: **For**$\{j, k, l, m, n, f\} \in Ni$ **do**
8:   {at time (t) there is a comparison between the available Rs as follows:}
9:   **if** $|RouteAvailable(j)| \neq 0$ and reward(j) > reward(i) for ($j \neq i$) **then**
10:      $L1 \leftarrow j$
11: **end if**
12: **if** $|RouteAvailable(k)| \neq 0$ and reward(j) > reward(k) > reward(i) for ($k \neq i$) and ($j \neq i$) **then**
13:      $L2 \leftarrow k$
14: **end if**
15: **if** $|RouteAvailable(l)| \neq 0$ and reward(j) > reward(k) > reward(l) for ($k \neq i$) and ($l \neq i$) **then**
16:      $L3 \leftarrow l$
17: **end if**
18: **if** $|RouteAvailable(m)| \neq 0$ and reward(l) > reward(m) > reward(i) for ($l \neq i$) and ($m \neq i$) **then**
19:      $L4 \leftarrow m$
20: **end if**
21: **if** $|RouteAvailable(n)| \neq 0$ and reward(m) > reward(n) > reward(i) for ($m \neq i$) and ($n \neq i$) **then**
22:      $L5 \leftarrow n$
23: **end if**
24: **if** $|RouteAvailable(f)| \neq 0$ and reward(f) < reward(i) for ($k \neq i$) for $f = j, k, l, m, n$ and ($f \neq i$) **then**
25:      $L6 \leftarrow f$
26: **end if**
27: **End for**
28: **if** $L1 \neq 0$ **then**
29:      The next Route is chosen from $L1$ of the largest Q-value
30: **else if** $L1 = 0$ and $L2 \neq 0$ **then**
31:      The next Route is chosen from $L2$ of the largest Q-value
32: **else if** $L1 = L2 = 0$ and $L3 \neq 0$ **then**
33:      The next Route is chosen from $L3$ of the largest Q-value
34: **else if** $L1 = L2 = L3 = 0$ and $L4 \neq 0$ **then**
35:      The next Route is chosen from $L4$ of the largest Q-value
36: **else if** $L1 = L2 = L3 = L4 = 0$ and $L5 \neq 0$ **then**
37:      The next Route is chosen from $L5$ of the largest Q-value
38: **else if** $L1 = L2 = L3 = L4 = L5 = 0$ and $L6 \neq 0$ **then**
39:      The next Route is chosen from $L6$ of the largest Q-value
40: **else**
41:      Ni is empty
42:      At time (t) a negative reward is generated for that Ri
43: **end if**
44: **end if**
45: At time ($t + 1$) the reward for every Ri is computed after the interval (τ) based on step (2)
46: At time (t) the Q-value is updated based on Eq. (1)

3.3 Support Vector Machine Classification

SVM, a supervised machine learning tool, is used for classifying linear and non-linear dynamic systems. SVM decisions can be considered for the classification problems and the regression aspect. The best classification results from finding the best hyperplane fits data separation with error free and maximal distance between the closest vector to that hyperplane.

In this work, SVM is used to distinguish among two classes. Class one refers to the best route(s) to certain destination while the other class refers to the alternative routes that are available for use to the same destination.

The aim is to find a data separator between the two mentioned classes whatever linear or non-linear. From theoretical point of view, linear SVM matches our scenario as it is a question of only two classes. Assuming the routes data set (i.e. the collected traces) is D as:

$$D = \{(x^1, y^1), \cdots, (x^n, y^n)\}, x \in R^n, y \in (0, 1) \quad (3)$$

Where:

- x is the set of training vectors.

- n is the number of days quarters under study that is equal to 25 period.

- y is the classs labels.

And the hyperplane is:

$$\langle w, x \rangle + b = 0 \quad (4)$$

Considering a canonical hyperplane [4], where the parameters w, b are constrained by,

$$min_i|\langle w, x^i \rangle + b| = 1 \quad (5)$$

The following constraints must be considered for a separating the canonical hyperplane, which is the most fitting assumption for non-linear classification,

$$y^i[\langle w, x^i \rangle + b] \geq 1, i = 1, \cdots, l \quad (6)$$

The distance $d(w, b; x)$ of a point x from the hyperplane (w, b) is:

$$d(w, b; x) = \frac{|\langle w, x^i \rangle + b|}{\|w\|} \quad (7)$$

For optimal hyperplane, the margin, ρ has to be maximized based on the constraints of (6) and given by:

$$\rho(w, b) = min_{x^i, y^i = -1} d(w, b; x^i) + min_{x^i, y^i = 1} d(w, b; x^i)$$

$$= \frac{2}{\|w\|} \quad (8)$$

So, the optimal hyperplane, the data separator that minimizes is:

$$\phi(w) = \frac{1}{2}\|w\|^2 \quad (9)$$

The optimization problem solution of (9) using the constraints of (6) is given by the saddle point of the Lagrange function:

$$\phi(w, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{l} \alpha_i(y^i[\langle w, x^i\rangle + b] - 1) \quad (10)$$

where α is the Lagrange multiplier that must be minimized with respect to w, b and maximized with respect to $\alpha \geq 0$. This dual problem is given by:

$$max_\alpha W(\alpha) = max_\alpha(min_{w,b}\phi(w, b, \alpha)) \quad (11)$$

To get the minimum w and b of the Lagrangian, ϕ is given by,

$$\frac{\partial \phi}{\partial b} = 0 \Rightarrow \sum_{i=1}^{l} \alpha_i y_i = 0$$

$$\frac{\partial \phi}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{l} \alpha_i y_i x_i \quad (12)$$

So, from (10), (11), and (12), the dual problem expressed by:

$$max_\alpha W(\alpha) = max_\alpha - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j$$

$$\langle x_i, x_j\rangle + \sum_{k=1}^{l} \alpha_k \quad (13)$$

and the solution of this problem will be given as follows:

$$\alpha^* = arg\ min_\alpha \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j \langle x_i, x_j\rangle - \sum_{k=1}^{l} \alpha_k \quad (14)$$

where $\alpha_i \geq 0$, for $i = 1, \cdots, l$.

By solving (14) according to its constraints defines the Lagrange multipliers then, obtains the optimal separating hyperplane as well as follows:

$$w^* = \sum_{i=1}^{l} \alpha_i y_i x_i$$

$$b^* = -\frac{1}{2}\langle w^*, x_r + x_s\rangle \quad (15)$$

where x_r and x_s are the support vectors from each class satisfying:

$$\alpha_r, \alpha_s > 0, \quad y_r = 0, y_s = 1 \quad (16)$$

and the hard classifier is:

$$f(x) = sgn(\langle w^*, x\rangle + b) \quad (17)$$

The input vectors classification that refer to the routes parameters can be solved as a nonlinear problem that fits our case as well. This depends on the input vectors distribution. For a general form in high dimensional feature space, lets suppose SVM maps the input vector x into a high dimensional feature space, z. So, the optimization problem becomes:

$$\alpha^* = arg\ min_\alpha \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{k=1}^{l} \alpha_k \quad (18)$$

where K is the kernel function for non-linear mapping into feature space with the same limitations, which are:

$$0 \leq \alpha_i \leq C, i = 1, \cdots, l \sum_{j=1}^{l} \alpha_j y_j = 0 \quad (19)$$

where C is a regularization parameter of the kernel function and represents the upper bound on the permissible values. Solving (18) with its constraints in (19), resulting the Lagrange multipliers, and a hard classifier as follows:

$$f(x) = sgn(\sum_{i \in SV_s} \alpha_i y_i K(x_i, x) + b) \quad (20)$$

where:

$$\langle w^*, x\rangle = \sum_{i=1}^{l} \alpha_i y_i K(x_i, x)$$

$$b^* = -\frac{1}{2}\sum_{i=1}^{l} \alpha_i y_i [K(x_i, x) + K(x_i, x_r)] \quad (21)$$

4 Performance Validation

This section demonstrates the proposed GTDS model performance evaluations. It is evaluated based on using Q-Learning and SVM learning methods. *Matlab 2012* tool is used for the simulation part with the following conditions:

- The model use case under study considers Paris region in France. The departure station is (*Gare De Paris Montparnasse*) and the arrival destination station is the (*Gare De Paris Nord*) as the most two famous stations in this Parisian region.

- The routes *Rs* available between the source and destination stations are many but, we processed and filtered them to form six routes only of closed results for simple comparisons.

- Time shift under study is divided into two sets. Set 1 is representing the two days of the weekend (i.e. Saturday and Sunday) and set 2 is representing the two selected working days which are Monday and Friday.

- Time shift under study starts at 7:00 AM till 13:00 PM

- The ticket price of *Ri* is supposed to be unity and its value is redeemed comparatively based on the transport mean used. The route cost is supposed to be the same as this fits the case in France that depends on dividing the destinations to zones.

- The collected traces of travel duration, waiting and connection times are gathered from the real-time schedules of the *SNCF Transilien* official site [5].

- The passengers comfortability reactions/feedbacks of the transport means under-study follows the chart in Figure 3. This comprises many parameters such as the capacity, the design satisfaction, the rate of flow (periodicity), and the ease of catching each of them.

- The actual CO_2 emission is measured in gram per person of each Ri and gathered from *Vianavigo* site [6].

- Number of stops: It is dynamic parameter. Its value depends on the route combinations that may comprise a combination of different transport means.

The simulations are categorized in two parts; part one is concerned about the SVM learning algorithm, and part two focuses on the Q-learning algorithm. In part one, the used traces of the determined source and destination are gathered for two weeks just for a proof of concept from [5]. We considered the first week traces as the training set and the second week traces as the testing set.

After the validation of SVM, the obtained testing set results are shown in Figure 4 and Figure 5. They illustrate the SVM accuracy calculated for both the week-end and working days respectively.

As illustrated in these figures (Figure 4 and Figure 5), the SVM classification accuracy is looking good in week-end days but has less accuracy in the working days. These are anticipated results as working days may have rush hours that influence on choosing the best route. Despite the results disorder, the system performs steady as the error is so small that is closed to one over six.

Figure 5: The output SVM system testing accuracy during working tested days, for the X-axis is represented the time shift from 7:00 AM to 13:00 PM as testing period

The second simulation part is concerned to the Q-learning machine results of choosing the best route Ri. The same traces and system settings are used for the Q-learning simulation. This simulation considers two use cases as follows:

A) The week-end days: it concerns to the delay of using each transport mean on one of the week-end days (Saturday) as shown in Figure 6. The use of bus compared to Metro and Train has the largest delay at the same day quarter although roads are not easily suffering of congestion during the week-ends. Moreover, the metros are not far from buses statistics due to their limited circulations in the weekend days.

B) Working-days: this simulation concerns to one working day like Friday traffic situation. The highest rush hours of the day are simulated. Figure 7 shows each transport mean delay. As clear in the figure, the use of the bus causes the largest delay with respect to Metro and Train at the same day quarter due to the traffic jams on normal routes.

Figure 4: The output SVM system testing accuracy during week-end tested days, for the X-axis is represented the time shift from 7:00 AM to 13:00 PM as testing period
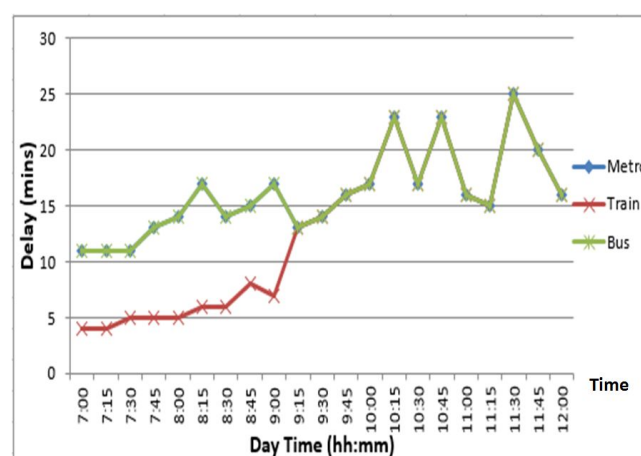
Figure 6: Saturday travel delay due to using either Metro, Train, or Bus. The hh: is the hour while mm: represents the minutes part
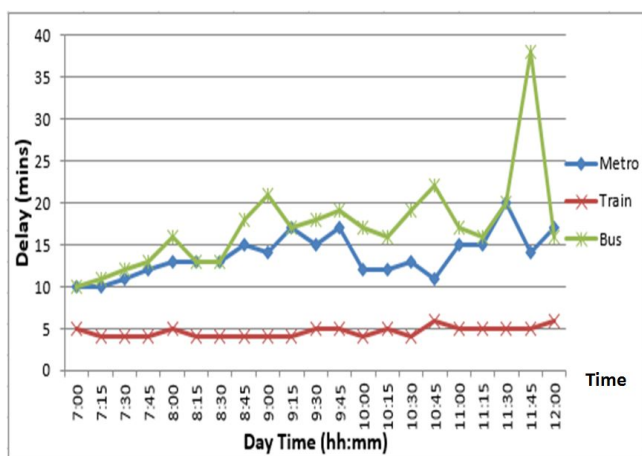
Figure 7: Friday travel delay due to using either Metro, Train, or Bus. The hh: is the hour while mm: represents the minutes part
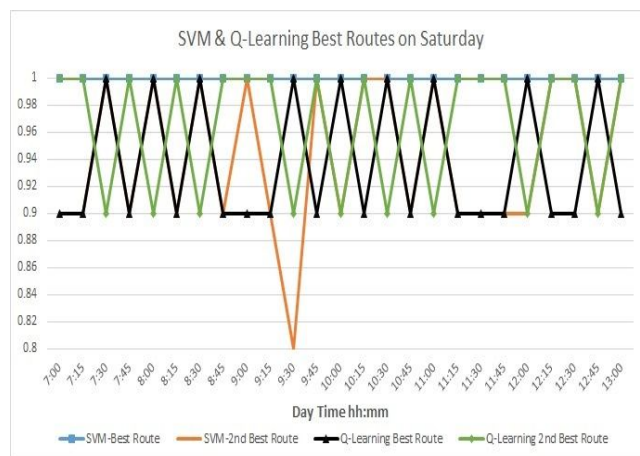
Figure 10: SVM versus Q-Learning algorithms results for the week-end selected day as Saturday
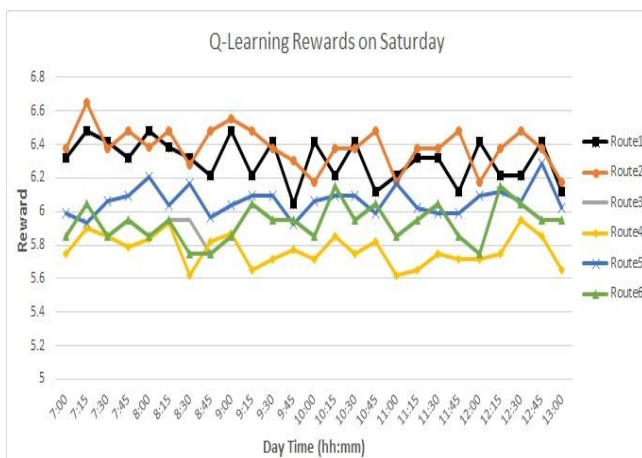
Figure 8: The rewards for Q-Learning algorithm based Saturday testing set. The hh: is the hour while mm: represents the minutes part
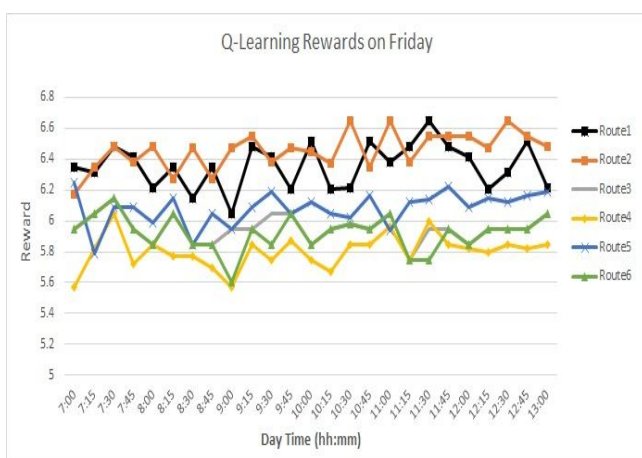
The results shown in Figure 8 and Figure 9 demonstrate the Q-values which stands for the rewards of the optional routes of each day quarter on Saturday and Friday respectively. As clear in the figures, the best route decision changes with respect to the day quarter time and routes conditions. These routes are composed of different two or more traffic means. Therefore, these figures illustrate that, the recommended routes to be used along the studied day hours depends on the conditions and delays for each route which is composed of different connections between the three-transport means (metro, bus and train).

Figure 10 and Figure 11 show the best two routes suggested by the two proposed machine learning systems (i.e. SVM and Q-Learning) for Saturday and Friday respectively. As appear in these figures, each learning algorithm decision is various due to the difference of the learning technique of both algorithms. Also, due to the difference of the input parameters weight. And finally, because SVM is a supervised learning model but, Q-learning is based on reinforcement learning.

The main purpose from these two figures is to clarify the criteria of each algorithm output and shows that each algorithm may fits according to certain applications. If the ITS system is merely in stable with no fluctuates, SVM would be a perfect fit for these scenarios. In contrast, Q-learning could suitable for infrequent cases such as accidents. This is because of Q-learning depends on conditions of the current collected data in addition to its decision history of the previous day while SVM depends on the training sets.

5 Conclusions

The framework of route selection in the ITS multimodal stations is investigated in this paper using context awareness decision model based on machine learning tools. Through this work, we used two different models; the Q-Learning as a reinforcement model and the SVM as a supervised model.

Figure 9: The rewards for Q-Learning algorithm based Friday testing set. The hh: is the hour while mm: represents the minutes part
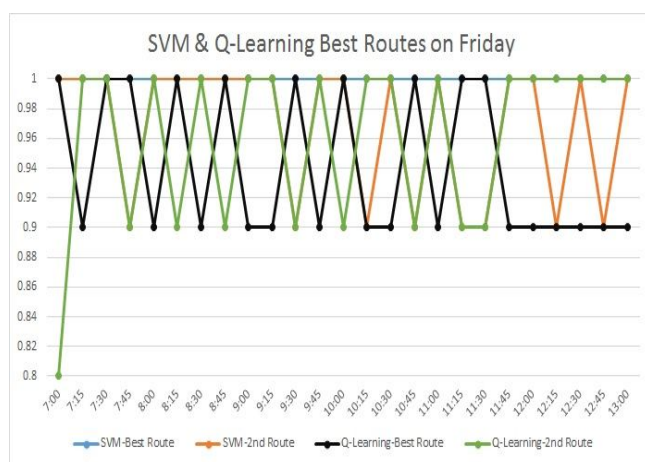
Figure 11: SVM versus Q-Learning algorithms results for the selected work-day as Friday

In the first phase, we built the Q-Learning to select among six studied routes the best one. After that, we used SVM model in order to first classify the proposed routes, second to calculate the model's accuracy, and third to predict the best solution for the upcoming routes. The obtained results from the simulation part of both approaches indicated the advantages and disadvantages for each technique and when (time during the day or day during the week) or where (according traffic status) we can use them in ITS stations.

Actually, Q-Learning approach is best choice that can be used when the ITS system is not stable and subject to many disturbances while, SVM model is better used when the ITS system is stable and the variations over the time in data sets are steady. That is why we recommend using hybrid solutions to face such cases when searching the optimal classification based on different time intervals and with green parameters. On the positive side, ITS will enable the Green Transport Decision System (GTDS) to decide based on context-sensitive information the best route, to improve the efficiency of multimodal stations usage and to reduce the environmental impacts by including the CO_2 effects.

In the future work, it will be more realistic to focus on three dimensional model approach that will consider in addition to user-location; the time. This spatio-temporal model can be applied in different universal applications. But, the problem is the difficulty of having such kind of statistics as it is subject to country regulations related to security and privacy issues. However, it can provide a new generation of traffic flows modeling that will add dynamic, predicative and adaptive control for multimodal traffic.

Acknowledgment

References

[1] A. M. Said, E. Abd-Elrahman and H. Afifi, "A comparative study on machine learning algorithms for green context-aware intelligent transportation systems," 2017 International Conference on Electrical and Computing Technologies and Applications (ICECTA), Ras Al Khaimah, 2017, pp. 1-5.

[2] A. M. Said, A. Soua, E. Abd-Elrahman, and H. Afifi, "Context-Aware Multi-modal Traffic Management in ITS: A Q-Learning based Algorithm," International Wireless Communications and Mobile Computing Conference (IWCMC 2015), Dubrovnik-Croatia, Aug. 2015, pp. 674-679.

[3] V. N. Vapnik, "The Nature of Statistical Learning Theory," New York: Springer, 1995.

[4] S. R. Gunn, "Support Vector Machines for Classification and Regression," Technical report:Faculty of Engineering, Science and Mathematics School of Electronics and Computer Science, 1998.

[5] SNCF Transilien 2018. [Online]. Available: https://www.transilien.com/en. [Accessed: 29-Sep-2018].

[6] Vianavigo 2018. [Online]. Available: https://www.vianavigo.com/en/home. [Accessed: 29-Sep-2018].

[7] S.J. Russel and P. Norvig, "Artificial Intelligence: A Modern Approach," Third Edition, Prentice Hall, 2010.

[8] D. Fagan and R. Meier, "Intelligent time of arrival estimation," IEEE Forum on Integrated and Sustainable Transportation System (FISTS), Vienna, 2011, pp.60 - 66.

[9] S. El-Tantawy and B. Abdulhai, "Multi-Agent Reinforcement Learning for Integrated Network of Adaptive Traffic Signal Controllers (MARLIN-ATSC)," 15th International IEEE Conference on Intelligent Transportation Systems (ITSC), Anchorage, 2012, pp. 319-326.

[10] S. El-Tantawy, B. Abdulhai, and H. Abdelgawad, "Multiagent Reinforcement Learning for Integrated Network of Adaptive Traffic Signal Controllers (MARLIN-ATSC): Methodology and Large-Scale Application on Downtown Toronto," IEEE Transactions on Intelligent Transportation Systems, Vol.14, No. 3, 2013, pp. 1140-1150.

[11] S. El-Tantawy, and and B. Abdulhai, "An agent-based learning towards decentralized and coordinated traffic signal control," 13th International IEEE Conference on Intelligent Transportation Systems (ITSC), Funchal, 2010, pp. 665-670.

[12] I. Dusparic and V. Cahill, "Distributed W-Learning: Multi-Policy Optimization in Self-Organizing Systems," Third IEEE International Conference on Self-Adaptive and Self-Organizing Systems SASO '09, San Francisco, 2009, pp. 20-29.

[13] J. Song and Z. Jin, "Q-learning based multi-intersection traffic signal control model," International Conference on System Science, Engineering Design and Manufacturing Informatization (ICSEM), Vol.2, 2011, pp. 280-283.

[14] J.Song, Z. Jin, and W. Zhu, "Implementing traffic signal optimal control by multiagent reinforcement learning," International Conference on Computer Science and Network Technology (ICCSNT), Vol. 4, Harbin, 2011, pp.2578-2582.

[15] A. Karatzoglou, D. Meyer, and K. Hornik, "Support Vector Machines in R," Journal of Statistical Software, Volume 15, Issue 9, 2006, pp.1-28.

[16] C. Wu, J. Ho, and D. T. Lee, "Travel-Time Prediction With Support Vector Regression," IEEE Transactions on Intelligent Transportation Systems, Vol. 5, No. 4, 2004, pp. 276-281.

[17] C. Wu, C. Wei, D. Su, M. Chang, and J. Ho, "Travel Time Prediction with Support Vector Regression," IEEE Intelligent Transportation Systems, Vol.2, 2003, pp. 1438-1442.

---

[1] Programme Imhotep (2016-2017): http://www.campusfrance.org/fr/imhotep

[18] A. Yusuf, and V. Madisetti, "Configuration for Predicting Travel-Time Using Wavelet Packets and Support Vector Regression," Journal of Transportation Technologies, Vol.3, No. 3, 2013, pp. 220-231.

[19] Y. Chen, H. J. van Zuylen, and Y. Qipeng, "Travel Time Prediction on Urban Networks Based on Combining Rough Set with Support Vector Machine," International Conference on Logistics Systems and Intelligent Management, Vol. 1, 2010, pp. 586-589.

[20] S. Zhong, J. Hu, S. Ke, X. Wang, J. Zhao, and B. Yao, "A Hybrid Model based on Support Vector Machine for Bus Travel-Time Prediction," PROMET Traffic and Transportation Journal, Vol. 27, No. 4, 2015, pp. 291-300.

[21] C. Bai, Z. Peng, Q. Lu, and J. Sun, "Dynamic Bus Travel Time Prediction Models on Road with Multiple Bus Routes," Computational Intelligence and Neuroscience (Hindawi Publishing Corporation), Vol.2015, No.432389, 2015.

[22] H. Zhou, Y. Qin, and Y. Li, "A Partial Least Square Based Support Vector Regression Rail Transit Passenger Flow Prediction Method," International Journal of u- and e- Service, Science and Technology, Vol.7, No.2, 2014, pp. 101-112.

[23] S. Maldonado-Bascn, S. Lafuente-Arroyo, P. Gil-Jimnez, H. Gmez-Moreno, and F. Lpez-Ferreras, "Road-Sign Detection and Recognition Based on Support Vector Machines," IEEE Transactions on Intelligent Transportation Systems, Vol.8, No.2, 2007, pp. 264-278.

[24] S. Sathiya, M. Balasubramanian, and S. Palanivel, "Pattern Recognition Based Detection Recognition of Traffic Sign Using SVM," International Journal of Engineering and Technology (IJET), Vol. 6, No. 2, 2014, pp. 1147-1157.

[25] S. B.Wali, M. A. Hannan, A. Hussain, and S. A. Samad, "An Automatic Traffic Sign Detection and Recognition System Based on Colour Segmentation, Shape Matching, and SVM," Mathematical Problems in Engineering (Hindawi Publishing Corporation), Vol.2015, No.250461, 2015.

[26] S. Lafuente-Arroyo et al., "A decision support system for the automatic management of keep-clear signs based on support vector machines and geographic information systems," Expert Systems with Applications, Vol.37, 2010, pp. 767-773.

[27] D. Kim, "Prediction Performance of Support Vector Machines with Fused Data in Road Scene Analysis," International Journal of Transportation, Vol.3, No.3, 2015, pp.41-48.

[28] C.C Lin, C.W Lin, D. Huang, and Y. Chen, "Design a Support Vector Machine-based Intelligent System for Vehicle Driving Safety Warning," 11th International IEEE Conference on Intelligent Transportation Systems, 2008, pp. 938-943.

Recent Advances in Intelligent-Based Structural Health Monitoring of Civil Structures

Satyam Paul[*,1], Raheleh Jafari[2]

[1]*School of Engineering and Sciences, Tecnologico de Monterrey, Monterrey, Nuevo Leon, 64849, Mexico*

[2]*Centre for Artificial Intelligence Research (CAIR), University of Agder, Grimstad, 4879, Norway*

A B S T R A C T

This survey paper deals with the structural health monitoring systems on the basis of methodologies involving intelligent techniques. The intelligent techniques are the most popular tools for damage identification in terms of high accuracy, reliable nature and the involvement of low cost. In this critical survey, a thorough analysis of various intelligent techniques is carried out considering the cases involved in civil structures. The importance and utilization of various intelligent tools to be mention as the concept of fuzzy logic, the technique of genetic algorithm, the methodology of neural network techniques, as well as the approaches of hybrid methods for the monitoring of the structural health of civil structures are illustrated in a sequential manner.

1. Introduction and Preliminary Concepts

The main intention of structural health monitoring methodology is to the expose of the area of damage and then to evaluate its location and severity thus predicting the life. Visual inspection and experimental measurement techniques are an important aspect of damage detection methodology. More and more research on the field of global detection methods in the identification of damages in the structure during and aftermath of seismic events has resulted in the investigation of methodology used in mechanical, nuclear, and aerospace engineering [1, 2]. The civil engineering structures like multistory buildings, towers and bridges tend to damage gradually with time elapse or due to sudden environmental disasters. So structural health monitoring is an important area of research that verifies the safety of civil structures from various issues. The stability and reliability of civil structure becomes a matter of concern after a strong earthquake and so more and more studies are carried on structural damage detection using different nondestructive computation methods. The vibration techniques are another popular method to identify the location of the damages with accuracy and thus considered to be global technique. The vibration methods associated with structural health monitoring detects the changes in the modal properties of the structure.

The constraints associated with these techniques are: a) the influence of factors producing same to same effects on the evaluation parameters which are not easy segregate, b) civil

engineering structures has superfluity that give rise to low sensitivity of the method in case of localized damaged, c) the uncertainties in the model, d) the noises in the measurements, e) in case of training data, uncertainties may incorporated in the data due to adverse conditions and d) in case of artificial neural networks, there is completeness of the conceived damage scenarios and training data sets [3, 6]. The method of damage detection of civil structures using non-destructive techniques (NDE) is an area of interest among the researchers. Its applications extents from aviation industries, manufacturing techniques, space shuttles, architectural industry, devices related to power plants, etc. The damage detection in structures cab be subdivided into local and global techniques. The local methodology are mainly applied for tiny and generalized structures for example pressurized vessels. The limitation of damage identification method involving local concepts is that it fails to detect the damages in case of large and complicated structures. In case of onsite structural damage detection, engineers can apply local detection method only to investigate some special component of the structures. The analysis of entire structure including some large and complicated structure is possible through global vibration-based identification of damages in structure [7, 8]. The variation in the parameters of structural modal is considered to be an early signal of the occurrence of damage in the structures [9]. The validation of reliability in function of numerous crucial technical structures by tracking online and detection of damage in continuous mode is possible using vibration based methodology. As a result of which vibration based method is growing area of interest among the

[*]Satyam Paul, Monterrey, Mexico, satyam.controlsystems@gmail.com

researchers. The main concept of vibration based damage detection is that the changes associated with the physical properties like mass, damping and stiffness will result in a change in the modal properties like natural frequency, modal damping and mode shapes. The past three decades have witnessed the contribution of the researchers in all three domains of time, frequency and modal domains. The effort of the researchers in this field will continue as existing methods cannot rectify all the damages that is incorporated in the structures [10]. In [11], the author has broadly distinguished between various methods on the basis of the level of identification attempted as shown in Table 1:

Table 1: Level of identification

Level	Various Methodology
1	To determine that the damage exists in the structure
2	To determine the damage geometric location
3	To estimate the level of extent of damage
4	To forecast the remaining life of the structure

Vibration-based structural damage detection can be broadly classified into traditional and modern type which is still under development and innovation. The description and comparison of traditional method and modern methods have been vividly described in forthcoming sections. The modern-type methods of structural health monitoring can be classified as wavelet analysis, fuzzy logic concept, neural network techniques, genetic algorithm methodology. In this paper, the intelligent technique related to structural monitoring techniques are discussed and analyzed. The intelligent methodologies can sub divided as neural network techniques, Fuzzy logic theory, concept of genetic algorithm and hybrid intelligent techniques. The researchers are taking lot of interest in the optimization methodology involving genetic algorithms [12] and neural networks [13] in order to investigate localized damage variation as a function of modal data. In the area of research associated to civil structures to the optimization of structure[14] as well as scheme related to the identification of structure [15] there has vast use of genetic algorithm for handling optimization problems. The detection of structural damages using this methodology is growing vastly. The synergistic combination of algorithms associated with Eigen sensitivity and genetic algorithm for optimizing the discrete variables of damage area as well as their extent utilizing natural frequencies is illustrated by [16]. The methodology of genetic algorithm in the detection of structural damages was presented by [17]. They utilized optimization fitness function techniques associated to displacements which is static in nature. Good linear mapping capability, inverse problems solving ability and high degree of robustness is the criteria of artificial neural network that makes it an important methodology in the field of structural health monitoring and detection. Several investigators utilized artificial neural network for the detection of localize as well as to compute the damage involved. The popular type of neural network termed as multi-layer perception (MLP) is utilized in health monitoring of structures as well as well in the situations involving identification [18, 20]. In the area of fault diagnosis and structural damage detection, there is a wide popularity of fuzzy logic for its embedded abilities in extracting precise and reliable data from imprecise as

well as uncertainty features [21, 22]. Fuzzy logic has been employed in wide areas of fault diagnosis and structural damage detection and is a growing area of interest. The developed fuzzification technology associated with fuzzy concept is counted for uncertain factors as well as the methodology of training approaches is uncounted in numerable phase involving noisy determinants in comparison to the neural approaches for accounting of uncertainty [23]. In [24], the authors demonstrated a concept to evaluate structural damage utilizing vibration data as well as fuzzy clustering. A novel methodology was demonstrated by [25] which is a combination of fuzzy logic, neural network and wavelet transformation for the detection of damages in structures. Researchers are taking widespread interests in the field of structural health monitoring that uses combined intelligent methodologies. In [26], the authors utilized a fuzzy-neuro hybrid technique to monitor the health condition of bridges. The decision system of fuzzy logic utilizes fuzzy clustering methodology in order to determine the damages in the bridge. A neural network forecasting technology makes use of back propagation algorithm to predict the correct amount of damage. An innovative filter design on the basis of combined neural and genetic algorithm concept for the improvisation health monitoring technique was proposed by [27].

In the next section, focused is vested on the traditional techniques associated to structural health monitoring. This section is followed by a modern methodology involved in health monitoring od structures which is followed by a brief overview of wireless sensors and then detailed analysis on intelligent techniques are illustrated. Finally conclusion is provided to give an overview of the paper.

2. Methods based on traditional techniques

The conventional methodology for the detection of damages in structure utilizes the mechanical approaches related to the structural attributes. The methodology involved in this type of techniques are transfer function computation or experimental modal evaluation. This techniques are not widely used in online detection of structures because the experimental data requires instruments which are diversified or involves manual operations. Therefore this technology has been not taken up with greater interest in recent studies. The main advantage of natural frequency analysis technique is that it is easily measurable in case of structures and does not depend on the measured position. Generally the preciseness of computation utilizing natural frequency methods is superior when compared with the mode shape analysis or modal damping. In [28], the authors proposed an innovative technique for extracting the information on the size and location of structural damage by utilizing the concept of natural frequency. In the current method, the location of the crack is obtained using Armon's Rank-ordering method and selecting the first four natural frequencies. In [29], the researchers proposed an innovative technique for the detection and computation of the damaged layout in structural devices utilizing a limited natural frequency. An algorithm with the properties of localized damages criteria is generated for tracking damages in terms of natural frequency variance as well as quantifying damage algorithms in order to predict the dimension of the crack. The method solely verifies the presence of large damage, but it is a matter of difficulty

to extract the damage location due to the fact that structural damage in different location may result in same frequency alteration. The changes in natural frequency will result in detection of damages in structure but in order to access the location of structural damage, information on vibration modes is utter necessary. In [30], the authors handled the constraints associated with the extraction of damages utilizing outcome based computation of vibration in the grip of varying environmental conditions. Using the computation techniques, two variety of features are extracted categorize under Eigen properties of the structure with help of automatized stochastic identification of the subdivided space methodology as well as with the help of peak indications calculated based on Fourier transform associated with the filters related to the modal. The method of modal analysis for the research on the damages on a wooden wall was investigated by [31]. They use the technique of damage-sensitive parameters computation for the investigation. The damaged zone is detected by inspecting visually the various period of damaged states related to the deformation in mode dimensions. The analysis of modal residue as well as variations in stiffness is an requirement for superior illustration of the damaged zone. The structural vibration mode is lower when compared with the natural frequency associated with the preciseness of computation although vibration mode generate more information related to damages. The damage detection methodology based on flexibility matrix change depends on the mode shape as well as reciprocal natural frequency and also the generalized criteria is satisfied by the structural vibration mode. Hence, in flexibility matrix the high frequency element will decrease in a rapid manner with the increasing of natural frequency. In [32], the author proposed a methodology for diagnosing damage in structure on the basis of change in dynamically computed flexibility and structural stiffness. The detection of damage is not only the primary objective, the detection of location is also very much essential. The implementation of the identification methodology utilizing covariance-driven subspace gives an idea of structural modal parameters. The main idea is to extract the dimensions of the matrix that are flexible and is related to the computed DOF. The concept of pseudo inversion will yield stiffness matrix. The damage localization is extracted with combined analysis of changes in two calculated matrices considering a reference state until the damaged state. In most cases, when some damage is visible in a structure, it is convenient to extract more information from stiffness matrix than the mass matrix. The one of the properties of stiffness is that it changes in astonishing manner when major damage occurs in a structure and therefore the change of stiffness matrix can be used to detect damages. But this method fails to do justice in case the damage is very small. Non-inclusion of mathematics model, vibration mode measurement and experiential knowledge in frequency response function methods makes it an advantageous model. So this model can be used for monitoring online. In [33], the researcher proposed and developed new monitoring techniques considering higher order spectra for the frequency response functions in order to monitor involved nonlinearities in structure as well as non-Gaussian types of signal in case of damage as well as the computation of related to the phase-coupling of harmonics signal associated with the structures. The proposed methodology is in general traditional functions associated to the frequency response for the investigation of spectrum for actual value transformations

in higher order. A damage identification technique relied on artificial neural network in combination with frequency response functions was investigated by [34]. This technique can effectively analyze nonlinear damages suitable to some level of excitation. The important concept of their work is to illustrate a practical methodology associated to vibration dependent structural health monitoring. This innovative technique reduces the dimension of the initial frequency response function values and converts it into latest damage indices and engages technique of artificial neural network for the exposure of different level of nonlinearity with the help of damage patterns recognition. The main disadvantages of this methodology is that the quantity as well as the zone of measured point have an impact on the accuracy of structural damage detection. A newly developed concept for the identification of structural damages based on the dual methodology of instant identification of damping coefficient and wavelet transform was presented by [35]. The combination of experimental results and numerical simulations justifies that in structural systems with incorporated damages results in crucial variations in damping. Thus, the parameters linked to the structural damping can be used as sensitivity to damages associated to system properties. In [36], the authors proposed a technique where sensitivity-based finite element (FE) model updating methodology is implemented for damage detection. The optimization algorithm is utilized for the minimization of the objective function and to initiate the damage detection techniques. In [37], the authors developed a novel damage indicator termed to be as mode shape area index on the basis of mode shape changes. This developed methodology found out to be as the most sensitive damage detection approach. A statistical technique with combined uncertain frequency as well as mode shape data for the detection of damages in structure is suggested by [38]. The suggested technique is implemented to a laboratory tested steel cantilever beam and frame structure. The experimental results validate that all the damages are detected suitably with high probabilities of damage existence. The utilizing of frequency response functions for the detection of damages in structure was proposed by [39]. The suggested technique iteratively maximizes the elements related to the damages by not including the elements which were validated as undamaged from the sets of good elements. The mentioned approaches was applied to numerical simulations using a structural setup having 2D frame, also the efficiency was established. These conventional techniques have some constraints to be mention as [40]:

- Due to its dependency on experimental results to be mention as modal shape computation as well as the calculation involving damping but it incurs computation cost. This creates a negative impact on online damage detection techniques.
- In terms of damage location identification, this methodology is not popular because of its complexity to develop a recognized methodology considering several structures due to its severe dependency on the acquirement of the individual structures which are required for the detection.
- The conventional techniques fail to unearth the minute damages associated with the structures.

3. **Detection techniques based on modern methodologies**

Modernized pattern of vibration reliable structural damage detection which is also known as intelligent damage extraction

technique is a renowned methodologies for the detection of damage utilizing online computed structural vibration responses. The methodology utilize combined technology of current signal processing concepts as well as artificial intelligence analyze mechanisms. The methods fall under this category is wavelet analysis, neural network, genetic algorithm, etc. Wavelet analysis is an upcoming research filed mainly in the area of signal processing. It has been implemented with success in many applications such as transient signal analysis, image analysis, structural health monitoring system etc. The existence of damage in structure can be identified using the spectrum graph obtained using wavelet transform. A wavelet based concept for the detection of structural damages was proposed by [41]. In their work, they illustrated the analysis of vibration signals characteristics associated with wavelet transformations. The concept is implemented to simulated data extracted from an complex structural model under the effect of harmonic excitation. The model is made up of numerous breakable springs, of few may undergo irreversible damage when the value of response exceeds the value of threshold or the number of cycles of motion is stored beyond their fatigue span.

In [42], the authors demonstrated an innovative concept on the basis of fuzzy concept involving wavelets phenomena for the identification of damages in structure. The methodology is the combination of Wavelet Packet Transform which is used for feature extraction and the capabilities of fuzzy sets to model vagueness as well as uncertainty. The operation are carried out in two steps mainly the training of pattern and the monitoring of health. The general difference illustrating the importance of each technique is given in Table 2. The elaborate explanation about the intelligent methodologies associated with structural health monitoring has been depicted vastly in this survey paper.

4. Importance of wireless sensors for the health monitoring of structures

The technology of wireless sensing is an important area of research preferred by the researcher in the structural health monitoring. The wireless sensing unit is designed and developed for reliable communication of response measurements as well as for power efficiency. The extraction of signals for structural health monitoring is an important step and so a short illustration of wireless sensors is provided in this section. In [43], the authors had illustrated a brief review on the experiences of the investigators working on wireless sensors as well as sensor networks for maintaining records on structural performance and health. The structural monitoring systems involving wireless technology is a popular area of research because of its positive criteria such as low cost instillation since wiring is not essential between the sensors and data acquisition device. The stated methodology plays superior roles in the processing of structural response data since it is implemented to validate data involving signs of damages associated with the structure. The analysis on results related to international cooperative analysis on smart wireless sensors was carried out by [44]. In the initial phase of this research, the concept associated with the smart wireless sensor is investigated. In the next phase, the subsystems associated to the smart wireless sensor are illustrated. Apart from that, available wireless sensor platforms manufactured in the research companies are investigated. In the final phase, three smart wireless SHM technologies were

demonstrated by the investigators which has vast scope in the identification of structural damages. In [45], the reseacher developed a novel wireless sensing unit device for intelligence monitoring of remote systems. The sensor is utilized for bridge as well as infrastructure SHM. The design is carried out based on cost reduction and superior performance. In [46], the authors illustrated an innovative wireless sensing device for structural health monitoring of historic landmarks. This technique minimizes the problems connected to the wired monitoring systems. In [47], the authors illustrated a design of an innovative low powered academic wireless sensor model for the monitoring of the structure. The effectiveness of the sensor design is assured via performance analysis. In [48], the researchers suggested a wireless sensing model incorporated with multitasking capabilities using recently widely available incorporated system components.

Table 2: Comparison of methodologies

Fuzzy Logic	Neural Network	Genetic Algorithm
Fuzzy logic is superiorly capable of handling uncertainties and complexities in structure with ease and effectively.	In case of arbitrary complex functions situations, neural network is considered to be as generalized function predictors.	Genetic algorithm is considered to be a powerful tool when implemented to solve combinatorial problems and so is advantageous in case of structural health monitoring.
Fuzzy logic is simple, easy to implement and robust in nature.	Neural Network has the capabilities of predicting the structural health during the operation without the necessity of interruption or the termination of the process.	Genetic algorithm is capable of determining the location of damage more precisely and computes the magnitude of damage more significantly.

5. Involvement of intelligent methodologies in structural health monitoring

5.1. Genetic algorithm

Genetic algorithm (GA) is advantageous because it offers easy implementation due to the fact that only an objective function is necessary and derivatives or other supplementary details are not required. So this attributes of GA makes it strong and most trust worthy in the field of structural optimization [49, 50]. The technique of GA have been used in model updating and detection of damage by few researchers in recent studies [51]. The technique of optimization utilizing the concept of genetics which is considered to be a zone of great interest among researchers investigating on the localized damage alteration as a function of **sets** involving modal [52]. In [53], the author had presented a technique of real number encoded genetic algorithm which is utilized in analysis of damages in structure utilizing method of objective function minimization. On the basis of available vibration data 3 separate attributes to be mentions as frequency variation, mode dimension variation and combination of this two

are taken into consideration. The direct collation is done by the objective function by the changes in the measured vibration extracts via damaged as well as undamaged structure before and after update. In their work, they used laboratory certified cantilever beam and frame to verify the efficiency of real-coded GA for the verification of structural damage. In [54], the researchers implemented the methodology of genetic algorithm in order to identify damages in vibrating structures. They utilized the concept of residual force matrix in combination of optimal methods for the identification of damages. The mentioned methodology is verified by the use of a doubled centered crossover dual embedded genetic algorithm complied with selected tournament for reproducing population. The optimal solution through GA is computed by utilizing C program for generating programs which utilizes the concept of eigenvalue forecasting code at par with sub program termed as force function residue. The methodology involves an appropriate selection of crossover as well as mutation operators. Considering the samples of 40 individuals with crossover probability 1.0, mutation probability 0.001., and each structural parameter β_i was illustrated as a 10 bit binary number having variable limits from 0 to 1. The main complexity involved is in obtaining the minimal force residue. Also the function related to the fitness is selected in the current matter termed to be as inverse function is illustrated as:

$$M = \frac{C_1}{C_2 + f(\beta_1, \beta_2, \dots \beta_n)} \quad (1)$$

C_1 and C_2 are valued with number 1 and considered to be constants. The identification of damage parameters for various structures is done from associated residual vectors minimization problems that are handled by genetic algorithms. This technique explored the excellent bonding with those chosen for the mechanical simulation of these damaged structures. Tso and Shen [55] have used residual forced vectors in identification process based on neural network to display the properties of good location and severity assessment. In [13], the author had presented the concept of vectors associated with residual force in order to point out an objective function for a technique of optimization, which is imposed using genetic algorithm. The main intention involving the detection of structural damage is the formulation of parameters based on objective function. The formulation rules of objective function should include the process of obtaining the maximum value during the process of evaluation taking into consideration the true parameters. This is a matter of importance as the technique of optimization like the concept of genetic algorithm may be implemented to extract the parametric values using iterative methodology by choosing values for maximizing the objective function. The structural situation about the damaged states incurred can be established using the procedure of optimizing to get the solution values of the parameters. For addressing the identification of damage at par with constraint optimization related to the concept of GA, it is utter essential to broaden up the objective function. The suitable objective, which is required to be maximized by using the general rules of residual forces and taking into consideration several practical aspects, is given as follows:

$$J_{obj}(\beta_1 \dots \beta_i \dots) = \frac{C_1}{\sum_{f=1}^{r} R_f^*(\beta_1^* \dots \beta_1^* \dots)^T R_f^*(\beta_1^* \dots \beta_1^* \dots) + C_2} \quad (2)$$

where C_1 implies a constant used to control the value of the objective function C_2 represents a constant used to construct a well established function for the ideal case and r represents the number of modes taken into consideration which is determined experimentally or used in simulated examples. In [56], the authors had used the approach of wave propagation having a synergistic combination with genetic algorithm in order to identify the damages in beam structures. In [57], the authors has proposed a technique of training a neural network taking into consideration the input parameters as strain and frequency whereas the output is taken to be the damage level. The neural network parameters are selected using GA. In [58], the authors had presented a GA-based methodology for the extraction of crack zone as well as aluminum made beam depth. The location of crack and depth in a cantilever beam is identified and constructed as an optimization problem. The optimal location and depth is found out by using binary and continuous genetic algorithms by implementing cost reduction scheme on the basis of extracted and computed changes in frequencies. Initially the chromosomes are arranged on the basis of their fitness. The selection probability P_n for each chromosomes considering rank n is:

$$\frac{N_{keep} - n + 1}{\sum_{n-1}^{N_{keep}} n} \quad (3)$$

Single point crossover is used for binary GA. In case of continuous genetic algorithm, the used variable value considered is unit offspring. P_{new} is considered to be the outcome of binary associated variable illustrated as:

$$P_{new} = \beta P_{mn} + (1 - \beta) P_{dn} \quad (4)$$

The interval [0, 1] is selected for β and is termed to be a random value. Also, P_{mn} is considered to be the nth mother chromosome variable. Finally, P_{dn} is nth variable associated with father chromosome. The methodology of genetic algorithm is implemented in this investigation for the identification of beam cracks. The identification methodology associated with the approach use unit and continuous GA for the identification of depth and crack. In [59], the researcher briefly defined a technique for damage identification of structure based on micro GA which utilizes test data involving noises and incompleteness. The computation of modal data is carried out by installing some sensors. In the primary phase, the expansion of mode shape data is executed in phase with all DOF related to the finite element model. The next step is to implement elemental energy quotient difference in order to investigate the damage location approximately. The final step includes the utilization of a micro-genetic algorithm for the quantification of the damage by minimizing the errors between the computed data and numerical outcomes. Ruotolo and Surface [60] used the methodology of genetic algorithms to identify structural damage in structure by the modal residuals optimization. In [61], the author had presented a combined methodology of genetic algorithm and finite element method for detecting and locating damage. In [62], the researchers had revealed a methodology based on nondestructive global damage detection and assessment computation associated with frequency changes and mode shapes of vibration of a structural system. This is an elementary level methodology based on finite-element model.
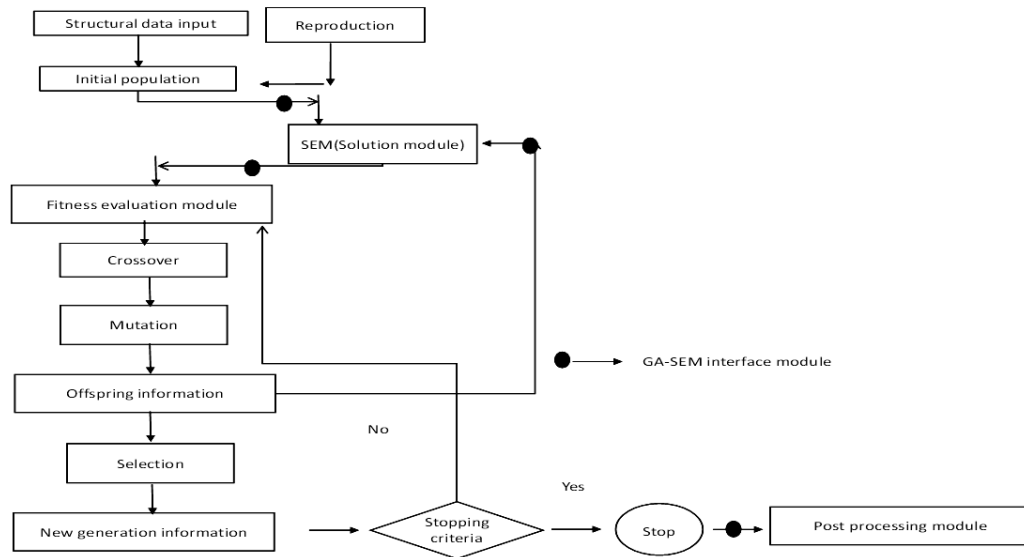
Figure 1: Flow chart of genetic search

Continuum damage mechanics suggest that the representation of damage should be done using a reduction factor of the element bending stiffness. In order to localize the area of damage in structures, an optimization approach with non-classical touch involving the implementation of genetic algorithm is laid down. Innumerable cases of damage instances has proved the efficiency of this method when applied to simulated beams and also by use of direct data from experiment which is extracted from the vibration tests of a beam. The proposed technique of genetic algorithm generates an appropriate location of damage and extreme detection while offering number of benefits in comparison to classical methods. In [63], the authors had presented an effective procedure involving optimization technique in order to locate multiple damage in structures. One indication that the structure is under the grip of damage is the natural frequency variation of the structure. The extraction of information on the basis of natural frequencies is accomplished by using a finite element analysis (FEA). A novel technique involving tweaked genetic algorithm(MGA) having novel operators termed as health as well simulation components are illustrated in significant way for detecting range and location of damages. The efficient combination of MGA and the ECBI that result in a effective robust tool for detecting structural damage is validated from the numerical results for different cases of damage with different sample. This methodology can be used to detect the real damage sites and extent thus decreasing the total number of FEA in comparison to the available tools like SGA-MDLAC. In [64], the authors had suggested a technique to compute the extent of damage as well as precise location using the concept of natural frequency shifts. The combination utilizes real genetic coded algorithm as an optimization tool to verify natural frequency with those extracted by using the technique of parametric finite element code via modified MDLAC index. A Binary-coded genetic algorithms (BCGA) concept was implemented for the detection of structural damage associated with GA technique [65]. In recent days, the trend of studies in the field of real-coded genetic algorithms (RCGA) is on the up rise [66, 67]. This approach works better with variables that are continuous and there is less requirement of storage than BCGA. An improvised GA technique for the detection of frame structure damages was presented by [68]. In their work they suggested discrete values to illustrate the damage. This methodology is composed of initial population of heuristic generation, two dynamically changing fitness functions on the basis of modal data and two appropriate mutation operators. In [69], the author suggested an algorithm in two stages in order to sense structural damage. The first step is to define an search algorithm in order to search for expected damages. In second case, the algorithm scanning is accomplished in order to uplift the level of knowledge. GA is utilized in all the cases. In [70], the authors had proposed a hybrid type real-coded genetic algorithm for the detection of structural damage. The objective function is incremented with a damage penalization in order to discard fake damage detection caused by experimental disturbance or error in computation. The algorithm verification is carried out using a three dimensional space frame structure taking onto consideration both single and multiple damage cases. This research reveals that the genetic algorithm approach used gives more accurate results than conventional optimization methods. The successful detection of real damage was made possible while neglecting the false damage detection. This methodology is widely used for large or complex damage criteria and preferred over traditional optimization techniques which fails to extract the global optimum. A hybrid methodology involving optic fiber sensor and GA was developed by [71] for structural health monitoring. The methodology involving random generation is utilized to develop the preliminary population having three individuals. The researchers have utilized the concept of objective function which is illustrated by addition of squared differences associated with the beam eigenvalues extracted obtained by a sensor termed as POF. Compilation of the frequency associated with the modal includes randomly generated individuals related to the initial population. The objective function is given as:

$$f(m,d) = \sum_{i=1}^{i=3} (f_{data}^i - f_{exp}^i)^2 \qquad (5)$$
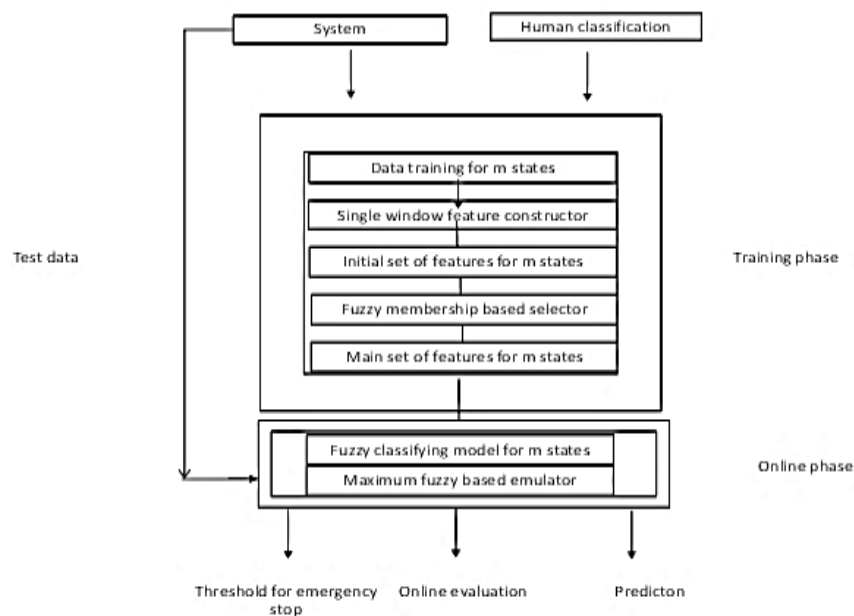
Figure 2: Fuzzy filter approach

objective function is carried out so that solution variables having minimal value will have the higher fitness value. GA is utilized in this investigation which is a novel approach. The one with the strongest feature is permitted to shift to the forthcoming generation. The approach of reimplementation associated with the chromosomes are embedded into the new population related to the forthcoming generation thus not permitting any variations in the GA. The approach of joint general GA along with sensor made of plastic optic fibre are depicted for the health monitoring of structures. The methodology utilized in this research can be implemented for identifying damages in structures made of concrete as well as composite materials. The identification of delamination in composite beams as well as joint structures utilizing an efficient identification concept was illustrated by [72]. The predicting technology based on modal analysis combined with damaged spectral element for extracting the structural response damage associated with the frequency domain is proposed. A novel type of curved genetic algorithm (GA) associated with a with finite element code for automatic is utilized for structural damage identification. The techniques involved in genetic search is illustrated in Figure 1. A technique for inverse problem is suggested for the detection and localization of damage in composite beam like structures by [73]. The developed methodology utilizes Genetic Algorithm (GA) as well as Particle Swarm Optimization (PSO) techniques. The basis of the objective function is the computation of natural frequencies and Modal Assurance Criterion. In [74], the author suggested a hybrid methodology based on of standard GA in order to improvise the stability associated with the EM algorithm involved in the findings of clusters optimizing number and also the parametric values thereby expanding the damage classified results. The importance of the illustrated approach is proved by utilizing real time sets of data extracted from the Z-24 Bridge situated in Switzerland. A varied feature GA was illustrated by [75] in order to detect the accurate locations as well as extents of the imparted minimal

damages in the structure. The proposed methodology detects the damages to the suitable extent.

5.2. Fuzzy logic

In 1965, The concept of fuzzy theory was proposed by Zadeh [76]. Fuzzy algorithm is capable of processing simulated thoughts and judging capabilities of humans. One of the essential attributes of fuzzy control rule is its linguistic criteria which can be tweaked very easily and can be understood with ease**.** The nonlinear mapping can be handled effectively by the utilization of fuzzy set theory as well as fuzzy logic. The main attributes of fuzzy logic system is to deal uncertainty easily and with more flexibility thus making it popular among researchers. Fuzzy logic manages the oddly situations in a suitable way, also it is very sensitive to information such as it is possible for a peak may to get established as fault in the same time without creating problems [77]. Regarding the in-between laid dependencies, it is proved that the concept of fuzzy approaches results in superior outcomes [78, 79]. One aspect that makes fuzzy system a convenient tool for analysis is that it has convincingly less number of rules when compared to classical expert system [80]. Fuzzy logic technique, has astonishing capabilities of damage detection in structures along with ambiguous data processing capabilities [81, 82]. In [83], the authors had shown an intelligent approach to alter the human procedure with a computer by characterizing a system based on fuzzy concept by utilizing decision approach based on human attributes. The feature of developed technology is that it will grant affordable way of encoding methodology related to the rules set. The designing of fuzzy expert system is laid down on the basis of FE model associated with simple beam. Also, it is **s**uccessful in validating reliable source of damage detection for every scenario of damage analysis. There is 100% accuracy in the detection of damages by this system and no fake positives or negatives on the shapes of the mode obtained by impact testing and or by using vibrometer procedure that are non-contact. The analysis reveals

that the expert system involving fuzzy portrayed in this paper displays a decisive and precise results in recognition of various cases of damage. The performance of thresholding based methods is overshadowed by the performance of this system mainly in the case of multiple damage. The elaborately defined example in this paper suggest that the thresholding based technique is capable of detecting only one of the two damaged areas but the expert fuzzy system proposed is well enough to detect both of the damaged areas. The decision process can be processed very easily by installing the damage detection system as an expert system. The intervention of fuzzy logic results in comfortable solution of the unreliability and denial within the rules set and while taking any decision. In [84], the authors had proposed a technique of conditioning monitoring and approach based on classification using fuzzy-filtering. There is a necessity of defining various models for differentiating different states of operation in a machine. This problem is discarded by model based on qualitative methods and signal based methods due to the involvement of innovative concept principle of modeling. In their work, the technique of qualitative model associated with fuzzy concept and statistic stated signal properties is combined synergistically. A conditioning monitoring system is designed on the basis of important signals, alternate measurements experimental data as well as the classifying knowledge of the system which is required to be verified. The concept behind the fuzzy filter is stated by a flow chart illustrated with Figure 2. The filtering methodology is illustrated relying on the concept of qualitative model based technique utilizing fuzzy concept in combination with statistics concept associated with signal based methods**.** The mentioned methodology gives the raw data conversion to a novel presentation for approaching the classification. In [85], the authors illustrated a technique for the validation of overlapping health fuzzy sets by utilizing presumed distributions of statistics associated with damage attributes. The superior impacts of implementing progressive stages of damages to be mention as crack, defect as well as flaw for laying down a systematic investigation of damage using the concept of supervised as well as unsupervised learning was mentioned by [86]. In [87], the authors implemented the methodology of fuzzy similarity for monitoring structural health utilizing damage pattern recognition. The health features are illustrated by fuzzy sets that are developed by utilizing the methodology of healthy observations. The representation of healthy as well as damaged states are accomplished by fuzzy membership functions which is utilized as a combination with piecewise linear functions. The model steel bridge is consider as a case study for identifying damage pattern in their work. The precise estimation of damage patterns is achieved by regularizing this methodology. The fuzzy set associate with health which is a left shouldered triangular membership function is illustrated as:

$$\mu_H(\lambda) = \begin{cases} 1, \lambda \leq \bar{\lambda}_H \\ \frac{(\lambda H^* + \sigma_H) - \lambda}{\sigma_H}, \lambda H^* \leq \lambda \leq \lambda H^* + \sigma_H \\ 0, \lambda \geq \lambda H^* + \sigma_H \end{cases} \quad (6)$$

$\mu_H(\lambda))$ is stated as membership linked to the fuzzy set illustrating a good forma with a mean of λH^* having a spread of σ_{λ_H}. The membership function with good health illustrate that healthy fuzzy set considers a data interval which are completely leaned towards

healthy outcome. The technique of pattern recognition associated with damaged states with unknown capability sets is implemented utilizing the concept of fuzzy similarity as well as maximum convex degree. The methodology depicted in this investigation can precisely detect the damages in the structure. The proposed technique is nonspecific for any SHM system due to the fact that it is independent of the damage feature. An advanced methodology was illustrated by [88] for the identification of location as well as damages extremity in the area of stochastic structures relied on ARMA parametric model combined with fuzzy classification. The method suggested in this paper has a high significance as it uses fuzzy viewpoint on stochastic structural damage analysis. In this case, the feature vector is considered to be ARMA parameters. The ability of this technique is also minimize the stage at par with the optimization which is utilize to extract the values related to the membership functions of fuzzy technique by implanting changes of predicted ARMA values associated with membership functions. In this current investigation, the fuzzy classified inputs are ARMA model's parameters and the outputs are specified as elemental damaged numeric as well as the level of extremity. Considering all features, the antecedent fuzzy sets (membership function) is illustrated as:

$$\alpha(f_i^j) = \exp[0.5] \left(\frac{f_i^j - m_i^j}{\sigma_i^j} \right)^2 \quad (7)$$

where, $\alpha(f_i^{\ j})$ = ith membership function of jth feature, $f_i^{\ j}$ = absolute value, m_i^j = related mean value (midpoint) and σ_i^j = variance of the jth feature in ith class. In [89], the authors suggested a noteworthy methodology on the basis of fuzzy clustering combined with computed frequency response function samples minimized via principal component projection in order to investigate damage detection in truss bridge model using vibration based technique. The preliminary data utilized in this investigation are the FRFs associated with the healthy as well as the damaged structure. The normalization of FRF data is carried out to eliminate the effects caused by the environmental and operational variability. The analysis by principal component at par with kernel principal component investigation technique are laid down for handling the compression of data. In extraction methodology of damage features, the principal components of median values are used. Finally, fuzzy c-means(FCM) clustering algorithm is applied for the categorization of structural damage detection features. In [90], the authors had suggested an approach of novel binary phase detection of damages in structures utilizing the theory of fuzzy-neuro methodology and techniques of data fusion. This method is widely used for the identification of damage considering the condition that computed sample is vast and involves uncertainties. Primal phase of damage detection involves the collection of modal attributes of structure from the structural vibration responses that are embedded as an input to the FNN. The output values associated with the FNN are defuzzified for delivering a rough evaluation damages in structure**.** In the second phase, FNN model reveals three output values which are inputted to the center of data fusion in order to execute fusion computation. The result is then filtered using threshold function in order to carry out final fusion decision. Thus a superior quality damage assessment methodology is achieved. The innovative approach is implemented in 7 DOF building for the validation of damage identification of structure.

The results are effective, reliable and satisfactory. In [91], the researchers had proposed an modernized approach to detect damage in structures using a new algorithm based on the fuzzy cognitive map (FCM). In this research, damage in structures is modeled with the help of continuum mechanics approach. The technique of modeling using finite element method for cantilevers is utilized in order to compute the variation of first six beam frequencies due to structural damage. In [92], the researchers illustrated the concept of structural damage detection based on data fusion as well as semi-supervised fuzzy C-means clustering. This method performs superiorly when compared with other methods for detecting damages in structures. The experimental results associated with a benchmark model validates the effectiveness of the proposed methodology. In [93], the authors illustrated a fuzzy inference model for the extraction of building damage hazard, normally associated with the non-technical building houses considering special sort of earthquake event. The laid down model consists of 3 level model consisting of fuzzy rules which utilizes numerable data associated to the damaged buildings obtained from the impingement of earthquake occurred in 2013 at Indonesia. The important contribution associated with this research is vested towards the development of the function as well as membership related to fuzzy taking into consideration the variable of each determinant associated to building house damage warning as well as three levels of fuzzy inference procedure for extracting building house damage hazard as an earthquake occurrence. A system based on decision making and associated with an adaptive network-based fuzzy inference system (ANFIS) model was proposed for the extraction of damage level of building by [94]. The suggested methodology was tested over earthquake data set of Bam city in Iran. The proposed methodology was accurate up to 76.36% for the detection of damage degree of building. A technique for damage identification in buildings using the approach of superior dimension remote-sensing imaging combined with fuzzy reasoning concept was presented by [95].

5.3. Concept of Neural Network

A neural network consists of inter connected neurons and layers and it is one of the most important means of computing weights which are relied on the connections illustrated as learning algorithms. In [96], the authors has vividly described the concept of common anatomy of neural network by sub dividing it into seven important zone as follows: i) processing unit set, ii) activation set of processing unit, iii) calculation of processing unit output using a function, iv) connectivity pattern alongside the processing unit, v) implementation of propagation rule, vi) instillation of activation function and vii) confirmation of learning rule. In [97], the author had investigated the primary research stage inclined at the development of automatic monitoring methods for the detection of damage in structure. In their study, the technique of self-organization and learning capabilities of neural network are explored for the assessment of structural damage. The concept involved is the training approach of neural network to identify the undamaged structure behavior alongside the behavior of the structure considering numerous possibility of damage states. The detection of any existing damage can be made possible by impinging the training network for the computation of responses from structure. This concept was implemented to a simple structure to get a motivated results. The initial outcome of results

verifies that the neural network are capable enough to learn the nature of damaged and undamaged structures and thus can be used to identify the damaged member and the extent of the damage by the approach of patterns in the frequency response of the structure. In [98], the author utilized the concept of counter propagation neural network for revealing the beam damages, also damages in a frame as well as support movements of a beam along axial direction. The research involves a wide range of diagnostic parameters, along with static displacements, natural frequencies, mode shapes and other mode shapes parameters. The fetching results are quite interesting showing that these parameters can be utilized as a diagnostic parameters for artificial neural network applied to structural engineering. In [99], the researchers developed an artificial neural network technique associated with system identification concepts extracting the characteristics of composite structures damage. The methodology implemented in this research consider a training phase as well as recognition phase. The different types of damage considered to be as patterns are associated with structures candidate model in the training phase. The area and the extremeness of the damage decides the arrangements of patterns into pattern classes. In order to extract the transfer functions, the identification of system are carried out which is an important attributes associated with the structural systems. The extracted transfer functions is used as input patterns for training which are transferred to multi-layer perceptron. Numerical validation proves the efficiency of the developed technique. In [100], the authors implemented an architecture on the basis of multiple layer perceptron in combination with the learning algorithm based on backpropagation for modeling a bridge by utilizing data from damaged states. By using the methodology of general structure analyzer, the patterns of the training were laid down by predicting the zones of structural damages. By creating a comparison between output and the performance data that are created algorithmically without considering it in the training process, the working of the network was displayed. The problems associated to the working of the network is thoroughly investigated. For the estimation of structural damage it can be accomplished from various investigation done that the perceptron model is completely applicable. For the described case, the functioning of the architecture of an individual layer is not more sophisticated than the performance with double hidden layers. The investigation done on the calculated data based on the implications of engineering is considered by the evidence that for the identification recognition task barely a certain area in the structure required to prepare the network. In [101], the authors conducted an experiment five storey frame shake table and validated that there are problems leading to damage states occurred during the event of implementation of neural concept that is trained utilizing trained samples obtained by using FEM. The obtained results proves the superiority of the technique. In [102], the authors illustrated a novel network associated with proposition for the identification of damages in structure. In the preliminary part, the technique initiates with identification of system utilizing networks on the basis of neural concept recognition in order to extract the damaged as well as undamaged states. The next phase deals with the structural damage identification using neural concept in combination with damage identification network for the area detection as well as severity of the structural damage. Considering n hidden layer in the network having output with k^{th} node:

Step 1: Training of decentralized neural networks

Step 2: Decentralized damage detection
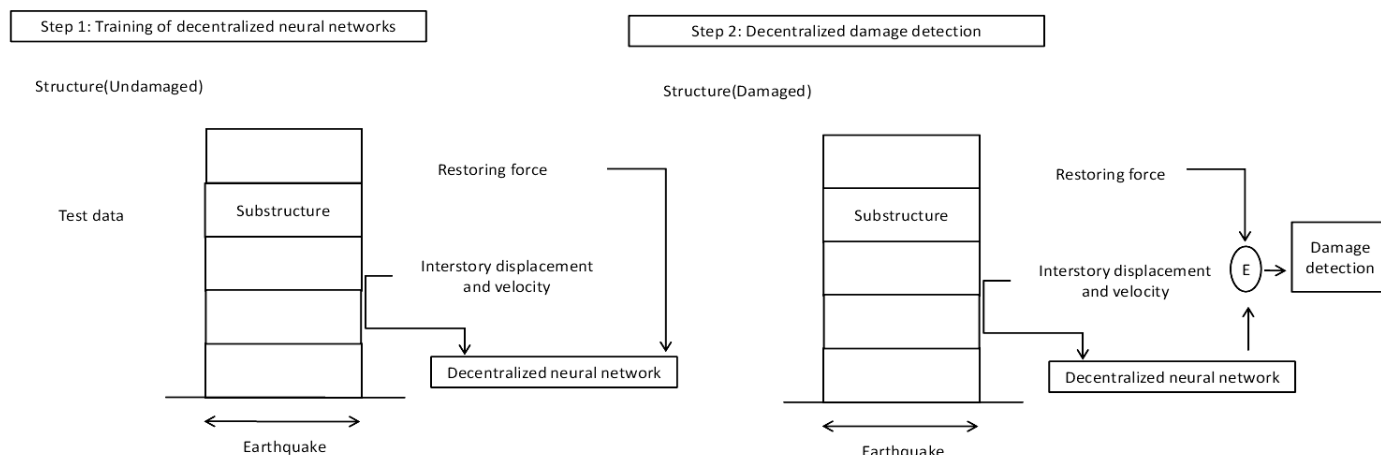
Figure 3: Decentralized damaged detection

(8)

H_{nj} is the illustrated output at par with j^{th} node having n^{th} hidden layer. Again, θ_{ok} is the threshold parameter related to k^{th} node of the output. Also, $W_{hnj,ok}$ is termed to be the weight of the jth node embedded in n^{th} hidden layer of the kth node of the output. The activation function is illustrated by f. The approaches implemented in this research is highly effective in structures involving nonlinearities, as neural concept can handle linearity and non-linearity with same technique. An innovative methodology of decentralized information structures related to the neural network concept was proposed by [103]. An identification technique involving parametric damage with decentralized pathway is implemented to the structures with multi degrees of freedom. In preliminary stage, detection of damages associated with substructures of a multi degree of freedom system is implemented and is applied by the method of neural networks. The training of the decentralized neural network required for the detection of analogous substructure is carried out based on displacements as well as velocity computations obtained from a good health substructure system as well as from the substructure restoring force. The differences in forces between the damaged substructures as well as the undamaged substructures is computed by means of trained decentralized identification neural concept. The figure 3 demonstrates the block diagram related to decentralized damage detection with the application of neural networks. The significance quality level of this methodology was inspected using numerical simulations. In this current research, the investigator has shown the effectivity of this method by demonstrating it as a practical tool for damage detection which can be utilized to structure-unknown smart civil structures. In [104], the authors laid down the technique of modular neural network approach in for the extraction of crack zone as well as depth associated with a cantilever beam. In the mentioned investigation, principal component analysis network is implemented for size mitigation as well as for unearth the concealed structure associated with the data set. In the adopted modular network approach, the crack location is estimated using one neural network whereas the other is used for computing depth of crack. In [105], the researchers presented an algorithm for the detection of damage in beam like structures with prediction of damaged ranges and zone

by implementing the novel methodology of global and local vibration dependent analysis samples as input in neural concepts. Performance sensitivity analysis give rise to the essential features for damage detection. Various input-output sets are incorporated to various ANNs. The verification of input robust features used in the methodology as well as the actuation of the experimental uncertainties are done by generating artificial random noise and adding noise free data during the training procedure of the ANNs. The procedure of sensitivity analysis is carried out on the extracted features by using different modes of vibration taking into account the effect of the area of damage as well as severity before introducing them to ANNs. The various steps involved in the investigation starting from vibration based analysis for feature extraction in order to use as an input for the ANNs, noise level on these features, the methodology of the experiments, measuring instruments to the effectiveness verification at different vibration modes plays a significant role in extremity and location forecasting of the damage in beam like structures. In [106], the authors had proposed a methodology for damage detection using neural networks in association with modal data which impart the parameters of input to the modeling errors. In their study, to implement the methodology practically, the estimation of damage in bridge structures is demonstrated utilizing the technique of ambient vibration samples generated by traffic loadings. By initiating two numerical processes, the effectivity of the laid down concept is validated. The next step involves tests in laboratory as well as field tests on the structure of the bridge. The results are promising and effective which validates the efficiency of the neural network technique for damage identification, locations detection and severity analysis. In [107], the researchers illustrated a MLP neural network relied concept for damage estimation of structure related to the truss bridge joints, which is obtained from modal specification of the structure. Two mathematical investigation samples are illustrated to confirm the superiority of neural network concept. In [108], the authors presented an innovative approach of neural network based structural damage detection. The proposed methodology can be subdivided into categories. The first step includes system identification techniques that uses neural system identification networks (NSINs) to track the undamaged and damaged states of a structural system. The second step is associated to structural damage detection that uses trained NSINs to create free vibration responses taking into

consideration the same initial condition or impulsive force. The level of changes can be made available by comparing the periods and amplitudes of the free vibration responses of the damaged and undamaged states. In [109], the authors demonstrated a statistics based neural network approach by considering finite element modelling error as well as measurement noise for the detection of structural damages. The investigation on numerical as well as experimental outcomes reveals that the superiority of statistical neural network concepts approach in comparison to the normal neural network approach generates more precise identification of damages in structure. Once the model training of neural network is over, the testing samples are then used to identify the locations and severities of any damages. In [110], the authors investigated a damage detection procedure with the help of identification of pattern methodology associated with vibration. The strategy is evaluated utilizing a FEM of a suspension bridge which is a century old. The response spectra give rise to feature vectors which was installed to binary unsupervised neural networks for evaluation. Investigation reveals that the tweaking of the sensitivity of the neural networks is accomplished for achieving a superior rate of damage detection along with the existence of noisy signals. In [111], the authors suggested the methodology of novelty detection and damage computation. In this process, signal parameters obtained from reference structure are used as a data set. Also soft computing methods are used to notify about the damage visualization, location and its type. The results obtained from the investigation shows that the suggested methodology makes automation of structure testing possible and can be applied to SHM systems. In this investigation, a two-stage SHM algorithm was developed which can predict structure damaged height and width. Neural Network (NN) trained novelty detection is used for the first level identification. The second level of identification is conducted using the NN trained for damage prediction. In [112], the authors proposed a damage detection methodology on the basis of computational method to identify damage in structures using neural networks. In this process, the training of the neural network is done for the recognition of the pattern of static value associated with the undamaged structure as well as of the structure having several possible damage extent and location which were modeled as random states. The damages in the cantilever beam was identified using this methodology. This study reveals the fact that the proposed methods using strain is more effective than the displacement method. In [113], the researchers presented an approach for the detection of small damage in structures using ANN technique with progressive substructure zooming. In this work, a combination of substructure technique along with a multi-stage ANN models in order to detect the area and extremity of damage. The input for ANN are considered to be modal parameters such as frequencies and mode shapes. The higher performance of the technique is validated using a two-span continuous concrete slab structure and a three-story portal frame. The experimental results shows that this approach is able to identify all the simulated damages in the structure. An identification methodology for the extraction of damages in structure by implementing the technique of parametric identification combined with iterative neural network (NN) concept was proposed by [114]. A variable back-propagation learning algorithm is proposed for tackling possible saturation of the sigmoid function as well as to run the training process. The structural parameters are predicted by utilizing the trained NN model which is implemented in measured dynamic attributes. The extracted structural parameters is then impinged in the finite element model for computing the dynamic characteristics. The obtained structural parameters is utilized for the identification of the location as well for examine how severe is the structural damage. In [115], the authors illustrated a particle swarm optimization-based methodology in order to train the NN (NN-PSO). The suggested approach is well capable of handling the problem of predicting failures in structure associated with multistoried reinforced concrete buildings. The superiority of the methodology is validated with the help of experimental results. In [116], the researchers developed a damage detection technique that uses 3 alternate neural networks. In order to analyze the model, a tiny dataset having ground-level image data of damaged buildings was extracted. Experimental results demonstrated the efficiency of the proposed method. In [117], the authors proposed a novel approach for the detection of structural damage as well as estimation utilizing incomplete noisy modal data and artificial neural network (ANN). A feed-forward back propagation network is suggested for the estimation of the structural damage location and severity. In [118], the author illustrated an innovative, fast and precise identification of damages in structure using 1D Convolutional Neural Network. The illustrated technique performs vibration based damage identification as well as localization of the damage in real-time and generates superior outcomes.

5.4. Hybrid Techniques

Fuzzy logic, artificial neural network and genetic algorithm are the three main important categories of intelligent techniques that tries to implement the perception of human intelligence using different methodologies. There is positive and negative attributes of each methodology that differentiates among each other. The positivity of neural network is that it can learn accurately from data and establish highly nonlinear mapping among various platforms but it is concealed to the users and it has very low reasoning capabilities. This drawbacks can be compensated by the techniques of IF-THEN rules and logical inference from the concept of fuzzy logic. The linguistic models is the main approach of fuzzy logic that uses the same decisions making capabilities of human beings as in real life. Linguistic models inherits high reasoning and nonlinear mapping capabilities. But the constraint embedded to fuzzy systems is that it cannot cope up with the ability of learning and cannot shape up according to a new ambient. The information processing systems like classifiers and controllers are the main objectives of artificial neural network and fuzzy logic unlike genetic algorithm. The genetic algorithm has a higher degree of parallel searching capabilities that enables it to search multi-peak spaces. Another attributes of genetic algorithm is that it is not dependent on continuity of the parameter space. Apart from the positivity of genetic algorithm, the drawback of genetic algorithm is that it cannot be used alone as modeling techniques. It is very much efficient when it is made to work in parallel with other modeling techniques. The combination of all three intelligent concepts will yield maximum favorable outcome and vastly accepted as a hybrid methodologies [119]. A learning technique at par with the neural network is developed by [120] which is relied on the genetic fuzzy combination. The approach is to facilitate training to the radial basis function associated with the neural network. The unique approach is a blend of genetic algorithm with

fuzzy logic in order to tweak the centers as well as widths linked to the RBFNN. By utilizing the technique of linear least square method, the optimization of neural network connection weights is achieved. The output linked to the RBF fuzzy neural network is stated as:

$$Y_{pi} = \sum_{k=1}^{36} W_{ik}\overline{\alpha}_{pk}, \ i{=}1,2 \tag{9}$$

The matrix α depicts the firing strength of all rules. Also, W is illustrated as the output linked weight matrix. Again, Y is considered to be output matrix. The equation depicting the fitness function with genetic algorithm is:

$$F_{i=}\begin{cases}\bar{J} + 2\sigma - J_i, if \ J_i{<}\bar{J} + 2\sigma, i = 1,2,\dots 200 \\ 0, otherwise\end{cases} \tag{10}$$

The mean value associate linked to the objective function is defined by $\bar{J}$. Also, σ is considered to be standard deviation and J_i is the parameter associated with objective function. A hybrid neuro-genetic algorithm for evaluating the damages in the structure is presented by [57]. Current investigation reveals that two hidden layer network was chose taking into consideration back propagation algorithm with momentum coefficients as training algorithm. The network parameters in this case which are number of neurons associated to hidden layer, learning as well as momentum coefficients are taken as variables. Also, genetic algorithm is implemented for choosing the appropriate sets of the parameter. Superiority of the developed algorithm is proved by implementing dual structures consisting of a frame and a beam. The equation representing the objective function linked to the optimization problem is:

$$E(X_k, W) = \frac{1}{p}\sum_{k=1}^{p}\sum_{m=1}^{N_n}(y_{k,m} - o_{k,m})^2 \tag{11}$$

the desired and obtained outputs for the mth output and k^{th} training instance are $y_{k,m}$ and $o_{k,m}$ respectively. In [121], the authors presented a damage detection algorithm on the basis of neuro fuzzy hybrid system for the study of location as well as severity predictions of cracks associated to beam-like structures. In [122], the authors proposed an appropriate fuzzy inference system (FIS) which is also termed as an Adaptive Neuro-Fuzzy Inference System for the detection of damage level in roads. The extracted results display the effectiveness and preciseness of the Neuro-Fuzzy systems for road damage detection. In [123], the authors developed a semi-automated Mamdani based fuzzy decision-making system for the identification of damages in buildings by the utilization of textural features. For the improvisation of the level of automation in Mamdani fuzzy system, a genetic algorithm was utilized to find its optimized parameters associated with membership functions.

6. Conclusions

Structural health monitoring is a wide are of research and has different intentions depending on the various scenario. But approaches that need to be implemented have common and specific component classes at the levels mentioned as follows: a) Need for sensors, b) storage of data, c) data transmission technique, d) database management system in order to carry out feature extraction, e) development of model from the extracted data, f) past data analysis for knowledge and g) steps related to decision making on the basis of features that is identified in

combination of known models. The design state of structural health monitoring is an important phase where consideration needs to be implemented in order to monitor the substructures related to uncertain or critical performance [124]. This critical survey revealed various positive attributes and capabilities of intelligent technique in the detection damaged structures associated to civil engineering. The genetic algorithm can generate good approximate solutions related to complex optimization problems. Fuzzy logic displayed its capabilities on handling the structural systems associated with complexities. The fuzzy methodology can handle uncertainty with ease. The neural network established its identity as it has above well capabilities in detecting cracks in structures. It also shows superior performance in detecting damage index of cases associated to the free decay of structures. It is to be noted down that hybrid techniques are more effective than other conventional artificial intelligent methods in structures as it induces the quality of individual intelligent techniques in a single methodology. The intelligent technique is considered to be the most popular tools for damage identification due to its precise, reliable and low-cost solution nature. The main intention of this survey paper is to provide an in-depth knowledge to the researcher about the application of various intelligent approaches in the area structural health monitoring. This research will impart a vast knowledge of intelligent tools among the researchers who wants to conduct analysis and monitoring of structures under various conditions. This survey paper will generate a brief idea among the researchers, the several ways of initiating research in the area of structural health monitoring.

References

[1] Richardson, M.H. Detection of damage in structure from changes in their dynamic (modal) properties---a survey. 1980; NUREG/CR-1431,Washington, DC: US Nuclear Regulatory Commission.

[2] Rytter, A. Vibration based inspection of civil engineering structures. PhD Dissertation. 1993; Department of Building Technology and Structural Engineering, Aalborg University, Denmark.

[3] Chang, P.C.;Flatau, A.; Liu, S.C. Review paper: health monitoring of civil infrastructure. Struct Health Monit, 2003; Volume 2, Issue No. 3, pp.257--67.

[4] Doebling, S.W.; Farrar, C.R.; Prime, M.B.; Shevitz, D.W. Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: a literature review. 1996; Report LA-13070-MS, Los AlamosNationalLaboratory, Los Alamos, NM.

[5] Ubertini, F., Comanducci, G., Cavalagli, N., Vibration-based structural health monitoring of a historic bell-tower using output-only measurements and multivariate statistical analysis. 2016; Volume 15, Issue No.4, pp.438-457.

[6] Bhuiyan, M. Z. A., Wang, G., Wu, J., Cao, J., Liu, X., Wang, T., Dependable Structural Health Monitoring Using Wireless Sensor Networks, 2017; IEEE Transactions on Dependable and Secure Computing, Volume 14, Issue No. 4, pp. 363-376.

[7] Carden,E.P.;Fanning,P.Vibration based condition monitoring: A review, Structural Health Monitoring,2004; Volume 3, Issue No. 4, pp.355--377.

[8] Auweraer, H.V. International research projects on structural damage detection, Damage Assessment of Structures Key Engineering Materials, 2001; Volume 204, Issue No. 2, pp.97—112.

[9] Shi, Z.Y.;Law, S.S.; Zhang, L.M. Structural damage detection from modal strain energy change, Journal of Engineering Mechanics-ASCE , 2000; Volume 126, Issue No.12,pp.1216--1223.

[10] Gawronski, W.; Sawicki, J.T. Structural damage detection using modal norms, Journal of Sound and Vibration, 2000; Volume 220, Issue No. 1, pp.194-198.

[11] Rytter, A. Vibration based inspection of civil engineering structures, PhD Dissertation.Department of Building Technology and Structural Engineering, Aalborg University, Denmark; 1993.

[12] Tsou, P.; Shen, M.H.H. Structural damage detection and identification using neural networks, AIAA J, 1994; Volume 32, pp. 176-183.

[13] Mares, C.; Surace, C. An application of genetic algorithms to identify damage in elastic structures, J Sound Vib,1996; Volume 195, pp. 195-215.

[14] Fuat, E.; Oguzhan, H.; Ilker, T.; Hakan, K.Optimal design of planar and space structures with genetic algorithms, Compos.Struct., 2000; Volume 75, pp. 209-224.

[15] Koh, C.; Chen, Y.; Liaw, C. A hybrid computational strategy for identification of structural parameters, Computers and Structures, 2003; Volume 81, Issue No. 2, pp. 107-117.

[16] Friswell, M. I.; Penny, J. E. T.; Garvey, S. D. A combined genetic and eigensensitivity algorithm for the location of damage in structures, Compos.Struct., 1998; Volume 69, pp.547-556.

[17] Chou, J.; Ghaboussi, J. Genetic algorithm in structural damage detection, Compos.Struct., 2001; Volume 79, pp. 1335-1353.

[18] Chen, D.; Wang, W. J.Classification of wavelet map patterns using multi-layer neural networks for gear fault detection, Mechanical Systems and Signal Processing, 2002; Volume 16, Issue No. 4, pp. 695-704.

[19] Garg, A. K.; Roy Mahapatra, D.; Suresh, S.; Gopalakrishnan, S.; Omkar, S. N.Estimation of composite damage model parameters using spectral finite element and neural network, Composites Science and Technology, Volume 64, Issue No. 16, pp. 2477-2493.

[20] Ko, J. M.; Sun, Z. G.; Ni, Y. Q. Multi-stage identification scheme for detecting damage in cable-stayed KapShuiMun Bridge, Engineering Structures, 2002; Volume 24, Issue No. 7, pp.857-868.

[21] Gao, J.Introduction to intelligent information processing. Beijing: China Machine Press, 2004.

[22] Pedrycz, W.Computational intelligence. Boca Raton, FL: CRC Press, 1997.

[23] Ganguli, R.A fuzzy logic system for ground based structural health monitoring of a helicopter rotor using modal data, Journal of Intelligent Material Systems and Structures, Volume 12, Issue No. 6, pp. 397-407.

[24] Silva, S.D.; Dias, M.; Lopes, V.; Brennan, M.J.Structural damage detection by fuzzy clustering, Mechanical Systems and Signal Processing, 2008; Volume 22, Issue No. 7, pp. 1636-1649.

[25] Kong, F.; Chen, R. A combined method for triplex pump fault diagnosis based on wavelet transform, fuzzy logic and neuro-networks, Mechanical Systems and Signal Processing, 2004; Volume 18, Issue No. 1, pp. 161-168.

[26] Meyyappan, L.; Jose, M.;Dagli, C.; Silva, P.; Pottinger,H. Fuzzy-neuro system for bridge health monitoring, 22ndInternational Conference of the North American Fuzzy Information Processing Society, NAFIPS 2003, Chicago, IL, 2003; pp. 8-13.

[27] Roy, N.; Ganguli, R. Filter design using radial basis function neural network and genetic algorithm for improved operational health monitoring, Applied Soft Computing, 2006; Volume 6, pp. 154-169.

[28] Lee, Y.S.; Chung, M.J.A study on crack detection using Eigen frequency test data, Computers &Structures, 2000; Volume 77, Issue No. 3, pp.327-342.

[29] Kim, J.T.; Ryu, Y.S.; Stubbs, N. Damage identification in beam-type structures: frequency-based method vs. mode-shape based method, Engineering Structures,2003; Volume 25, Issue No. 1, pp. 57-67.

[30] Deraemaeker, A.; Reynders, E,; De Roeck, G.; Kullaa, J.Vibration-based structural health monitoring using output-only measurements under changing environment, Mechanical Systems and Signal Processing, 2008; Volume 22, Issue 1, pp. 34-56.

[31] Khoo, L.M.; Mantena, P.R.; Jadhav, P. Structural damage assessment using vibration modal analysis, Structural Health Monitoring, 2004; Volume 3, Issue No. 2, pp. 177-194.

[32] Yan, A.M.; Kerschen, G.; De Boe, P.;Golinval, J.C. Structural damage diagnosis under varying environmental conditions---Part I: A linear analysis, Mechanical Systems and Signal Processing, 2005; Volume 19, pp. 847-864.

[33] Gelman, L.The new frequency response functions for structural health monitoring, Engineering Structures, 2010; Volume 32, Issue No. 12, pp. 3994-3999.

[34] Bandara, R. P.; Chan, T. H. T.; Thambiratnam, D.P. Structural damage detection method using frequency response functions, Structural Health Monitoring, 2014; Volume 13, Issue No. 4, pp. 418-429.

[35] Curadelli, R.O.; Riera, J.D.; Ambrosini, D.; Amani, M.G. Damage detection by means of structural damping identification, Engineering Structures, 2008; Volume 30, Issue No. 12, pp. 3497-3504.

[36] Jaishi, B.; Ren, W-X. Damage detection by finite element model updating using modal flexibility residual, Journal of Sound and Vibration, 2006; Volume 290, Issues No. 1–2, pp. 369-387.

[37] Huth, O.; Feltrin, G.; Maeck, J.; Kilic, N.; Motavalli, M. Damage Identification Using Modal Data: Experiences on a Prestressed Concrete Bridge, Journal of Structural Engineering, 2005; Volume 131, Issue No. 12, pp.1898-1910.

[38] Xia, Y.; Hao, H.; Brownjohn, J. M. W.; Xia, P.-Q. Damage identification of structures with uncertain frequency and mode shape data. Earthquake Engng. Struct. Dyn., 2002; Volume 31, pp.1053–1066.

[39] Furukawa, A.; Otsuka, H.; Kiyono, J. Structural Damage Detection Method Using Uncertain Frequency Response Functions, Computer-Aided Civil and Infrastructure Engineering, 2006; Volume 21, pp.292–305.

[40] Yan, J.; Cheng, L.; Wu, Z.Y.; Yam, L.H. Development in vibration-based structural damage detection technique, Mechanical Systems and Signal Processing, 2007; Volume 21, Issue No. 5, pp. 2198-2211.

[41] Hou, Z.; Noori, M.; Amand R. St.Wavelet-Based Approach for Structural Damage Detection, Journal of Engineering Mechanics, 2000; Volume 126, Issue No. 7, pp. 677-683.

[42] Escamilla-Ambrosio, P.J.; Liu, X.; Lieven, N.A.J.; Ramírez-Cortés, J.M. Wavelet-fuzzy logic approach to structural health monitoring, Fuzzy Information Processing Society (NAFIPS), 2011; Annual Meeting of the North American, El Paso, Texas, USA.

[43] Lynch, J. P.; Loh, K. J. A Summary Review of Wireless Sensors and Sensor Networks for Structural Health Monitoring, The Shock and Vibration Digest, 2006; Volume 38, Issue No. 2, pp. 91-128.

[44] Cho, S.; Yun, C-B.; Lynch, J. P.; Zimmerman, A. T.; Spencer Jr., B. F.; Nagayama, T. Smart Wireless Sensor Technology for Structural Health Monitoring of Civil Structures, Steel Structures, 2008; Volume 8, pp. 267-275.

[45] Aoki, S.; Fujino, Y.; Abe, M. Intelligent Bridge Maintenance System Using MEMS and Network Technology, In Smart Systems and NDE for Civil Infrastructures, San Diego, CA, Proceedings of the SPIE, 2003; Volume 5057, pp. 37–42.

[46] Casciati, F.; Faravelli, L.; Borghetti, F.; Fornasari, A. Tuning the Frequency Band of a Wireless Sensor Network, In Proceedings of the 4th International Workshop on Structural Health Monitoring, Stanford, 2003; pp.1185–1192.

[47] Ou, J.; Li, H.; Yu. Development and Performance of Wireless Sensor Network for Structural Health Monitoring," Smart Structures and Materials, , Proceedings of the SPIE, 2004; Volume 5391, pp.765–773.

[48] Wang, Y.; Lynch, J. P.; Law, K. H. Wireless Structural Sensors Using Reliable Communication Protocols for Data Acquisition and Interrogation, Proceedings of the 23rd International Modal Analysis Conference, 2005.

[49] Camp, C.; Pezeshk, S.; Cao, G. Optimized Design of Two-Dimensional Structures Using a Genetic Algorithm, Journal of Structural Engineering, 1998; Volume 124, Issue No. 5, pp. 551-559.

[50] Jenkins, W. M. Towards structural optimization via the genetic algorithm, Computers & Structures, 1991; Volume 40, Issue No. 5, pp. 1321-1327.

[51] Rajeev, S.; Krishnamoorthy, C. S. Discrete optimization on structures using genetic algorithms, J. Struct. Eng., 1992; Volume 118, Issue No. 5, pp. 1233-1250.

[52] Larson, C. B.; Zimmerman, D. C. Structural model refinement using a genetic algorithm approach, Proc., 11th Int. Modal Analysis Conf., Society for Experimental Mechanics, Kissimmee, Fla., 1993; pp. 1095-1101.

[53] Yong, X.; Hong, H. A Genetic Algorithm for Structural Damage Detection Based on Vibration Data. School of Civil and Structural Engineering, Nanyang Technological University.

[54] Rao, M. A.; Srinivas, J.; Murthy, B. S. N. Damage detection in vibrating bodies using genetic algorithms, Computers and Structures, 2004; Volume 82, pp. 963-968.

[55] Tso W.K.; Zhu, T. J. Design of torsionally unbalanced structural systems based on code provisions I: ductility demands, Earthquake Engineering & Structural Dynamics, 1992; Volume 21, Issue No. 7, pp. 609–627.

[56] Krawczuk, M. Application of spectral beam finite element with a crack iterative search technique for damage detection, Finite Element Anal. Des., 2002; Volume 38, pp. 537-548.

[57] Sahoo, B.; Maity, D. Damage assessment of structures using hybrid neuro-genetic algorithm, Appl. Soft Comput., 2007; Volume 7, pp. 89-104.

[58] Vakil-Bagmisheh, M-T.; Peimani, M.; Sadeghi, H.; Ettefagh, M. M. Crack detection in beam-like structures using genetic algorithms, Applied Soft Computing, 2008; Volume 8, pp. 1150-1160.

[59] Au, F. T. K.; Cheng, Y. S.; Tham, L. G.; Bai, Z. Z. Structural damage detection based on a micro-genetic algorithm using incomplete and noisy modal test data, , Journal of Sound and Vibration, Volume 259, Issue No. 5, pp. 1081-1094.

[60] Ruotolo R.; Surace C.Damage assessment of multiple cracked beams: numerical results and experimental validation, Journal of Sound and Vibration, 1997; Volume 206, Issue No. 4, pp. 567-588.

[61] Khatir, S.; Belaidi, I.; Serra, R.; Benaissa, B.; Saada, A. A.Genetic Algorithm Based Objective Functions Comparative Study for Damage Detection and Localization in Beam Structures, Journal of Physics: Conference Series, 2015; Volume 628.

[62] Perera, R.; Torres, R. Structural Damage Detection via Modal Data with Genetic Algorithms, Journal of Structural Engineering, 2006; Volume 132, Issue No. 9, pp. 1491-1501.

[63] Nobahari, M.; Seyedpoor, S.M. Structural damage detection using an efficient correlation-based index and a modified genetic algorithm, Mathematical and Computer Modelling, 2011; Volume 53, pp. 1798-1809.

[64] Gomes, H.M.; Silva, N.R.S.Some comparisons for damage detection on structures using genetic algorithms and modal sensitivity method, 2008; Volume 32, pp. 2216-2232.

[65] 65. Perera, R.; Manzano, C. Performance assessment of multicriteria damage identification genetic algorithms, Computers & Structures, 2009; Volume 87, Issue No. 1-2, pp. 120-127.

[66] 66. Hao, H.; Asce, M.; Xia, Y.Vibration-based damage detection of structures by genetic algorithm, Journal of Computing in Civil Engineering, 2002; Volume 16, Issue No. 3, pp. 222-229.

[67] He,R-S.; Hwang, S-F. Damage detection by an adaptive real-parameter simulated annealing genetic algorithm, Computers and Structures, 2006; Volume 84, Issue No. 31-32, pp. 2231-2243.

[68] Borges, C.; Barbosa,H.;Lemonge,A.A structural damage identification method based on genetic algorithm and vibrational data, International Journal for Numerical Methods in Engineering, 2007; Volume 69, pp. 2663-2686.

[69] Kouchmeshky, B.; Aquino, W.;Bongard,J.; Lipson, H. Co-evolutionary algorithm for structural damage identification using minimal physical testing, International Journal for Numerical Methods in Engineering, 2007; Volume 69, Issue No. 5, pp. 1085-1107.

[70] Meruane, V.; Heylen, W. Structural damage assessment under varying temperature conditions, Structural Health Monitoring, 2012, Volume 11, Issue No. 3, pp. 345-357.

[71] Kuang, K. S. C.;Maalej, M.; Quek, S. T.An Application of a Plastic Optical Fiber Sensor and Genetic Algorithm for Structural Health Monitoring, Journal of Intelligent Material Systems and Structures, 2006; Volume 17, Issue No. 5, pp. 361-379.

[72] Nag,A.;Mahapatra, D. R.;Gopalakrishnan, S. Identification of delamination in composite beams using spectral estimation and a genetic algorithm, Smart Mater. Struct., 2002; Volume 11, pp. 899-908.

[73] Khatir, S.; Belaidi, I,; Khatir, T.; Hamrani, A.; Zhou, YL.; Wahab, MA. Multiple damage detection in unidirectional graphite-epoxy composite beams using particle swarm optimization and genetic algorithm, MECHANIKA, 2017; Volume 23, Issue No. 4, pp.514-521.

[74] Santos, A.; Figueiredo, E.; Silva M.; Santos R.; Sales C.; Costa, J. C. W. A. Genetic-based EM algorithm to improve the robustness of Gaussian mixture models for damage detection in bridges, Struct. Control Health Monit., 2017; Volume 24, Issue No. 3.

[75] Cha, Y.-J.; Buyukozturk, O. Structural Damage Detection Using Modal Strain Energy and Hybrid Multiobjective Optimization, Computer-Aided Civil and Infrastructure Engineering, 2015; Volume 30,pp. 347–358.

[76] Zadeh, L. A. Fuzzy sets, Information and Control, 1965; Volume 8, Issue No. 3, pp. 338-353.

[77] Hajek, P.; Paris, J.; Shepherdson, J. The liar paradox and fuzzy logic, J. Symb. Logic, 2000; Volume 65, pp. 339-346.

[78] Raju, K.; Majumdar, A.Fuzzy functional dependencies and lossless join decomposition of fuzzy relational database systems, ACM Trans. Database Syst., 1988; Volume 13, Issue No. 32, pp. 129-166.

[79] Hajek, P.; Havranek, T.; Jirousek, R.Processing Uncertain Information in Expert Systems, CRC Press, Boca Raton, 1992.

[80] Juuso, E. K. Intelligent systems design with linguistic equations, 9th Workshop Fuzzy Control des GMA-FA, Dortmund, Germany, 1999; pp. 177-196.

[81] Liu, X.; Ma, L.; Mathew, J. Machinery fault diagnosis based on fuzzy measure and fuzzy integral data fusion techniques, Mechanical Systems and Signal Processing, 2009; Volume 23, Issue No. 3, pp. 690-700.

[82] Rodríguez, P. V. J.; Arkkio, A.Detection of stator winding fault in induction motor using fuzzy logic, Applied Soft Computing, 2008; Volume 8, Issue No. 2, pp. 1112-1120.

[83] Sazonov, E. S.; Klinkhachorn, P.; Gangarao, H. V. S.; Halabe, U. B. Fuzzy Logic Expert System for Automated Damage Detection from Changes in Strain Energy Mode Shapes, Nondestructive Testing and Evaluation, 2002; Volume 18, Issue No. 1, pp. 1-20.

[84] Aljoumaa, H.; Soffker, D. Condition Monitoring and Classification Approach based on Fuzzy-Filtering, Proceedings of the World Congress on Engineering and Computer Science, San Francisco, USA, 2010.

[85] Reda Taha, M. M.; Lucero, J. Damage identification for structural health monitoring using fuzzy pattern recognition,Engineering Structures, 2005; Volume 27, Issue No. 12, pp. 1774-1783.

[86] Worden, K.; Delieu-Barton, J. M.An overview of intelligent fault detection in systems and structures, Structural Health Monitoring, 2004; Volume 3, pp. 85-98.

[87] Altunok, E.; Reda Taha, M. M.;Epp, D. S.; Mayes, R. L.; Baca, T. J. Damage Pattern Recognition for Structural Health Monitoring Using Fuzzy Similarity Prescription, Computer-Aided Civil and Infrastructure Engineering, 2006; Volume 21, pp. 549-560.

[88] Ettefagh, M.M.; Sadeghi, M.H.;Khanmohammadi, S.Structural Damage Detection, Using Fuzzy Classification and ARMA Parametric Modeling, Mech. & Aerospace Eng. J., 2007; Volume 3, Issue No. 2, pp. 85-98.

[89] Yu, L.; Zhu, J-H.; Yu, L-L. Structural Damage Detection in a Truss Bridge Model Using Fuzzy Clustering and Measured FRF Data Reduced by Principal Component Projection, Advances in Structural Engineering, 2013; Volume 16, Issue No. 1, pp. 207-217.

[90] Jiang, S-F.; Zhang, C-M.; Zhang, S. Two-stage structural damage detection using fuzzy neural networks and data fusion techniques, Expert Systems with Applications, 2011; Volume 38, pp. 511-519.

[91] Beena, P.; Ganguli, R. Structural damage detection using fuzzy cognitive maps and Hebbian learning, Applied Soft Computing, 2011; Volume 11, pp. 1014-1020.

[92] Liu, Z.; Zhou, Q; Chi, Q.; Zhang, Y.; Chen, Y.; Qi, S. Structural Damage Detection Based on Semi-Supervised Fuzzy C-means Clustering, The 9th International Conference on Computer Science & Education (ICCSE 2014), 2014.

[93] Irwansyah,E.; Hartati, S.; Hartono. Three-stage Fuzzy Rule-Based Model for Earthquake Non-Engineered Building House Damage Hazard Determination, JACIII, 2017; Volume 21, Issue No. 7, pp. 1298-1311.

[94] Janalipour, M.; Mohammadzadeh, A. Building Damage Detection Using Object-Based Image Analysis and ANFIS From High-Resolution Image (Case Study: BAM Earthquake, Iran), IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2016; Volume 9, Issue No. 5, pp. 1937-1945.

[95] Ye, X.; Qin, Q.; Liu, M.; Wang, J.; Wang, J. Building damage detection from post-quake remote sensing image based on fuzzy reasoning, IEEE Geoscience and Remote Sensing Symposium, Quebec City, QC, 2014; pp. 529-532.

[96] Rumelhart,D. E. .; McClelland, J. L.; the PDP Research Group, Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume I: Foundations, MIT Press, Cambridge, MA, 1986.

[97] Wu, X. .; Ghaboussi, J.; Garrett Jr., J.H. Use of neural networks in detection of structural damage, Computers & Structures, 1992; Volume 42, Issue No. 4, pp. 649-659.

[98] Zhao, J.; Ivan, J. N.; DeWolf, J. T. Structural Damage Detection using Artificial Neural Networks, Journal of Infrastructure Systems, 1998; Volume 4, Issue No. 3, pp. 93-101.

[99] Rhim, J.; Lee, S.W. A neural network approach for damage detection and identification of structures, Computational Mechanics, 1995; Volume 16, Issue No. 6, pp. 437-443.

[100] Pandey, P. C.; Barai, S. V. Multilayer Perceptron in Damage Detection of Bridge Structures, Computers & Structures, 1995; Volume 54, Issue No. 4, pp. 597-608.

[101] Elkordy, M. F.; Chang, K. C.; Lee, G. C. Neural networks trained by analytically simulated damage states, Journal of Computers in Civil Engineering, 1993; Volume 7, Issue No. 2, pp. 130-145.

[102] Hung, S-L.; Kao, C. Y. Structural damage detection using the optimal weights of the approximating artificial neural networks, Earthquake Engineering & Structural Dynamics, 2002; Volume 31, pp. 217-234.

[103] Wu, Z.; Xu, B.; Yokoyama, K. Decentralized Parametric Damage Detection Based on Neural Networks, Computer-Aided Civil and Infrastructure Engineering, 2002; Volume 17, pp. 175-184.

[104] Suresh, S.; Omkar, S. N.; Ganguli, R.; Mani, V. Identification of crack location and depth in a cantilever beam using a modular neural network approach, Smart Materials and Structures, 2004; Volume 13, pp. 907-915.

[105] Sahin, M.; Shenoi, R.A. Quantification and localization of damage in beam-like structures by using artificial neural networks with experimental validation, Engineering Structures, 2003; Volume 25, pp. 1785-1802.

[106] Lee, J. J.; Lee, J. W.; Yi, J. H.; Yun, C. B.; Jung, H. Y. Neural networks-based damage detection for bridges considering errors in baseline finite element models, Journal of Sound and Vibration, 2005; Volume 280, pp. 555-578.

[107] Mehrjoo, M.; Khaji, N.; Moharrami, H.; Bahreininejad, A. Damage detection of truss bridge joints using Artificial Neural Networks, Expert Systems with Applications, 2008; Volume 35, pp. 1122-1131.

[108] Kao, C. Y.; Hung, S-L. Detection of structural damage via free vibration responses generated by approximating artificial neural networks, Computers and Structures, 2003; Volume 81, pp. 2631-2644.

[109] Bakhary, N.; Hao, H.; Deeks, A. J. Structure Damage Detection Using Neural Network with Multi-Stage Substructuring, Advances in Structural Engineering, 2010; Volume 13, Issue No. 1, pp. 95-110.

[110] Yeung, W.T.; Smith, J.W. Damage detection in bridges using neural networks for pattern recognition of vibration signatures, Engineering Structures, 2005; Volume 27, pp. 685-698.

[111] 111. Nazarko, P.; Ziemiański, L. Application of artificial neural networks in the damage identification of structural elements, Computer Assisted Mechanics and Engineering Sciences, 2011; Volume 18, pp. 175-189.

[112] Haryanto, I.; Setiawan, D.; Budiyono, A. Structural Damage Detection Using Randomized Trained Neural Networks, Intelligent Unmanned Systems: Theory and Applications, 2009; Volume 192, pp. 245-255.

[113] Bakhary, N.; Hao, H.; Deeks, A. J. Damage detection using artificial neural network with consideration of uncertainties, Engineering Structures, 2007; Volume 29, pp. 2806-2815.

[114] Chang, C. C.; Chang, T. Y. P.; Xu, Y. G. Structural Damage Detection Using an Iterative Neural Network, Journal of Intelligent Material Systems and Structures, 2000; Volume 11, pp. 32-42.

[115] Chatterjee, S.; Sarkar, S.; Hore, S.; Dey, N.; Ashour, A.S.; Balas, V.E. Particle swarm optimization trained neural network for structural failure prediction of multistoried RC buildings. Neural Comput. Appl., 2017; Volume 28, Issue No 8.

[116] Nia, K. R.; Mori, G. Building Damage Assessment Using Deep Learning and Ground-Level Image Data, 14th Conference on Computer and Robot Vision (CRV), 2017; pp.95-102.

[117] Kourehli, S.S. Damage Assessment in Structures Using Incomplete Modal Data and Artificial Neural Network, International Journal of Structural Stability and Dynamics, 2015; 17 pages.

[118] Abdeljaber, O.; Avci, O.; Kiranyaz, S.; Gabbouj, M.; Inman, D. J. Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks, Journal of Sound and Vibration, 2017; Volume 388, pp. 154-170.

[119] Hefny, H. A.; Bahnasawi, A. A.; Abdel Wahab, A. H.; Shaheen, S. I. Logical radial basis function networks a hybrid intelligent model for function approximation, Advances in Engineering Software, 1999; Volume 30, pp. 407-417.

[120] Zheng, S-J.; Li, Z-Q.; Wang, H-T.; A genetic fuzzy radial basis function neural network for structural health monitoring of composite laminated beams, Expert Systems with Applications, 2011; Volume 38, pp. 11837-11842.

[121] Aydin, K; Kisi, O. Damage detection in structural beam elements using hybrid neuro fuzzy systems, Smart Structures and Systems, 2015; Volume 16, Issue No. 6, pp.1107-1132.

[122] Izadi, M.; Mohammadzadeh, A.; Haghighattalab, A. A New Neuro-Fuzzy Approach for Post-earthquake Road Damage Assessment Using GA and SVM Classification from QuickBird Satellite Images, Journal of the Indian Society of Remote Sensing, 2017; pp. 1-13.

[123] Janalipour, M.; Mohammadzadeh, A. A Fuzzy-GA Based Decision Making System for Detecting Damaged Buildings from High-Spatial Resolution Optical Images, Remote Sens. 2017; Volume 9, Issue No. 4.

[124] Brownjohn, J. M. W. Structural health monitoring of civil infrastructure, Phil. Trans. R. Soc. A, 2007; Volume 365, pp. 589-622.

Modified HOG Descriptor-Based Banknote Recognition System

Tamarafinide Victory Dittimi[*], Ching Yee Suen

Center for Pattern Recognition and Machine Intelligence, Concordia University, Montreal, Canada

A R T I C L E   I N F O

A B S T R A C T

This research presents a banknote recognition and counterfeit detection system that proposes a pre-processing approach before creating a new feature set by extracting Speeded-Up Robust Feature, creating image patches from its vertices and then computing the Histogram of Gradient descriptors from cells within the patch boundaries. A feature reduction approach is also designed based on a high correlation and Low variance filter to reduce the feature set by calculating the intra-feature value variability to ascertain the similarity between the elements to prune out highly correlated values and variances using a fixed threshold. For experimental evaluation, additional analysis was conducted using Principal Component Analysis, Bag of words, Low Variance filter, and High Correlation Filter method. Furthermore, the classification was performed using 10-fold cross-validation based One-Versus-One Support Vector Machine. Additionally, the developed technique was tested on counterfeit banknotes to authenticate genuine and fake currencies. Lastly, point matching method was applied to detect our suggested feature vectors to match images containing multiple notes to distinguish and determine the value of the bills.

1. Introduction

Technological advancement has made life more comfortable, but it also has its downside as criminals have taken advantage of this progression. Initially, only printing houses had access to the technology needed to produce banknotes but today with the aid of a personal computer and a good laser printer people can duplicate paper money easily in their residence. Banknote counterfeiting is one of the core crimes on earth. The Apex bank of every country is responsible for providing the security features and in Nigeria it is the responsibility of the Central Bank of Nigeria (CBN). Every Nigerian banknote contains several security features that can be used to identify and recognize different note denominations [1].

Over the years, differentiating between fake and real currency banknotes has become quite tricky due to the fast-scientific progress in color printing, cloning, and imaging and because fake notes are now reproduced via top-notch technology that uses security paper [2]. In the same view, counterfeit banknote detection has become an essential task in billing machines. Its implemented by using image analysis methods (transmittance analysis, reflectance analysis) with different light spectra (visible, infrared and ultraviolet). The drawback of this system is its high cost and its usage with only Automated Teller Machines (ATM), where an elevated level of reliability is required. Anti-

counterfeiting banknotes security features can be classified into three categories: printing, ink and substrate features. Printing features like intaglio printing, micro letters, serial number, signatures, Guilloches, Ink features such as UV glowing ink, pearl ink, infrared ink, color-shifting inks, and substrate features like planchets, security thread, see-through registers, and watermark [3].

This research presents a paper money recognition system that includes multiple banknote identifications and Counterfeit note detection that could aid in reducing banknote counterfeiting offenses and positively influence a country's economy. The rest of the article is divided into the following sections: Section 2 discusses the research background and surveyed literature. In Section 3, the proposed system including its architecture, design, and stages of the approach is studied. Section 4 debates the analysis and discussion of the results of the system functionality, and the conclusion of the research and future work is given in Section 5.

2. Background & Literature Survey

Banknotes are utilized in every business venture in as much as electronic currencies is becoming identified as a substitute legal tender, the note is still dominant and central in the day to day dealings. Currency Recognition and detection is a critical problem in computer vision. Although, electronic and online currencies exist banknote is still the preferred option in daily transactions such

[*]Tamarafinide Dittimi, Concordia University, t_dittim@encs.concordia.ca

as vending machine, the banking system, banks, shopping mall, railway ticket counter, ATM, vending machines, and currency exchange service. Furthermore, there are cases of money bundles withdrawn from ATMs contained a few counterfeit bills although banks are unwilling to accept these omissions. Even though this problem is majorly unavoidable, they are not also able to provide all branches with the needed counterfeit detection system [4].

One of the most straightforward methods that have been applied in paper money recognition system is the use of visible features found in notes like the size, texture, and color of banknotes. All banknotes have pictures and patterns impressed on them, such as portraits of historical persons or images of landmark buildings which are characterized as textures extracted as features using texture-modeling techniques. The drawback of this approach is that banknotes are mistreated so could get damaged and torn due to regular use leading to alteration of the notes thereby hurting the recognition rate [5]. On the other hand, [6] extracted the geometric features (length and width) and employed it in classifying bills with varying sizes such as Italy and Chinese banknotes. Although, this approach was simple and fast; it had significant limitations including its effectiveness for slant, stain, and tear. Furthermore, it was unsuitable for some banknote such as the US Dollars, which had several denominations of the same size.

Furthermore, many countries employ colors for differentiating banknotes although they could wear or tear it still retains its color [7]. [8] designed a system that identified the denomination of the Indian currency note based on its color. Likewise, another author also developed a technique that converted the image from RGB color space to HSV color space and then extracted the color feature. Their justification for employing the HSV color space was because of its nearness to the human conceptual understanding of color [9]. Additionally, a researcher recommended an approach based on characteristic point extraction of a unique part of the Euro paper currency that possesses the same color has also been conducted [10]. Similarly, [11] devised a US banknote counterfeit detection system that extracts its color feature and classifies the currency using fuzzy Hamming distance. Despite its high performance, if the system were tested on Euro banknotes. It would not be able to detect forgeries since its detection is based only on the color feature. Most countries have different colors distributed over the surface of its banknotes including the Nigerian Naira; all have only one dominant color. For instance, the dominant color of 10 Naira note is "Red." Although, some banknotes have the same dominant color due to wear and tear the color quality may degrade; hence leading to difficulties in using the color feature for automatic recognition. Also, using only the color as a discriminative feature may fail because some banknote denomination does have the same dominant color [12].

Likewise, [13] proposed a banknote classification technique based on discrete real-valued wavelet transform (DWT), they calculated the feature vector using the statistical properties of decomposed coefficients from the approximation and detail sub-bands. The technique was applied successfully in spatial and frequency domain, although it does not shift invariant and has no phase information. Another researcher also devised a new method called a hyper-complex wavelet transform (CWT) that is shift invariant and has a single phase. However, it also had its limitation like its ambiguity when used for 2-D signals [14]. Furthermore, an invariant shift extension of DWT and CWT called the quaternion wavelet transform (QWT) was also adopted. The technique had three phases and one magnitude, the first two stages were on the local image shifts, and the third focuses on image texture information [15]. [16] also suggested a new feature extraction system using the quaternion wavelet transform (QWT) that combined the partial and total Hilbert Transform (HT) using quaternion algebra to form one magnitude and three phases. It captures the statistical characteristics of QWT coefficients using the generalized Gaussian density (GGD) and classifies the banknote using the neural network. Furthermore, [17] presented a banknote recognition system that used a wavelet transform for extracting features and classified the features using Multi-layer perceptron (MLP) neural network; the method experimented on Persian paper currency. In another research work, a feature extraction technique was developed based on the Discrete Wavelet Transform (DWT). They employed Daubechies 1 (DB1), and the approximate coefficient matrix of the transformed image is extracted from the estimated efficient model and used for recognition, classification, and retrieval of currency notes [18].

Additionally, [19] formulated an approach that uses the edge and corners of an object in detection, the system was invariant to image rotation but was not able to handle issues associated with image scale. [20] then presented another technique that could automatically determine the scaling of the objects, it computed the Hessian matrix and used the Laplacian operator to locate the block like feature points in the image. Similarly, [21] also conducted further research on automatic scaling selection and implemented a Harris-Laplace detector in addition to the Hessian-Laplace to improve the automatic scale selection process. [22] also developed a SIFT method that finds the limits of the scale space in an object. The technique was invariant to scale and rotation and was very useful in extracting interest point descriptors which can be utilized in matching and classifying objects. [23] devised a method for the classification of Slovak banknotes based on SIFT detector and descriptors. Although, the approach posed its issues which were its computational complexity leading to high processing time limiting its acceptability in real-time applications.

In the same view, [24] also developed a colored SIFT approach that combines the color feature and the geometric data using the SIFT algorithm. The Mikolajczyk and Schmid framework was further applied to normalize introduced a nonlinear SVM method that juxtaposed images of different color space in addition to combining color space descriptors to evaluate the accuracy of the approach [25]. On the other hand, speeded up robust features (SURF) was created, the process was designed to improve on the SIFT method, the procedure retains the robustness of SIFT but with a decreased descriptor size and shorter processing rate and time.

SURF is also an interest point detector that is robust to scaling and rotation. It creates a unique, repetitive and vigorous process for capturing the interest points. The technique creates an integral image and uses it in place of the original image to diminish the processing time. It then uses the Hessian matrix to detect the highest and lowest second derivative value after this is achieved the interest point orientation is detected and rotated to align with the orientation of the predefined direction to complete the SURF descriptor creation [26]. A researcher applied SURF descriptor in
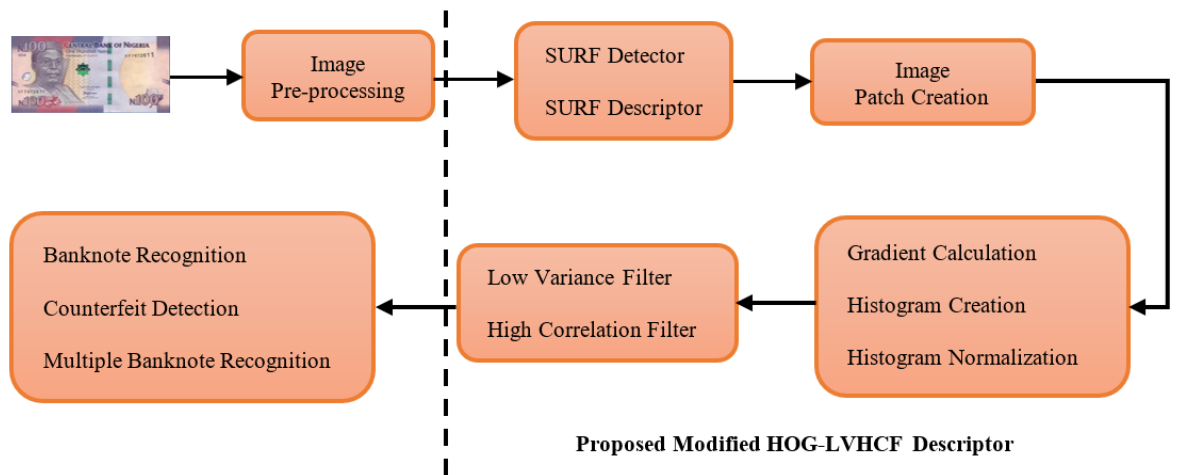
Figure: 1. Proposed Modified Hog-LVHCF based Banknote Recognition System

banknote recognition; the system was used to match reference regions with query images. It presented promising results and was robust to handle partial distortions as well as worn or wrinkled bills [27].

3. Proposed Methodology

In this research, we implemented a procedure for enhancing and cleaning up banknote images. It involves converting to grayscale then employing Gamma Correction and Histogram equalization to improve the image luminance and contrast of the intensity image. Furthermore, Weiner filter was applied to reduce the amount of noise after that the boundaries of the banknote are computed using Blob detection, cropped out and Gaussian pyramid utilized in standardizing the note. We also created a new feature set by extracting the Histogram of Gradient (HOG) Descriptors from image patches generated from the vertices of SURF interest points. After the HOG features have been obtained, the vector size may differ for each image in addition to being computationally demanding during processing. To minimize its effect on recognition/detection rate while maximizing the reduction of Feature dimensionality. A combination of dimensionality reduction techniques based on the descriptor correlation and variance is applied to reduce the Vector scale. A descriptor filter analysis using Pearson's Product Moment Coefficient ρ and variance σ^2 is employed to determine the best set of features for classification.

The approach answers two fundamental questions, determining the relevance of the descriptor to the feature set and its redundancy regarding other associated descriptors. Finally, we applied our technique to several datasets using OVO-SVM and Putative point matching to investigate the efficiency of the descriptor showing the effectiveness of Modified HOG/LVHCF descriptor for banknote identification, counterfeit detection, and multiple notes recognition. Figure 1 shows the phase of the proposed Modified Hog-LVHCF based paper money recognition system.

3.1. Image Pre-processing

During image acquisition, the irrelevant background image is also captured. This unwanted information must be removed to identify and recognize the notes. Furthermore, digitized banknote images sometimes have discontinuous particles, noise,

illumination, positioning, and unanticipated boundaries. Also, when matching the image, if the banknotes are of varied sizes it could affect the recognition rate. In this research, image pre-processing plays a fundamental step as it involves enhancing and cleaning up the images for efficient feature extraction. Thus, in this research, a new combination of techniques for image pre-processing is presented, and a flowchart of the procedure is shown in Figure 2.

Step 1: The image was first converted to grayscale by calculating the weighted sum of the three channels of the RGB image and representing the result as a matrix whose fragments are assigned values based on how dark or bright the color of the pixel is at the corresponding position. The value allocated to the pixel during grayscale conversion ranges from 0 to 255 representing the brightness of the image.

Step 2: Gamma Correction was applied to improve the image luminance by enhancing the quality of the image in dark or blurry areas while also compressing the intensity of bright regions.

Step 3: Histogram equalization was then employed in enhancing the contrast of the intensity image, the approach redistributes the image histogram to ensure that the brightness value is constant across the image.

Step 4: Weiner filter was applied to reduce the amount of noise present in a signal comparing with an estimate of the desired noiseless signal.

Step 5: Background subtraction was also achieved by first finding all blobs in the image using Blob detection, select the boundary coordinates of the blob containing the banknote and cropping out the note using the detected blob edges.

Step 6: Banknote denominations sometimes are of varied sizes, and the note can also change due to the acquisition manner thus image reduction and standardization were achieved using the Gaussian pyramid.

3.2. Feature Extraction (Proposed Modified HOG Descriptor)

Feature extraction is a crucial step in image processing, the selection of features is vital in determining the recognition of an object, and an ideal feature set should be stable, invariant to rotation, scale, and illumination changes. Several feature
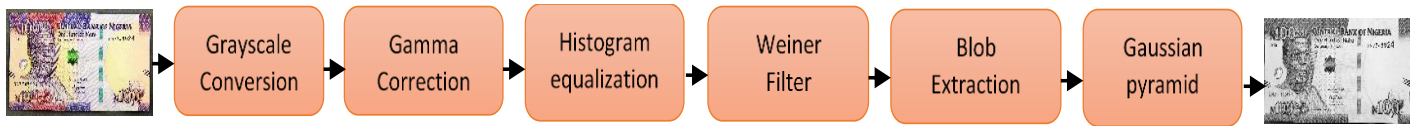
Figure 2: Block diagram of the procedures undergone in the Image Pre-processing

extraction methods have been proposed over the years. In this research, several useful descriptors and feature reduction approaches were investigated and compared to the proposed Modified HOG descriptor and Low Variance High Correlation Filter. In this study, six different descriptors were employed: Scale Invariant Feature Transform, speeded up robust features, gradient location and orientation histogram, Histogram of Gradient, shape context, and the proposed Modified HOG Descriptor. While to reduce the processing time several feature reduction methods were also implemented: Principal Component Analysis, Bag of Words, High Correlation filter, Low Variance Filter, and compared to the developed Low Variance and High Correlation Filter Algorithm. The proposed Modified HOG based descriptor technique is split into four phases: SURF point detection, SURF descriptor extraction, image patches creation from SURF points, and HOG feature computation.

3.2.1. SURF Point Detection

The SURF interest point detection is separated into four phases: integral image generation, calculation of the matrix, construction of the scale space and localization of the feature point.

Integral Image Generation: The Integral image generation was introduced to reduce the computational complexity of the square like convolution filters used in the future, it is mapped from the original image. The integral image is then computed by the sum of all pixels in the input image I within a given rectangle sector which can be deduced by Equation 1.

$$I_\Sigma(x, y) = \sum_{i=0}^{x} \sum_{j=0}^{y} I(i, j) \qquad (1)$$

After the integral image has been calculated, three additions are carried out to compute the summation of the intensities over any four-sided vertical area consequently the computational time is not dependent on the image size using Equation 2.

$$\sum = I_\Sigma(A) + I_\Sigma(D) - \left(I_\Sigma(C) + I_\Sigma(B)\right) \qquad (2)$$

Computation of the Hessian Matrix: The box filters are used as an approximation of the second-order Gaussian derivatives, this is computed on the integral image calculated earlier rather than the actual image hence reducing the processing time. A 9x9 square-like filter was employed in the computation of the Gaussian denoted by with a scale of 1.2 which led to the creation of a blob response map over different scales in the image. A set of simple weights w were applied to the square-like sector to reduce the computational overhead.

Construction of the scale space: The SURF approach is scale invariance, it applies the same filter at different sizes to the image at the same speed in parallel. The length of the dark lobe can only be increased by an even number of pixels to guarantee the

presence of a central pixel (top). The scale space is categorized as a series of octaves; an octave is a set of filter response maps gotten by convolving each input image with a filter in different sizes. For an octave layer with a size N * N, the scale σ, the convolution of each layer in the image is computed on the response map using equation 3.

$$\det(H_{approx}) = D_{xx}D_{yy} - (wD_{xy})^2 \qquad (3)$$

Localization of the interest point: The final step in the interest point detection is localizing the image over the different scales, a threshold is set for removing weak responses in the map. Several thresholds are checked until a middle ground is reached that takes into consideration the number of interest points and the strength of the detected interest point. Furthermore, a 3-dimensional scale space is applied to the non-maximum suppression, and only points bigger than all the neighbors is selected as the given interest points. Finally, scale space interpolation is applied to obtain the best interest points [28].

3.2.2. SURF Descriptor Extraction

SURF Descriptor computation is divided into three phases: Assignment of the interest point orientation, Computation of the descriptor using Haar Wavelet transform and Feature matching using Fast Indexing.

Assignment of the interest point orientation: Assignment of Interest points orientation is needed to attain rotation invariance, so Haar Wavelet transform is used at each interest point (x, y) to determine the direction of a circular neighborhood in addition to its scale. After computing the wavelet transform the feature point is weighted using a Gaussian filter, the dominant orientation is decided by calculating the summation of all the horizontal and vertical responses within the sliding windows; the most extended vector is then selected as the local orientation vector for the feature point.

Computation of the descriptor using Haar Wavelet transforms: The previous step used a box area centered around the point of interest in determining the orientation. In this phase, each sector is further divided into a 4x4 square like sector, to improve the reduced the geometric distortions, and localization inaccuracies that could occur the interest pointed are first weighted with a Gaussian filter then the Haar wavelet transform is calculated horizontally and vertically for each 5x5 paced interest points. Furthermore, the computed wavelet transform is then summed for each sector and use as the first level entry in forming the feature vector set. Moreover, to capture the change in intensity after each interest point the absolute value of the responses is also computed to create a 4-dimensional descriptor vector for each sub-sector. Finally, the sectors which are made up of a 4x4 sub-sector is concatenated resulting in a descriptor with a length of 64 dimensions.

Feature matching using Fast Indexing: This phase presents an additional functionality to the matching stage, it uses a fast indexing technique based on the previous computed Laplacian in the Haar wavelet transform stage. The Laplacian value can be used to differentiate lighter blobs on darker areas from darker blobs in brighter areas. Hence, used in removing features that do not share the same type of contrast thus reducing unstable interest points from the feature vector set thus reducing the processing time and as the Laplacian has already been calculated earlier it to does not add any extra computation overhead. Figure 3 presents SURF point descriptors of sample USD and CAD paper money.

a)   b)
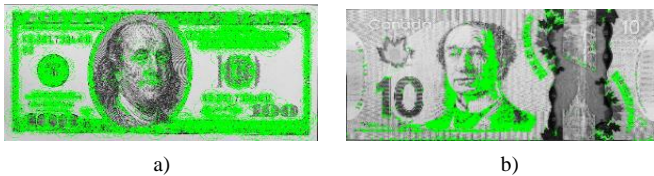
Figure 3. Detected SURF descriptors of a) the US, and b) Canadian Dollar

3.2.3 Creation of Image Patches

An image patch is designed from polygons defined by the coordinates of its vertices; it consists of one or extra polygons that might or might not be connected. In this research, the 43 x 43 patch is formed by specifying the coordinates of each unique vertex of the SURF square-like descriptors and a matrix that defines how to connect these vertices to create the faces. This technique is preferred for SURF points as its patches multifaceted, so vertices shared by more than one face need be defined only once thereby requiring fewer data to design the patches.

3.2.4 HOG Descriptor Computation

Histogram of Gradient (HOG) feature is a viable feature extraction technique with the ability to select feature points irrespective of the banknote image view. It is based on the histogram of pixel gradients neighbors for image sections or blocks thus it is hardly affected by changes in lighting conditions or image geometry. The HOG feature extraction technique can be grouped into three main stages: Calculation of Gradient for each pixel, computing its Histogram and by combining the histogram of all cells in a block to normalizing the histogram [29].

a)   b)
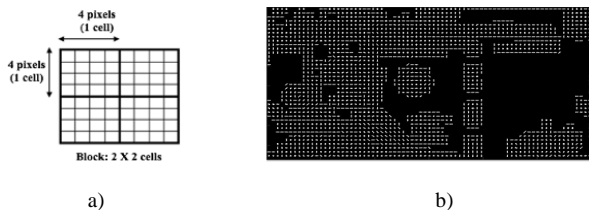
Figure 4. (a) Cell and block module (b) Sample HOG on Naira note

The feature depicts the gradient dispersion on the different positioning of a banknote image and is conducted to obtain the dimension and texture feature of the banknote. The HOG algorithm partitions all note image into small evenly-dimensioned sections called cells comprise several pixels which individually have a weight computed by getting the gradient magnitude and histogram. The histogram shows the dispersion of the pixel taking into cognizance the orientation of the gradient. Cell histogram is offered as a vector that is merged to get the HOG feature of the overall banknote image [30]. Figure 4 shows the terminology of

the HOG feature extraction. Finally, Table 1 shows the HOG method specification that provided the best result in this research. The Cell size of 16, four cells per block, 1 to 1 block overlap, nine bins, and 18x1 bin centers. The number of HOG feature is different for each Dataset because it is chosen based on the feature size that presented the best recognition and detection rate.

Table 1. Specification of the HOG Descriptor

Specification	Naira	USD	CAD	Euro
SURF Descriptor	124864	196240	209002	205855
Cell Size	16	16	16	16
Number of cells per block	4	4	4	4
Block Overlap	1,1	1,1	1,1	1,1
Number of bins	9	9	9	9
Number of HOG extracted	38102	56109	76464	70136
Bin centers	18 x 1	18 x 1	18 x 1	18 x 1

3.3. Feature Reduction (Proposed LVHCF)

After the HOG features have been extracted, the vector size is computationally demanding. Therefore, to unify and reduce the feature dimension; the techniques aim to minimize its effect on recognition/detection rate while maximizing the reduction of Feature dimensionality. The experiment conducts a descriptor filter analysis using Pearson's Product Moment Coefficient ρ and variance σ^2 to determine the best set of descriptors for classification. The technique answers two fundamental questions: determining which descriptor is relevant or not to the feature set and if the descriptor is redundant or not concerning other related descriptors.

$$\rho = \frac{m(\sum ij)-(\sum i)(\sum j)}{\sqrt{[m\sum i^2-(\sum i)^2][m\sum j-(\sum j)^2]}} \qquad (4)$$

$$\sigma^2 = \frac{\sum(x-u^2)}{m} \text{ where μ is the mean} \qquad (5)$$

The first question can be answered by just defining a threshold ρ value, but for the second question it is more complicated as it involves analysis of pairwise correlations between all descriptors. To answer both questions before filtering the descriptors we put forward an approach called a Low Variance and High Correlation Filter (LVHCF). In this experiment, it is assumed that if two descriptors are found to be redundant to each other one of them needs to be removed and removing the one that is less relevant to the descriptor-set concept keeps more information to predict the class while reducing redundancy in the data.

Table 2. Feature Reduction on Datasets

Technique	Naira	USD	CAD	Euro
PCA	680	740	690	755
BOW	600	650	650	670
LVF	1905	2805	2293	3507
HCF	1143	1685	2294	2104
LVHCF	581	661	764	701

As presented in Figure 5, given a dataset with M feature set and a descriptor set ds, the algorithm chooses a list of predominant descriptors D_{best} for a given ds. It comprises two phases, the first phase (line 2-17) computes the Pearson correlation moment coefficient p for all pairs of descriptors in the ds using equation 1, extracts relevant descriptors into $S1^i_{list}$ based on the predefined threshold t1. The $S1^i_{list}$ is further processed to remove additional correlated pairs and only keeps predominantly uncorrelated Dq and Dp among all the earlier selected descriptors in the $S1^i_{list}$ (line 13). After one round of filtering all Dq pairs based on Dp, the algorithm will take the currently remaining feature right next to Dp as the new reference (line 18) to repeat the filtering process. The algorithm stops until there is no more feature to be removed from $S1^i_{list}$. In the second phase (line 19 – 38), the variance σ^2 of

---

Input: S (D_1, D_2, …, D_m, ds)// training dataset

    t1 = 0.287   // a predefined threshold (ρ)

    t2 = 0.045   // a predefined threshold (σ^2)

output: S_{best} // an optimal subset

```
1    Begin
2      For i = 1 to M
3        For j = 1 to M
4      Do begin
5         Compute ρi, ds for all Di and Dj
6         if (ρi, ds >= t1)
7            Append Di to S1i list;
8      End;
9      Dp = getFirstDescriptor (S1i list);
10     Dq = getNextDescriptor (S1i list);
11       Compute ρ for all Dp and Dq
12       If (ρ >= t1)
13           remove Dq from S1i list;
14                Dq = getNextDescriptor (S1i list);
15       Else Dq = getNextDescriptor (S1i list);
16       End until (Dq == NULL);
17       Dp = getNextDescriptor (S1i list, Dp);
18      End until (Dp == NULL);
19     For i = 1 to M
20     Do begin
21        Compute σ2i, ds for S1i list;
22        if (σ2i, ds <= t2)
23           append Di to S2i list;
24     End;
25     Order S2i list in ascending σ2i, ds value;
26     Dp = getFirstDescriptor (S2i list)
27     Do begin
28         Dq = getNextDescriptor (S2i list, Dp);
29         if Dq <> NULL)
30           Do begin
31             D'q = Dq;
32             if (σ2p, q <= σ2q, ds)
33                remove Dq from S2i list;
34                Dq = getNextDescriptor (S2i list, D'q);
35             Else  Dq = getNextDescriptor (S2i list, Dq);
36           End until (Dp == NULL);
37         Dp = getNextDescriptor (S2i list, Dp);
38       End until (Dp == NULL);
39     Sbest = S2i list;
40   End;
```

Figure 5. Low Variance and High Correlation Filter (LVHCF) Algorithm

---

each descriptor in the $S1^i_{list}$ is gotten using equation 2. Furthermore, the descriptors are extracted into a new list $S2^i_{list}$ based on the predefined threshold t2 and ordered in ascending order according to their σ^2 values. The list is further processed to remove redundant descriptors and only keeps predominant ones among all the selected relevant descriptors. A descriptor Dp that has already been determined to be dominant can always be used to filter out other features that are ranked higher than Dp and have Dp as one of its redundant σ^2 peers. The iteration starts from the first descriptor in $S2^i_{list}$ (line 26) and continues as follows. For all the remaining descriptors, if Dp happens to be a redundant peer to a feature Dq, Dq will be removed from $S2^i_{list}$. After one round of filtering descriptors based on Dp, the algorithm will take the currently remaining descriptors right next to Dp as the new reference (line 37) to repeat the filtering process. The algorithm stops until there is no more feature to be removed from $S2^i_{list}$. The descriptors matching the variance remaining in $S2^i_{list}$ is then used as the final dataset B_{est}. Table 2 presents the feature reduction on the four datasets.

3.4. Classification

The classification stage is the decision segment of a recognition system, and it uses the features extracted in the previous phase. It involves identifying each banknote and assigning it the correct denomination class. Furthermore, the principal function of a pattern recognition system is to yield decisions concerning the class membership of the models with which it is confronted. Support Vector Machine (SVM) is a classification approach used for separating hyperplanes, it was designed in the 1990's for building nonlinear separating function, and real-valued estimation [31].

The Multi-Class SVM-based system uses a binary classification approach for the one versus one technique, it designs a classifier for each class and creates a hyperplane between the classes. Furthermore, the system assigns a class to a test sample based on a distance margin for the classes. Each data set is split into two groups, the training, and test groups. The system used a randomized approach for data selection, and the class of the test data was predicted based on the established class of the training mechanism. The Multi-Class SVM has two application method approaches; One Versus All (OVA) method and One Versus One (OVO) method. OVA differentiate one class from the remaining classes, and the OVO distinguishes each class from one class at a time while ignoring the remaining class [32].

For this research, the system used the OVO technique in combination with error correcting output codes. The error correcting output code divides a Multi-Class scheme into a double class system that constructs a classifier for each binary division. OVO technique compares all classes against each other and builds a binary classifier to distinguish each pair of class while ignoring the remaining classes. When classifying an unknown image to a class, the classifier conducts a vote, and the class with the highest majority is selected as the best choice to assign the banknote [33]. Further analysis was conducted using K-Nearest Neighbor (KNN), Weighted K-Nearest Neighbor (WKNN), Random Forest (RF) Decision Tree, and multilayer perceptron (MLP) Neural network. In conclusion, it is hard to decide which classification method is the best as research has only been able to prove that a

classifier could have a higher recognition and detection rate than others based on a specific problem.

4.   Result & Discussion

This system was implemented in MATLAB as its toolbox provides built-in support for SIFT, SURF, HOG, GLOH in addition to the several classifiers. The hardware utilized comprised of a Core i3, 2.4 GHz processor with 4 GB Ram, the operating system used was Windows 8.0 64Bit and image processing techniques, and other related MATLAB toolboxes employed in the development of the system. This research utilized 10-fold Cross-Validation for the analysis of the data. It is a statistical technique used in evaluating predictive models by dividing the data into training and validation sets. The approach divided the dataset into K-fold equal data sizes and crossed over each K-dataset partition in successive rounds such that each data point has a chance of being in the validation process. The Centre for Pattern Recognition and Machine Intelligence (CENPARMI) has developed a database for Banknote Identification (Experiment 1), Counterfeit Detection (Experiment 2), and Multiple Banknote recognition (Experiment 3). The database comprised of eight classes for the Naira Notes, US Dollars had six classes, five classes for the Canadian Dollars, and seven classes for the Euro Bills for both genuine and counterfeit bills [34, 35].
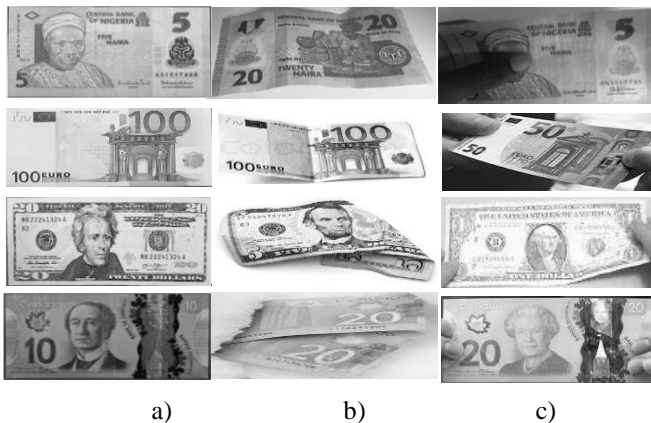
a)   b)   c)
Figure 6. a) Full notes b) Folded notes, and c) Occluded notes

4.1. Banknote Recognition (Experiment One)

The obverse and reverse of the banknotes were captured using a Canon T2i digital camera. Banknotes suffering from scratches, wrinkle, stains, wear and tear which represent the image quality of the paper monies in the real-world were captured for this project. The statistical distribution of the Dataset was Nigerian Naira, US Dollar, Canadian Dollar, and Euro data sets. It comprised 1790, 1120, 1000, and 1400 genuine banknotes and 1320, 823, 733, and 1020 fake banknotes respectively. Each dataset included images with multiple notes, clutters, overlapping and partial paper money. Furthermore, the datasets were made up of images captured in four directions for each banknote: the front (upward and downward) and back sides (upward and downward).

Figure 6 depicts pre-processed images employed in the single banknote recognition phase of the system drawn from the four datasets used namely Naira (First Row), Euro notes (Second Row), US Dollar (Third Row), and Canadian Dollar (Fourth Row). a) Comprises banknotes without clutters or occlusions captured with

minimal illumination interference, b) Consists of Partially Occluded Banknotes with clutters and distortions, and c) Made up of notes with wrinkling and folds held by a user in an open environment with minutiae illumination obstruction.

While Figure 7 measures the performance accuracy using SIFT, SURF, HOG, GLOH, Shape Context, and Modified HOG using KNN, WKNN, RF, MLP, OVO-SVM classifiers. The x-axis shows the accuracy while the y-axis gives the extracted features and the z-axis has the classifiers for the Naira, USD, CAD, and Euro Banknotes. From the results, the best accuracy for the Features and classifiers were SURF/Modified HOG and OVO-SVM while the worst performance was achieved with SIFT and KNN Respectively although the proposed Modified HOG presented comparatively higher accuracy in most cases it was barely higher than the SURF techniques. Furthermore, RF and MLP did produce promising results, but their accuracy was dependent on the descriptor and currency as RF outperformed MLP in some cases and MLP outweighed RF in performance accuracy in other feature and classifiers combination. Although, WKNN outperformed KNN they both still had low rates. Lastly, the most inferior combination accuracy was SIFT and KNN with an average of 82%, and the highest accuracy combination was Modified HOG and OVO-SVM with an average rate of 97.3% with the proposed SURF closely following behind with an accuracy of 97.29%.

Further experiments were conducted with feature reduction techniques to remove features that hinder the accuracy and precision of the proposed approach, and the error rate was utilized in measuring the performance of which is presented in Table 3 above. The table gives the Error Rate of OVO-SVM of the different Descriptors and Feature Reduction on the Four Datasets.

The proposed approach Modified HOG/LVHCF presented the lowest error rate within the average rate of 0.0270 after feature reduction while shape context/PCA had the worse error rate with an average of 0.0585. The rankings of the descriptor and Feature reduction technique ranging from best to worse are Shape Context, SIFT, GLOH, HOG, SURF, Modified HOG and PCA, BOW, LVF, HCF, LVHCF respectively. Additionally, each feature reduction approach was applied to all the descriptors to determine the best pair with the highest recognition rate. From the result, PCA had the lowest acceptance rate across all data sets while Modified HOG/LVHCF presented the best results across all datasets with relatively higher performance than SURF. Regarding datasets, Naira notes offered the best result while CAD had the worse rate although there were cases where CAD outperformed Euro and USD data sets when SIFT, SURF and Shape Context descriptors were employed. Lastly, the proposed approach presented promising results in comparison with existing techniques and can be engaged in descriptor dimension reduction.

4.2. Counterfeit Detection (Experiment Two)

Figure 8 depicts genuine and fake banknotes employed in this research, the first row consists of Naira notes, the second row has Euro paper money, the third row with the Canadian dollar, and the fourth/last row has US dollar. The dataset covers a wide variety of conditions, such as occlusion, rotation, changes in scaling, illumination, and viewpoints.
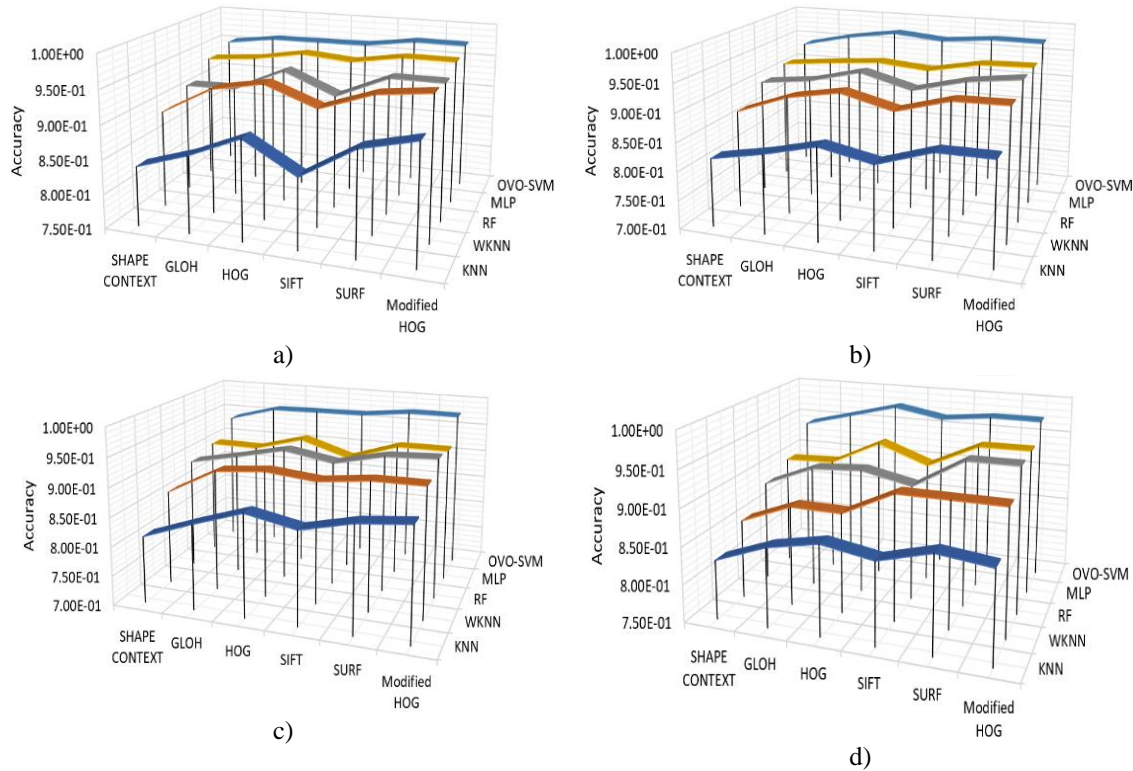
Figure 7: Accuracy measure of a) Naira, b) USD, c) CAD, and d) Euro using feature descriptors on different classifiers

The table presents the evaluation results of the proposed system on the data sets using OVO-SVM as it presented the best result during classification. For testing the system used 3110 samples for the Naira notes, 1943 instances for US Dollar, 1733 samples for Canadian Dollar and 2420 images of the Euro banknotes. Furthermore, the table also showed the Naira relatively had the highest correctly detected samples while the US dollar had the lowest successfully authenticated notes. However, all data sets had little incorrectly detected samples irrespective of the size of the data set; the system identified all images in the four datasets as paper money still, although some bills were wrongly classified. Finally, The Naira had the highest detection rate, the Euro came second, USA Dollar was third, and the Canadian dollar had the least authentication rate.

4.3. Multiple Banknote Recognition (Experiment Three)

For the multiple banknote recognition system, an object detection method in a cluttered scene using point feature matching of extracted HOG features based on image patches computed from the vertices of SURF points normalized using low variance and high correlation filter. It reads the test image and preprocesses it by applying grayscale conversion, noise removal, background subtraction, and normalization. Furthermore, the HOG Feature points are detected and used in matching the test image containing multiple banknotes is paired with the template image using the putative matching technique. The approach computes the transformation matched points while eliminating outliers. This aids in localizing the object in the scene. Lastly, the system locates the bounding polygon of the detected image and draws a black box to indicate the denomination.

a)   b)
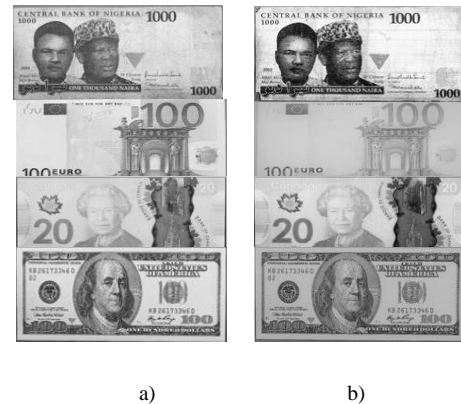
Figure 8. a) Genuine and b) Fake Banknotes

a)   b)   c)   d)
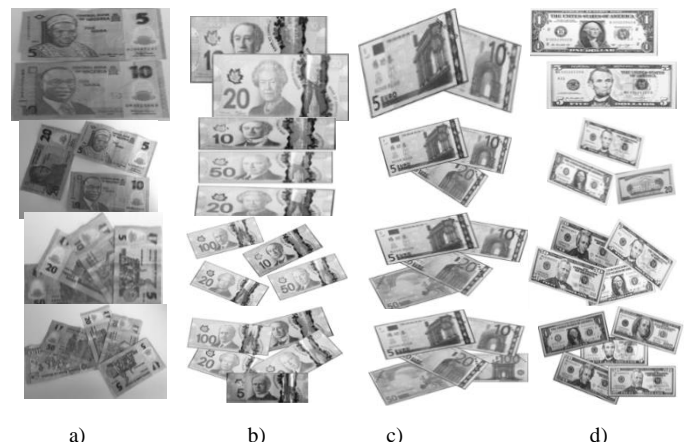
Figure 9. Multiple banknotes from a) Naira, b) CAD, c) Euro, and d) USD Dataset

Table. 3 Error Rate of feature reduction on different descriptors using OVO-SVM on the Four Datasets

No.	Descriptor	Feature Reduction	Naira	USD	CAD	Euro
1	Shape Context	No Feature reduction	0.0489	0.0627	0.0630	0.0594
		PCA	0.1500	0.1948	0.2104	0.1956
		BOW	0.1300	0.1553	1.7267	0.1678
		LVF	0.0950	0.1269	1.3056	0.1284
		HCF	0.0650	0.0754	0.0865	0.7253
		LVHCF	0.0700	0.0654	0.8721	0.6893
2	GLOH	No Feature reduction	0.0387	0.0421	0.0395	0.0415
		PCA	0.1300	0.1357	0.1243	0.1364
		BOW	0.0930	0.0892	0.7507	0.0885
		LVF	0.0600	0.0648	0.6340	0.0655
		HCF	0.0550	0.0540	0.5425	0.0574
		LVHCF	0.0575	0.0404	0.0409	0.0529
3	HOG	No Feature reduction	0.0379	0.0283	0.0379	0.0233
		PCA	0.1000	0.1092	0.1150	0.1144
		BOW	0.0900	0.0921	0.0988	0.1077
		LVF	0.0780	0.0795	0.0815	0.0921
		HCF	0.0550	0.0484	0.0502	0.0665
		LVHCF	0.0383	0.0285	0.0380	0.0238
4	SIFT	No Feature reduction	0.0371	0.0365	0.0368	0.0354
		PCA	0.1010	0.1204	0.1155	0.1288
		BOW	0.0750	0.0764	0.0843	0.0751
		LVF	0.0586	0.0582	0.0577	0.0549
		HCF	0.0400	0.0488	0.0422	0.0445
		LVHCF	0.0380	0.0469	0.0388	0.0363
5	SURF	No Feature reduction	0.0261	0.0276	0.0268	0.0275
		PCA	0.0907	0.1048	0.1166	0.1287
		BOW	0.709	0.0977	0.0928	0.0845
		LVF	0.558	0.0660	0.0702	0.0681
		HCF	0.350	0.0408	0.0502	0.0455
		LVHCF	0.0295	0.0389	0.3049	0.0315
6	Modified HOG	No Feature reduction	0.0266	0.0279	0.0268	0.0274
		PCA	0.0837	0.0884	0.0874	0.0865
		BOW	0.0649	0.0651	0.0662	0.0672
		LVF	0.0351	0.0257	0.0318	0.0328
		HCF	0.02660.	0.0384	0.0296	0.0299
		LVHCF	0265	0.0277	0.0268	0.0276

Figure 9 shows multiple banknote samples utilized in this research: a) comprises Naira notes, b) Canadain Dollar notes c) Euro bills, and d) US Dollar. The first row contains two bank notes; the second row contains three banknotes, the third row includes four bank notes, and the last/fourth row has five bank notes. The multiple notes dataset included clutters, distortions, rotations and overlapping captured in different illumination conditions.

While Figure 10 presents the database employed for the numerous banknote detection phases of the system, it consisted of 460 Naira images, 372 US dollar samples, 440 Canadian dollar images, and 365 Euro Samples. Table 6 shows the dataset of Naira, Euro, CAD and US notes used in the multiple banknote recognition experiments ranging from images with one banknote to five bills. It shows the precision and recall of this analysis; this was computed to investigate the effectiveness of our systems in distinguishing many banknotes in one image and determine which of the several banknote classes can best predict the note denomination irrespective of distortions.
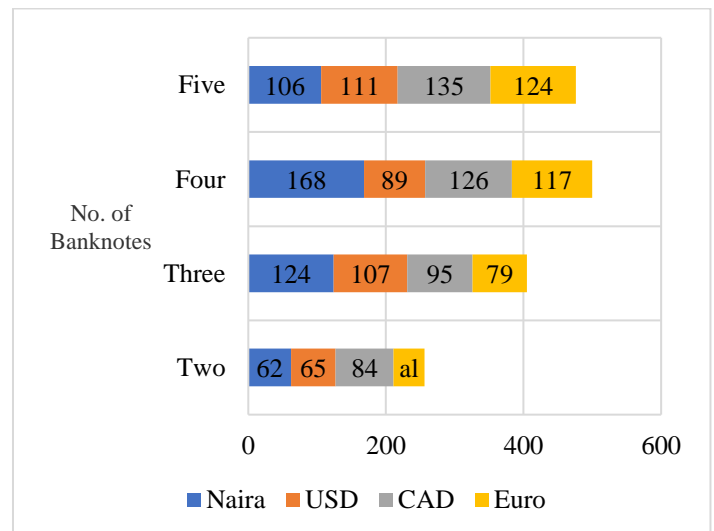
Fig. 10. Multiple banknotes Dataset
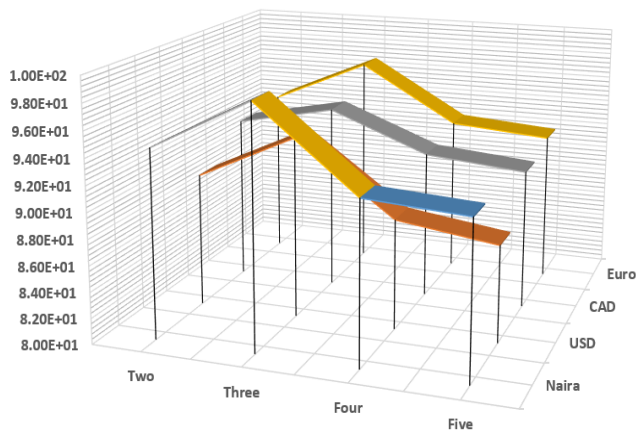
Figure 11. Recognition Rates of a) Naira b) USD c) CAD and d) Euro on multiple banknotes

Furthermore, in Figure 11 we have measured the Recognition Rate using One, two, three, four and five notes on the Naira, USD, CAD, and Euro Database. For simplicity, we have shown here only the recognition rate on the x-axis, while the number of the banknotes is fixed on the y-axis and the currency type on the z-axis. The reason for this experiment is to evaluate the effectiveness of the proposed feature extraction approach on multiple banknote recognition. The system recognized several currencies including same notes in the same image, separated, overlapped, cluttered and partially obstructed. After three bills the acceptance result reduced when the number of banknotes was increased. In all cases, the accuracy measure is more than 87%. Although one, two and three notes show more than 92% while four and five ranged between 87.50% and 92.45%. For this research, the best recognition rate across all datasets was three bills using Naira at 98.60%, and the worse was with five banknotes using USD at 87.50%.

Table 5. Qualitative Analysis of feature descriptors

Method	Time	Scale & Rotation	Illumination	Affine
Shape Context	Better	Better	Common	Good
SIFT	Common	Best	Common	Good
SURF	Best	Common	Common	Good
HOG	Common	Common	Best	best
GLOH	Good	Best	Common	Good
Modified HOG	Best	Better	Best	Best

In Table 5, we empirically compare the performance of all feature descriptors for multiple banknote identifications concerning its variants in five distinct scenarios: speed, scale, rotation, illumination, and affine invariance. The results of the experiment were analyzed qualitatively to give a general idea of the performance of each algorithm in distinct scenarios using four grades: Common, Good, Better, and Best. As shown in the table above, Shape context has significant advancements in speed, scale, affine, and rotation changes. But not illumination change. SIFT is slow; it is invariant to rotation and scale, maintains a stable degree of stability for affine transformation and a low threshold to illumination change. SURF is fast and has better performance than SIFT and GLOH, but it is not stable to the

rotation and illumination changes. HOG is robust to rotation, illumination, positioning, and affine modifications but very slow and scale variant. GLOH has excellent performance in scale and rotation, presented reasonable speed and affine changes and had a deficient sense of illumination invariance. The Modified HOG descriptor maintains a proper stability illumination, rotation, and affine transformation. It can also handle reasonable scale invariance and compared with the other algorithms; it has distinct advantages regarding speed. Thus, can be deduced that the proposed technique presented a promising performance in banknote recognition. Lastly, the result of this qualitative analysis is not constant in all cases as changes to the dataset, and classifiers can lead to a different conclusion.

4.4. Training Sample Needed

One critical issue is determining the number of notes by denomination and authenticity is needed to ensure a reasonable performance from the proposed technique. To solve this problem, the various databases are evaluated concerning the sizes by varying the training set dimension. A remarkable increment on the recognition and error rate is seen when the training data set increases from 300 to 800 for banknote recognition, 100 to 250 for counterfeit detection, and 50 to 200 for multiple currency recognition. After which growing the data further only produced asymptotical performance improvement but had no effect on its accuracy and error rate, and this could be attributed to the inclusion of more distorted notes.

5. Conclusion and Future Work

This research was conducted to determine the efficiency of selected Modified HOG/LVHCF descriptor for automatic recognition of paper money. The proposed system was implemented using the Nigerian notes, and for experimental evaluation of the system, a further test was conducted on USD, CAD, and Euro notes. The proposed method extracted the HOG descriptor from image patches created from the vertices of SURF interest points feature of the banknotes, Low Variance, and High Correlation feature reduction filter was employed to reduce the feature set and classified using OVO-SVM in identifying and authenticating fake and genuine bills. Lastly, point matching technique was applied to detect feature vectors to match images containing multiple banknotes to determine the value of the currencies. Future research would include incorporating new feature descriptors and reduction approach to reduce processing time in addition to scaling down the application to work on mobile devices and reducing the computational complexity to enable embedding in micro-controllers and microchips.

References

[1] D. Alekhya, G. Devi Surya Prabha, G. Venkata Durga Rao, "Fake currency detection using image processing and other standard methods," IJRCCT, 3(1), pp. 128-131, 2014.

[2] M. Thakur, A. Kaur, "Various Fake Currency Detection Techniques," International Journal for Technological Research in Engineering, 1(11), 2014.

[3] A. Berenguel, O.R. Terrades, J. Lladós, C. Cañero, "Banknote counterfeit detection through background texture printing analysis," 12th IAPR Workshop on Document Analysis Systems (DAS), pp. 66-71, 2016.

[4] M.S. Uddin, P.P. Das, M.S.A. Roney, "Image-based approach for the detection of counterfeit banknotes of Bangladesh," (ICIEV), 5th International Conference on Informatics, Electronics and Vision, pp. 1067-1072, 2016.

[5] H. Hassanpour, A. Yaseri, G. Ardeshiri, "Feature extraction for paper currency recognition," ISSPA 2007. 9th International Symposium on Signal Processing and Its Applications, pp. 1-4, 2007.

[6] N. Kato, S. Omachi, "A handwriting character recognition system using directional element feature," IEEE Trans. Pattern Anal. Mach. Intell, 21(3), pp. 258–262, 1999.

[7] F. García-Lamont, J. Cervantes, A. López, "Recognition of Mexican banknotes via their color and texture features," Expert Systems with Applications, 39(10), pp. 9651-9660, 2012.

[8] H. Aggarwal, P. Kumar, "Indian Currency Note Denomination Recognition in Color Images," International Journal on Advanced Computer Engineering and Communication Technology, 1(1), pp. 12-18, 2014.

[9] P.D. Pawar, S.B. Kale, "Recognition of Indian Currency Note Based on HSV Parameters," International Journal of Science and Research (IJSR), 3(6), pp. 132-137, 2014.

[10] J. Lee, J. Seong-Goo, K. II-Hwan, "Distinctive point extraction and recognition algorithm for various kinds of euro banknotes," International Journal of Control, Automation, and Systems, 2(2), pp. 201-206, 2004.

[11] M. Ionescu, A. Ralescu, "Fuzzy hamming distance-based banknote validator," FUZZ'05, The 14th IEEE International Conference on Fuzzy Systems, pp. 300-305, 2005.

[12] H. Hassanpour, M.F. Payam, "Using Hidden Markov Models for paper currency recognition," Expert Systems with Applications, 36(6), pp. 10105-10111, 2009.

[13] F. Poor, T. Mohammadpour, A. Kianis, "Persian Banknote Recognition Using Wavelet and Neural Network [A]," in Proceedings of the IEEE Conference on Computer Science and Electronic Engineering, 3, pp. 679-684, 2012.

[14] W. Selesnick, R.G. Baraniuk, N.G. Kingsbury, "The dual-tree complex wavelet transform," IEEE Trans. Signal Process, 22(6), pp. 123–151, 2005.

[15] W.L. Chan, H. Choi, and R.G. Baraniuk, "Coherent multiscale image processing using dual-tree quaternion wavelets," IEEE Transactions on Image Processing, 17(7), pp. 1069-1082, 2008.

[16] S. Gai, G. Yang, M. Wan, "Employing quaternion wavelet transform for banknote classification," Neurocomputing, 118, pp. 171-178, 2013.

[17] F. Ahangaryan, T. Mohammadpour, A. Kianisarkaleh, "Persian Banknote Recognition Using Wavelet and Neural Network," (ICCSEE), 2012 International Conference on Computer Science and Electronics Engineering, 3, pp. 679-684, 2012.

[18] A. Rajaei, E. Dallalzadeh, M. Imran, "Feature extraction of currency notes: an approach based on wavelet transform. (ACCT)," Second International Conference on Advanced Computing & Communication Technologies, pp. 255-258, 2012.

[19] C. Harris, M. Stephens, "A combined corner and edge detector" In Alvey vision conference, 15(50), pp. 147-151, 1988.

[20] T. Lindeberg, "Feature detection with automatic scale selection," International Journal of computer vision, 30(2), pp. 79–116, 1998.

[21] K. Mikolajczyk, C. Schmid, "Indexing based on scale invariant interest points," ICCV Proceedings. Eighth IEEE International Conference on Computer Vision, 1, pp. 525–531, 2001.

[22] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of computer vision, 60(2), pp. 91–110, 2004.

[23] T. Reiff, P. Sincak, "Multi-agent sophisticated system for intelligent technologies," ICCC International Conference on Computational Cybernetics, pp. 37-40, 2008.

[24] K. Van De Sande, T. Gevers, C. Snoek, "Evaluating color descriptors for object and scene recognition," IEEE transactions on pattern analysis and machine intelligence, 32(9), pp. 1582-1596, 2010.

[25] T.H. Rassem, B. E. Khoo, "Object class recognition using combination of color SIFT descriptors," IEEE International Conference on Imaging Systems and Techniques (IST), pp. 290-295, 2011.

[26] N.S. Pai, J.H. Hong, P.Y. Chen, J.K. Wu, "Application of design of image tracking by combining SURF and TLD and SVM-based posture recognition system in robbery pre-alert system," Multimedia Tools and Applications, 76(23), pp. 25321-25342, 2017.

[27] F.M. Hasanuzzaman, Y. Xiaodong, T. Yingli, "Robust and effective component-based banknote recognition for the blind," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42(6), pp. 1021-1030, 2012.

[28] W. Chen, S. Ding, Z. Chai, D. He, W. Zhang, G. Zhang, W. Luo, "FPGA-Based Parallel Implementation of SURF Algorithm," (ICPADS). 22nd International Conference on Parallel and Distributed Systems, pp. 308-315, 2016.

[29] Z.R. Wang, Y.L. Jia, H. Huang, S.M. Tang, "Pedestrian detection using boosted HOG features," ITSC 11th International IEEE Conference on Intelligent Transportation Systems, pp. 1155-1160, 2008.

[30] T.V. Dittimi, A.K. Hmood, C.Y. Suen, "Multi-class SVM based gradient feature for banknote recognition," (ICIT), IEEE International Conference on Industrial Technology, pp. 1030-1035, 2017.

[31] N.K. Verma, R. Abhishek, S. Al, "An optimized fault diagnosis method for reciprocating air compressors based on SVM," (ICSET), IEEE International Conference on System Engineering and Technology, pp 65-69, 2011.

[32] A. Rashid, A. Prati, R. Cucchiara, "On the design of embedded solutions to banknote recognition," Optical Engineering, 52(9), pp. 093106-093106, 2013.

[33] A. Govada, G. Bhavul, K.S. Sanjay, "Distributed multi-class SVM for large datasets," In Proceedings of the Third International Symposium on Women in Computing and Informatics, pp. 54-58, 2015.

[34] T.V. Dittimi, C.Y. Suen, "Mobile App for Detection of Counterfeit Banknotes," In Advances in Artificial Intelligence: 31st Canadian Conference on Artificial Intelligence, 31, pp. 156-168, 2018.

[35] T. Dittimi, C. Suen, A. Hmood, "Mobile Based Assistive Technologies for Authentication of Banknotes," ICPRAI 2018 - International Conference on Pattern Recognition and Artificial Intelligence, pp. 366-371, 2018.

ASTES

Characteristic, Thermochemical Behaviors and Kinetic of Demineralized and Torrefied Empty Fruit Bunches (EFB)

Nur Nasulhah Kasim[*,1,2], Khudzir Ismail[1,2], Alina Rahayu Mohamed[3], Mohd Azlan Mohd Ishak[1,2], Razi Ahmad[2,4], Wan Izhan Nawawi Wan Ismail[1,2,5]

[1]*Coal and Biomass Energy Research Group, Universiti Teknologi MARA (UiTM), 40450, Shah Alam, Selangor, Malaysia*

[2]*Faculty of Applied Sciences, Universiti Teknologi MARA Malaysia, 40450 Shah Alam, Selangor, Malaysia*

[3]*Faculty of Engineering Technology, Unicity Alam, UniMAP, Padang Besar, 02400 Perlis, Malaysia*

[4]*School of Environmental Engineering, Universiti Malaysia Perlis, 02600 Arau, Perlis, Malaysia*

[5]*Department of chemistry, University York, Heslington, YO10 5DD, United Kingdom*

A R T I C L E   I N F O

A B S T R A C T

A sequential pre-treatment of demineralization and torrefaction, was carried out on palm empty fruit bunches (EFB). EFB and demineralized EFB (DEFB) were torrefied in a vertical tubular reactor in the temperature range of 200 to 280 °C for 30 mins under nitrogen (flow rate:100 mL/min. The pretreated samples were characterized using proximate and ultimate analyses, fuel properties, and Fourier-transform infrared (FTIR) spectroscopy techniques. The thermal and kinetic study on the torrefied samples were carried out using thermogravimetric analysis. The results showed that sequential pre-treatment enhances the properties of solid EFB by increasing the carbon content and reducing the oxygen content with increasing the calorific value. Fuel properties of torrefied samples showed the mass and energy yield decreased, with an increase in energy density. In addition, the FTIR spectra showed the decomposition of hemicellulose occurring for torrefied samples as evidenced by the disappearance of the vibrational features belonging to hydroxyl and carbonyl groups. The kinetic study carried out using Coats-Redfern method on torrefied samples suggested that the activation energy can be transferred by the sequential pre-treatment, indicating that the abundant energy it has can be converted into bio oil of high quality. Apparently, torrefied samples bear high potential to be used as biofuel feedstock when exposed to further thermal decomposition and pyrolysis processes.

1. Introduction

This paper is an extension of work the originally presented in 4th IET Clean Energy and Technology Conference (CEAT 2016) [1]. The extended work was focused on sequential pre-treatment of demineralization and torrefaction of empty fruit bunches (EFB) feedstock. Characteristic and fuel properties of pre-treated EFB were investigated and its kinetic study was performed using Coats-Redfern method.

Depletion of energy resources is a crucial issue which has been challenged worldwide. Fossil fuel energy resources such as petroleum, natural gas and coal are expected to last for approximately 35 years, 37 years and 104 years, respectively [2]. Utilization of agricultural waste seems to assist the shortage of energy resources especially for countries having such resources [3]. Malaysia is the second largest producer of palm oil giving rise to abundant oil palm wastes during the refined, bleached and deodorized (RBD) palm oil processing[4].

Previously, many researchers have focussed on the utilisation of the oil palm waste such as empty fruit bunches (EFB) [5], oil palm fronds (OPF) [6], mesocarp fibre (MF) and palm kernel shell (PKS) [6,7,8] due to their potential benefits for energy sources. However, hygroscopic behavior and higher content of alkali and alkaline earth metals (AAEMs) particularly in EFB limits the

utilization of biomass in producing the energy due to vulnerable to biological attack and biodegradation [10]. In addition, nearly 60 % of water contained in wet EFB and the heating value of the dry EFB is roughly half that of palm oil [4].Thus, EFB is unsuitable to be directly burnt for energy production. Therefore, some pre-treatments need to be introduced on EFB prior to pyrolysis to produce high quality of bio oil that contain hydrocarbon content with low carboxylic acid molecules [9,10].

From previous research, feasible pre-treatments on EFB have been investigated such as demineralization and torrefaction in order to enhance the properties of biomass prior to thermal conversion [11,12,13]. Demineralization is an efficient pre-treatment that can be carried out using water, acid or alkaline solution to eliminate the inorganic compounds that assist the secondary reaction during the pyrolysis. Although water reduced about 50 % concentration of inorganic compounds in most of the biomass, chemicals like acid and alkali solution eliminated almost 100 % of the compounds [14,15].

Generally, torrefaction is the often-applied pre-treatment to improve the quality of biomass. The process is carried out in an inert condition at atmospheric pressure between the temperatures of 200-320 °C. This process is a mild version of slow pyrolysis with the aim of reducing the carbonyl, carboxyl, hydroxyl group compounds, and lowering moisture and oxygen contents. In addition, the grindability and hydrophobicity are improved by breaking down the cell wall and fiber structures of the biomass [11]. Washing, drying and size reduction on biomass is a simple pre-treatment that need to be prepared prior to pyrolysis [18]. Basically, an efficient pre-treatment was seen to be easier, economic and less time consuming with simple processing steps. Nevertheless, single pre-treatment such as demineralization or torrefaction alone was as an inappropriate method to obtain high quality bio oil [9,15,17] from raw EFB having high AAEMs and oxygen content.

Previously, application of demineralization and torrefaction sequentially was reported as a novel promising method to reduce the undesirable natural catalyst (AAEMs), moisture, acetyl and acid contents of the biomass samples [10,17]. Therefore, a combination of demineralisation and torrefaction processes was applied to reduce the oxygen, carboxyl, water and ash contents of biomass [19]. Similarly, demineralization using water washing technique combined with torrefaction pre-treatment on microwave pyrolysis of rice husk resulted in large elimination of inorganic compounds which thus, could enhance the quality of fuel characteristic [17].

Pyrolysis of biomass such as EFB, PKS, cotton stalk and rice straw has been performed under dynamic non-isothermal condition in thermogravimetric analyser. To our best knowledge, a few investigations on thermal behaviour and most kinetic studies of biomass pyrolysis have been done using thermogravimetric analyser [4,18,19,20]. In addition, more information or data on kinetic parameters, i.e., activation energy, pre-exponential factor, reaction order and mechanism of reaction can be generated from thermogravimetric analysis [23].

However, there is limited information about physicochemical properties of TEFB and TDEFB at various torrefaction temperatures and very scarce information is available on pyrolysis of demineralized torrefied biomass especially using model free method to calculate activation energy. Thus, the present study was performed to investigate the effect of the sequential application of demineralisation and torrefaction at 200, 220, 240, 260 and 280 °C to enhance the solid fuel qualities.

The basic characteristic of pretreated oil palm EFB including proximate and ultimate analyses, fuel properties and FTIR analysis were elucidated for further utilization of this biomass feedstock. Furthermore, the thermal behaviour and kinetic study of the selected torrefied EFB (i.e. TEFB 240) and torrefied demineralized EFB (i.e. TDEFB 240) were also searched during pyrolysis using thermogravimetric analysis. These data are of significant importance to understand the thermal decomposition and for further effective upgrading of the pre-treatment process. Besides, the thermochemical evolution profiles help designing suitable reactor and setting the optimum condition parameters for the system.

2. Methodology

2.1. Biomass samples preparation

The empty fruit bunches (EFB) were collected from the United Oil Palm Industries Sdn Bhd located in Nibong Tebal, Pulau Pinang, Malaysia. The biomass was rinsed with tap water to remove impurities and dried in oven at 105 °C for 24 h. It was ground using a hammer mill machine into small particle size and sieved using the sieve shaker to 0.5-1.0 cm. Finally, the dried biomass was stored in air-tight containers for further use.

2.2. Sample labels

Untreated EFB, demineralized EFB, torrefied EFB, torrefied demineralized EFB, EFB torrefied at 240 °C and torrefied demineralized EFB at 240 °C were represented as EFB, DEFB, TEFB240 and TDEFB240, respectively.

2.3. Demineralization of EFB (DEFB)

Demineralization of EFB was carried out to eliminate AAEMs that reduce the ash content in the sample. From our preliminary work, sonication for 10 min of EFB in 1 % nitric acid solution at room temperature gave positive results with minimum ash content of 0.74 % in comparison to 4.21 % in untreated EFB [1]. This demineralized EFB sample was used for torrefaction study here.

100 g of EFB was placed in a flask containing 500 mL of 1% nitric acid and was sonicated for 10 min in ultrasonic bath at room temperature. After acid leaching, the sample was filtered and dried at 105 °C for 24 h to a constant weight. Weight measurements for all samples before and after the pre-treatment were carried out using an electronic balance.

2.4. Torrefaction experiment

Torrefaction experiments were performed in a vertical tube fixed bed reactor shown schematically in Figure 1. Each experiment was conducted with 2.0 g of untreated and demineralized EFB samples in inert atmosphere under nitrogen flow (100 mL/min) from ambient temperature to various torrefaction temperatures of 200, 220, 240, 260 and 280 °C, bu keeping at the selected temperature for 30 min. Nitrogen flow was purged before and after 5 min of the experiment and the sample was removed from the reactor after cooling.

2.5. Characterization of EFB, TEFB and TDEFB samples

For each sample, the moisture and ash content, calorific value and ultimate and proximate analyses were carried out. The

moisture content was measured by drying the samples at 105 °C for 4 h [24]. The total ash content and calorific value were determined by ASTM E-1755-01 [25] and ASTM D 2015 – 96 [26] using LECO bomb calorimeter AC-350, respectively. The ultimate analysis of EFB samples was performed according to ASTM D3176 [27] using CHNS-O Flash 2000 with O content determined by difference. Proximate analysis of EFB, TEFB and TDEFB was carried out using Mettler Toledo thermogravimetric analyser (TGA). 10.0 mg of samples were pyrolyzed from ambient temperature to 900 °C at the heating rate of 20 °C/min. The volatile matter was determined using TG/DTG curves and fixed carbon was calculated by difference.
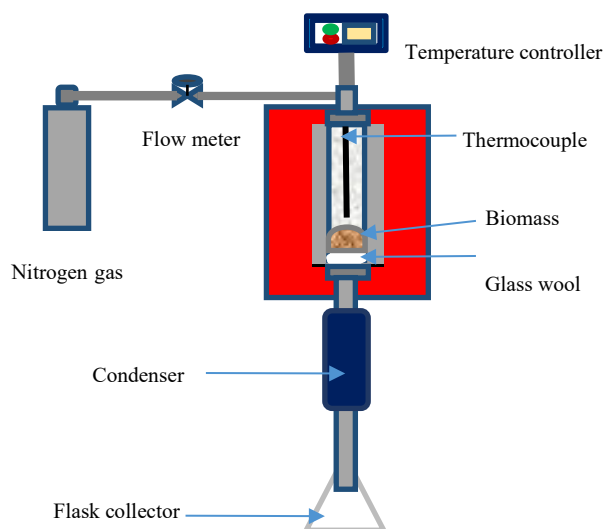
Figure 1. Schematic diagram of torrefaction reactor furnace

The Inductive couple plasma optical emission spectrometer (ICP-OES) was used to determine the inorganic metal content of the EFB and DEFB samples. 300 ± 10 mg of samples were digested with 2.5 mL of 65 % nitric acid and 2.5 mL of 30 % hydrogen peroxide (H_2O_2) in high performance microwave digestion system, Ethos One. The temperature was ramped to 90 °C over 15 min, held for 5 min, then ramped again to 180 °C over 10 min and held for 15 min [11]. The digested samples were analysed using Perkin Elmer Optima 8000 ICP-OES.

Qualitative analysis of functional groups in each EFB samples was conducted using Perkin Elmer Fourier-transform infrared spectroscopy (FTIR) technique. The samples were scanned in the range from 4000-600 cm^{-1} at a resolution of 4 cm^{-1}. For fuel properties analyses, the mass and energy yields of each samples were calculated using equations (1), (2) [28] and energy density equation (3) [29].

$$Mass\ yield\ (Y_{mass}) = \frac{Mass\ after\ torrefaction}{Mass\ of\ raw\ sample} \times 100\% \quad (1)$$

$$Energy\ yield\ (Y_{energy}) = Y_{mass} \times \frac{heating\ value\ of\ torrefied\ sample}{heating\ value\ of\ raw\ sample} \times 100\% \quad (2)$$

$$Energy\ density\ (Y_{energy\ density}) = \frac{Energy\ yield\ (Y_{energy})}{Mass\ yield\ (Y_{mass})} \quad (3)$$

Thermal and kinetic analyses of the EFB, TEFB240 and TDEFB240 samples were carried out using a Mettler Toledo

thermogravimetric analyser. 10.0 mg of each EFB sample pyrolysed at the heating rate of 20 °C/min under nitrogen flow (100 mL/min) from ambient temperature to 900 °C.

In this study, Coats-Redfern method which is widely used to follow the thermal decomposition of biomass [4,18,20,28] was applied for modelling the decomposition kinetics of raw EFB, TEFB240 and TDEFB240 samples. By this method, kinetic parameters such as activation energy, E_a and pre-exponential factors, ln A can be determined. In kinetic analysis, the rate of pyrolysis was assumed as:

$$\frac{d\alpha}{dt} = kf(\alpha) \quad (4)$$

In this equation, α represents the progress of reaction or the conversion which occurs between 0.05 to 0.95. α is the progress of reaction during the biomass decomposition which can be defined as:

$$\alpha = \frac{w_0 - w_t}{w_0 - w_f} \quad (5)$$

where w_0, w_t, and w_f are the initial, current (at time t) and final weight of the biomass samples. The rate is constant and dependent on the activation energy (E_a).

$$k = A\exp(-\frac{E_a}{RT}) \quad (6)$$

Considering the assumption of the Coats-Redfern method (2RT/E <<1), and following the first order reaction kinetic, equation (7) represents the kinetic equation used to calculate the activation energy (E_a) and pre-exponential factors, ln A.

$$\ln\left[-\frac{\ln(1-\alpha)}{T^2}\right] = \ln\left[\frac{AR}{\beta E_a}\left[1 - \frac{2RT}{E_a}\right]\right] - \frac{E}{RT} \quad (7)$$

According to kinetic equation (7), the plot of $\ln[-\ln(1-\alpha)/T^2]$ against 1/T give a correlation straight line with the first order kinetics. The activation energy and the pre-exponential factor, ln A can be determined from the slope and the intercept of the straight line, respectively [31].

3. Result and Discussion

3.1. Proximate analysis of EFB, DEFB, TEFB and TDEFB

The proximate analysis of EFB, DEFB, TEFB and TDEFB is shown in Table 1. Apparently, the ash content of DEFB dramatically decreased to 0.74 wt. % in comparison with EFB at 4.21 %.

Moreover, the effect of torrefaction on DEFB (represented as TDEFB) can be seen from the increase in the ash content changing in the range from 1.31 – 2.69 %. Likewise, for TEFB, as the torrefaction temperature increased from 200 to 240 °C, the ash content increased slightly and almost doubled at a higher torrefaction temperature of 260-280 °C in comparison with EFB. It was due to the induces partial removal of volatiles from biomass

after the torrefaction pre-treatment. Indeed, torrefaction has a marked effect on the ash content of both the TEFB and TDEFB samples. This implies that sequential application of demineralization and torrefaction on EFB reduced the ash content and offset the negative impact on the increased of ash content due to torrefaction [19].

Table 1. Proximate analysis of EFB, TDEFB and TDEFB

Sample	Proximate analysis (% db)		
	Volatile matter (VM)	Fixed carbon (FC)	Ash
EFB	53.85	41.93	4.21
DEFB	73.60	25.65	0.74
TEFB 200	72.95	22.49	4.54
TEFB 220	72.54	22.64	4.81
TEFB 240	45.95	49.12	4.91
TEFB 260	46.83	44.67	8.49
TEFB 280	46.81	43.59	9.58
TDEFB 200	59.23	39.45	1.31
TDEFB 220	56.27	42.35	1.37
TDEFB 240	63.04	35.08	1.87
TDEFB 260	69.09	29.13	1.76
TDEFB 280	47.98	49.32	2.69

With reference to volatile matter content, TEFB at low torrefaction temperature range of 200 and 220 °C showed a drastic increase by about 35 % in comparison to that of EFB. On contrary, fixed carbon at low torrefaction temperatures (200-220 °C) of TEFB decreased by 46 % which is significant. This trend can be explained by the dehydration and decarboxylation of hemicellulose which occurring at lower torrefaction temperature [30,31]. During the dehydration and decarboxylation, hemicellulose decomposes into H_2O, CO_2, CO and solid char as well as low molecular weight hydrocarbon [28].

Consequently, at the higher torrefaction temperature range from 240 to 280 °C, volatile matter of TEFB was found to decrease with the increase in fixed carbon. This observation is attributed to the initial thermal decomposition of cellulose [34] which completely destroys the cell structure of EFB and causes a further increase in the amount of fixed carbon during high torrefaction temperature [26,33].

As for TDEFB, the volatile matter content became slightly higher with the reduction in fixed carbon content in the torrefaction temperature range from 200 to 260 °C in comparison with EFB. These results indicated a slow rate decomposition of hemicellulose in DEFB during torrefaction. Similar observation was also reported in other studies investigated the influence of inorganic metals on the decomposition of biomass [36]. In contrast, at the torrefaction temperature of 280 °C of TDEFB, the reduction of volatile matter and the increase in fixed carbon content were observed.

This observation might be due to the extensive decomposition of hemicellulose into volatiles and solid products that took place at more severe torrefaction condition [37]. These results indicated that demineralization pre-treatment influenced the degradation of lignocellulosic components in EFB and facilitated the

decomposition of hemicellulose and cellulose at different temperatures during torrefaction pre-treatment.

3.2. Ultimate analysis and calorific value of EFB, DEFB, TEFB and TDEFB

The H/C ratio, O/C ratio and calorific value for EFB, DEFB, TEFB and TDEFB are listed in Table 2. The ultimate analysis revealed the alterations of chemical composition of EFB which occurred after the pre-treatment of demineralization and torrefaction. The carbon content was found to increase for single sequential pre-treatment of demineralization and torrefaction. When the torrefaction temperature was increased, the carbon content gradually increased whereas hydrogen and oxygen contents of TEFB and TDEFB samples decreased. These findings can be explained by the effects of dehydration, decarboxylation, and demethanation reactions of hemicellulose [38] and partial depolymerization of cellulose during torrefaction [34].

In the literature [35], the atomic ratio of O/C and H/C in biomass are found between 0.4 to 0.8 and 1.2 to 2.0, respectively. As shown in Table 2, H/C ratio of DEFB was slightly lower in comparison with EFB which decreased from 1.66 to 1.52. Similar trend was also observed for H/C and O/C ratio of TEFB and TDEFB when the torrefaction temperature was increased from 200 to 280 °C. This clearly indicated that, moisture and light volatile compounds which contain high hydrogen and oxygen were eliminated from EFB and DEFB, and consequently the carbon content of the materials increased after torrefaction. In general, sulphur and nitrogen content in biomass sample is much lower than that of coal [35].

Both demineralization and torrefaction pre-treatments could remove sulphur (S) and reduce the nitrogen (N) content of EFB (Table 2). This result was supported by previous studies reporting that sulphur and nitrogen content in biomass were reduced after torrefaction [39] and wet torrefaction [37,39].

Table 2. Ultimate analysis, atomic ratio and calorific value of raw and pre-treated empty fruit bunch (EFB)

Sample	Ultimate analysis				Calorific value
	O/C	H/C	N/C	S/C	(MJ kg^{-1})
EFB	0.79	1.66	0.03	-	17.80
DEFB	1.52	0.78	0.01		17.30
TEFB 200	0.78	1.53	0.00	-	18.50
TEFB 220	0.74	1.48	0.01		18.90
TEFB 240	0.69	1.27	0.00		20.60
TEFB 260	0.57	1.05	0.00		21.10
TEFB 280	0.48	0.99	0.00		23.50
TDEFB 200	0.73	1.50	0.00	-	17.90
TDEFB 220	0.70	1.43	0.00		18.20
TDEFB 240	0.68	1.39	0.00		19.70
TDEFB 260	0.63	1.27	0.00		21.10
TDEFB 280	0.45	0.99	0.01		23.50

Figure 2 compares the FTIR spectra of EFB with TEFB and TDEFB at various torrefaction temperatures. The FTIR spectra showed four prominent absorption peaks at 3345 cm^{-1} (i), 2920 cm^{-1} and 2850 cm^{-1} (ii), 1650 cm^{-1} and 1710 cm^{-1} (iii) and 1033 cm^{-1} (iv) which are assigned to the vibration of hydroxyl group (-OH), aliphatic methylene group, carbonyl groups (C=O) from aldehydes
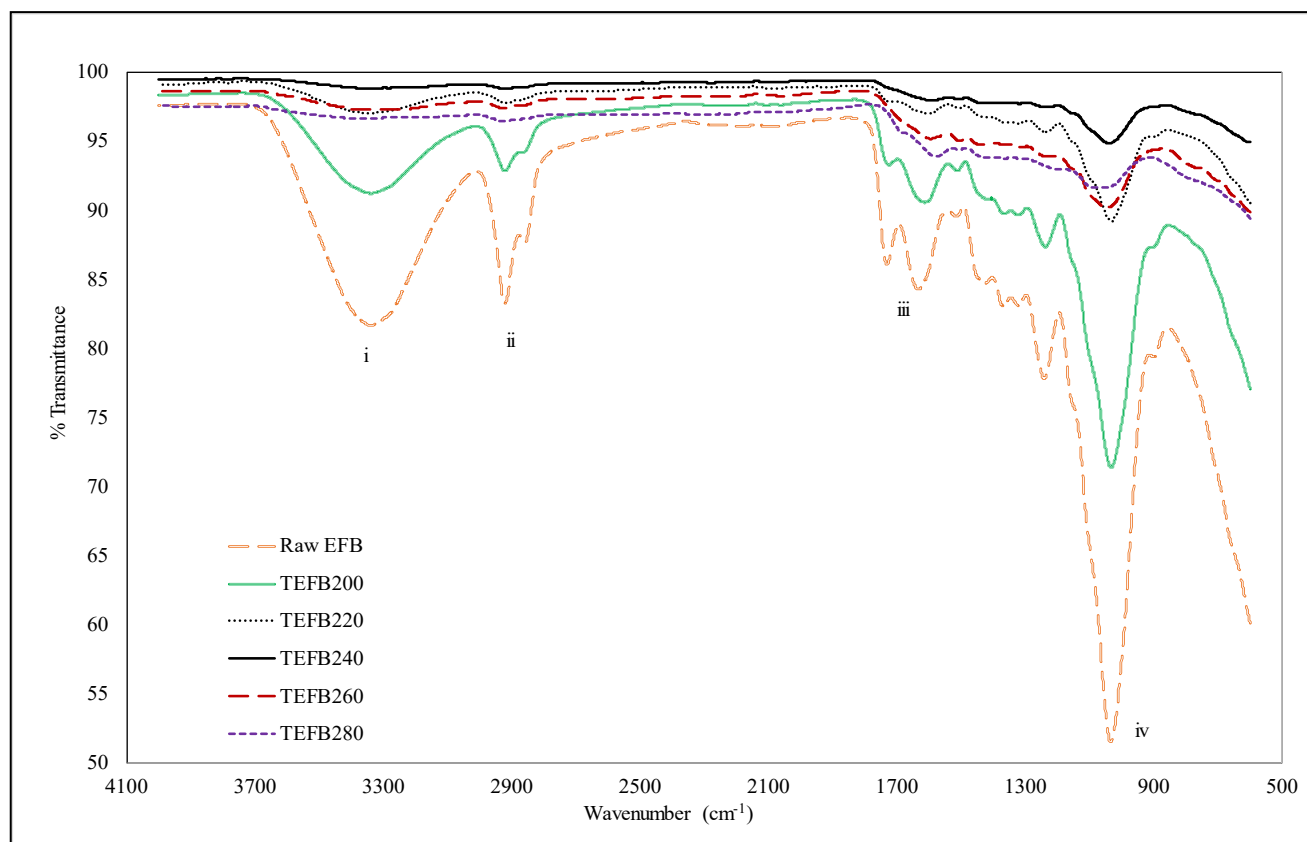
(a)
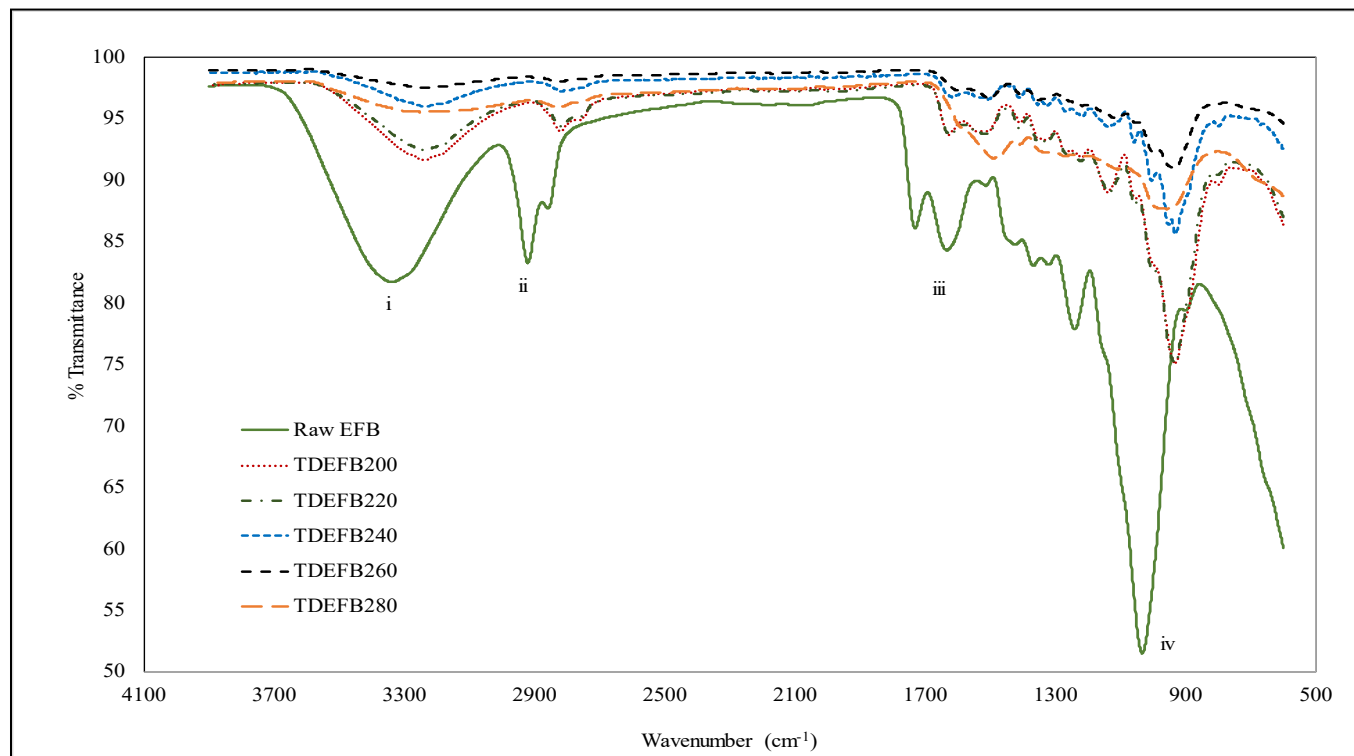
(b)

Figure 2. FTIR spectra of a) EFB and TEFB and, b) EFB and TDEFB at various torrefaction temperatures

and acids, and ethers, phenol and alcohol (C−O, C=C and C−C−O stretching), respectively. The spectra of EFB shows clearly a broad envelope of 3345 cm^{-1} which corresponds to the hydroxyl group. However, this peak was remarkably weakened as the torrefaction temperature was increased due to the removal of moisture through dehydration process.

The peaks at 2920 cm-1 and 2850 cm-1 corresponding to the hydrocarbons and carbonyl species, respectively show similar trend that represents the evolution of carbon dioxide and methane, and extent of the decomposition of hemicellulose for torrefied EFB. Apparently, the intensities of spectral peaks (iii) appearing at 1650 cm-1 and 1710 cm-1 which were assigned to alkene (C=C) and carbonyl (C=O), respectively decreased for both TEFB and TDEFB samples in comparison to EFB. This observation indicates that oxygen and hydrogen have been removed through decarboxylation and dehydration of hemicellulose during torrefaction and demineralisation of EFB. This is in agreement with similar findings reported by other researchers [39, 40].

Previous research [41] reported that torrefaction pre-treatment could reduce the acid content of bio oil due to the removal of acetyl group in hemicellulose which is the primary source of acetic acid. Torrefaction of EFB and DEFB samples above 240 °C decreased the intensity of the peak originating form acetyl group (1710 cm-1), which is observation attributed to deacetylation of hemicellulose. Therefore, torrefaction temperature of 240 °C were selected for TEFB and TDEFB in bio fuel feedstock. This is due to the fact that at this temperature, the percentages of torrefaction yield of TEFB240 and TDEFB240 were found higher than that of EFB and DEFB samples at the torrefaction temperatures of 260 and 280 °C with less acetyl group content

The Van Krevelen diagram can be used to analyse the characteristics of torrefied biomass based on the relative increase in carbon content while decreasing the relative amounts of hydrogen and oxygen [35].

With reference to Figure 3a, the Van Krevelen diagram of TEFB and EFB shows the reduction in elementary composition of oxygen and hydrogen with the increase of torrefaction temperature. Comparable tendency was also observed for TDEFB as shown in Figure 3b. The changes in elementary composition of both torrefied samples resemble sub-bituminous coal referring to the Van Krevelen plot. These results indicate that some alterations in fuel properties [28] that enhance the quality of torrefied EFB as a solid fuel feedstock [34].

3.3. The effect of sequential pre-treatment on fuel properties of EFB, TEFB and TDEFB

The mass and energy yields and energy density were calculated using equations 1, 2 and 3, respectively and the data are illustrated in Figure 4. Indeed, it is seen from Figure 4 that torrefaction pre-treatment had an important impact on the mass and energy yields and energy density. With the increase in torrefaction temperature from 200 to 280 °C, the mass and energy yields for TDEFB declined gradually from 77.88 to 34.96 % and 80.94 to 46.15 %, respectively. The same tendency of reduction was also observed for the mass and energy yields of the TDEFB sample in the range from 86.01 to 45.89 % and from 86.49 % to 60.59 %, respectively.

However, mass and energy yields of TDEFB were slightly higher than that of TEFB. These results may be associated with the trending of volatile matter of proximate analysis as shown in Table 1 in previous section. TEFB suffered from obvious decreases of mass and energy yields specifically at the temperatures above 240 °C. Meanwhile, TDEFB was able to retain a drastic loss of mass and energy yields prior to the pre-treatment. In addition, the energy density of TEFB and TDEFB increased in the range from 1.01 to 1.34 with the increased of the torrefaction temperature from 200 to 280 °C. This was due to the major decomposition of hemicellulose and partial decomposition of cellulose during torrefaction resulting in an increase in calorific value, and is similar to the results reported by other researchers [41,39,32].
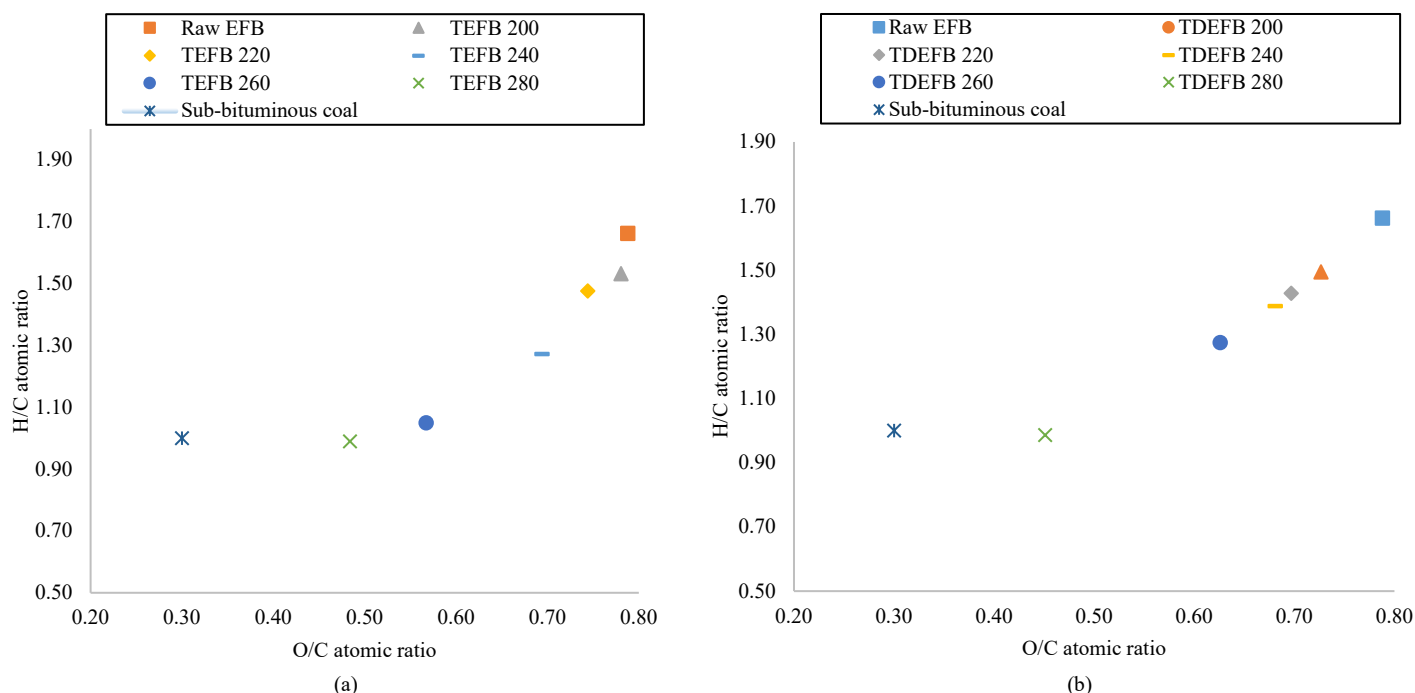
Figure 3. Van Krevelen diagram for chemical composition of a) EFB and TEFB at various torrefaction temperatures and b) EFB and TDEFB at various torrefaction temperatures (sub-bituminous coal as reference)
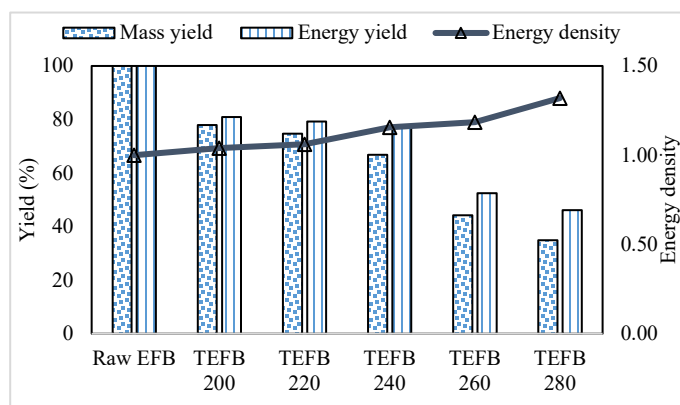
(a)
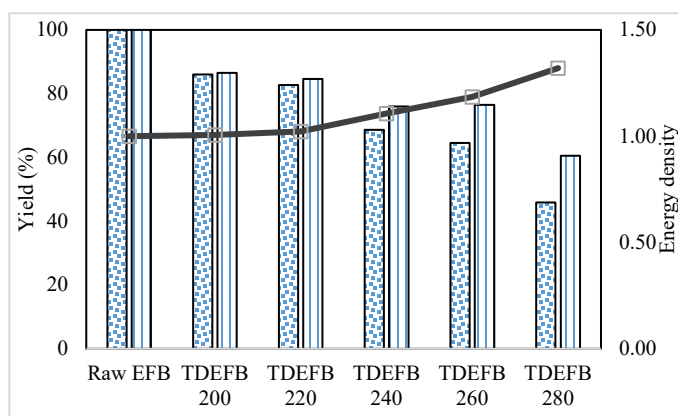
(b)

Figure 4. Mass and energy yield and calorific value a) TEFB and b) TDEFB

According to our previous work, pre-treated palm wastes such as EFB and palm kernel shell (PKS) using physical and chemical pre-treatments were identified the appropriate solid biofuel for further thermal conversions such as combustion, pyrolysis, gasification and other applications. Further, it also promote to the enrichment of energy security, decrease in CO2 emissions, sustenance to the development of renewable energy and high advantageous to the agricultural and forestry activity [44,45].

3.4. Thermogravimetric analysis

Based on the results obtained from proximate and ultimate analyses, fuel properties and FTIR data of EFB, TEFB and TDEFB, both TEFB 240 and TDEFB 240 were obtained following the pyrolysis process. Thermogravimetric analyser (TGA) was used to observe the thermal behaviour and to obtain the kinetic parameters of all samples. The weight loss (TG) and the differential weight loss (DTG) curves of EFB, TEFB 240 and TDEFB 240 are illustrated in Figure 5 and 6, respectively.

From Figure 5, the first weight loss is identified at approximately 128 °C for all samples. This corresponds to the removal of water molecules which are adsorbed on the surface of EFB and is shown by the first small peak in Figure 6. The second peak shows that the major weight loss occurs from 200 to 350 °C for EFB and from 290 to 350 °C for TEFB 240 and TDEFB 240. This could be due to the active pyrolysis of hemicellulose and cellulose portion in EFB. The cellulose and hemicelluloses contents of EFB are higher than those of TEFB 240 and TDEFB 240. Therefore, the active pyrolysis process began at a much lower temperature for EFB compared to TEFB 240 and TDEFB 240. This devolatilization stage of EFB resulted in changes via occurrence of simultaneous reactions such as depolymerization, decarboxylation, dehydration and decarbonylation [6, 42].

The TEFB 240 sample had undergone single pre-treatment of torrefaction at 240 °C whereas the TDEFB 240 sample had experienced sequential demineralization and torrefaction at 240 °C during pre-treatment stage which resulted in the cleavage of C-O bonds in hemicelluloses and decrease in cellulosic content of EFB. Therefore, the active pyrolysis processes of TEFB 240 and TDEFB 240 are contributed by the remaining lignocellulosic content of EFB that shifts the process to a maximum region. It can be seen from Figure 6 that, the peak temperature shifts from 318 to 336 °C first and finally to 354 °C for EFB, TEFB 240 and TDEFB 240, respectively. This trend is similar to previous study [20], where the main weight loss was observed in the range from 250 to 450 °C

with the corresponding DTG curve appearing from 350 to 390 °C and this showed the decomposition temperature regions represented by hemicellulose, cellulose and lignin.
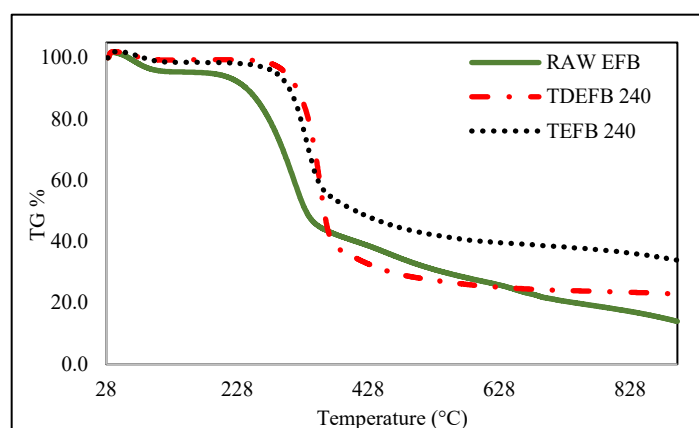
Figure 5. The weight loss (TG) of EFB, TEFB and TDEFB at heating rate of 20 °C min-1 under inert condition
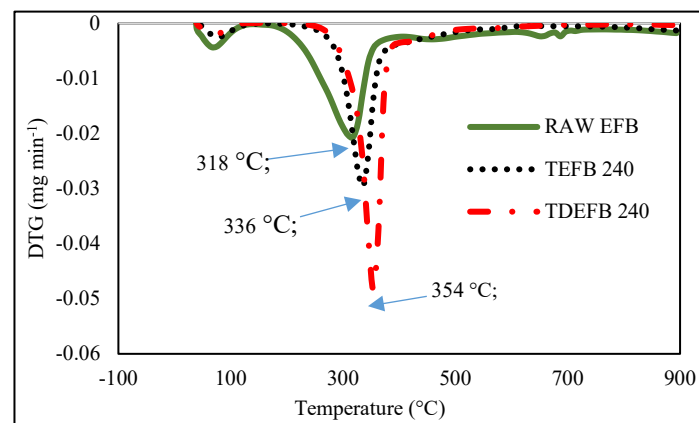
Figure 6. The weight loss rate (DTG) of EFB, TEFB and TDEFB at a heating rate of 20 °C min-1 under inert condition

The third stage is observed at a temperature range from 350 to 600 °C in Figure 5. This is a slow weight loss taking place over a wide range of temperature and is due to the composition of lignin in the samples. In summary, three temperature regions are clearly seen for decomposition of lignocellulosic of raw EFB, TEFB 240 and TDEFB 240

Table 3. The kinetic parameters for raw EFB, TEFB 240 and TDEFB 240 pyrolysis obtained by Coats-Redfern model

Samples	Equation	Activation energy, E_a (kJ mol^{-1})	R^2	Pre-exponential factor, ln A (min^{-1})
EFB	y = -8588x +1.8068	71.40	0.9956	10.87
TEFB 240	y = -14573x + 10.81	121.16	0.9898	20.40
TDEFB 240	y = -16592x + 13.419	137.42	0.9888	23.13

3.5. Thermal degradation kinetic analysis

The thermogravimetric data obtained at a heating rate of 20°C min^{-1} were used to calculate the kinetics parameters such as activation energy and pre-exponential factor, ln A for EFB, TEFB 240 and TDEFB 240 using the Coats-Redfern method of the data obtained in the temperature range from 200 to 400 °C. The plot of ln (-ln(1-α)/T^2) vs. 1/T is illustrated in Figure. 7. The corresponding values of activation energy (E$_a$), correlation coefficient (R^2) and intercepts are presented in Table 3.
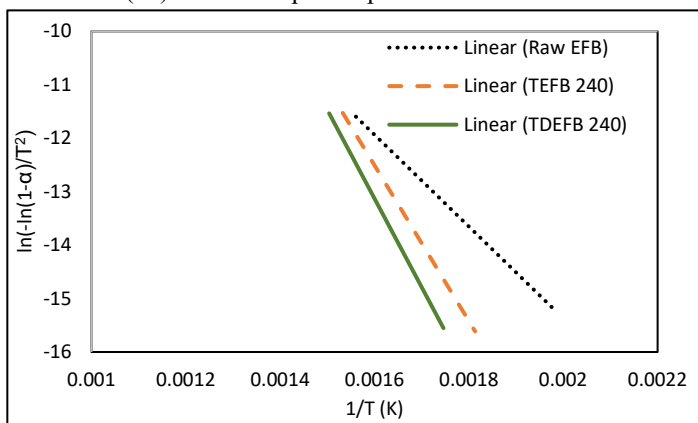
Figure 7. The plot of regression line was produced from value of ln[-ln(1-α)/T2 against 1/T for EFB, TEFB 240 and TDEFB 240

The values of R^2 are above 0.96 indicating the present of a strong correlation between dependable and undependable parameters. From Table 3, the activation energy values for pyrolysis of TDEFB 240 and TEFB 240 were determined as 137.42 and 121.16 kJ mol^{-1}, respectively. These values are higher than that for pyrolysis of EFB which is 71.40 kJ mol^{-1}. The low activation energy of EFB pyrolysis is attributed to the devolatilization of hemicellulosic component [47].

The higher values of activation energy for pyrolysis of torrefied samples may be explained by the fact that pyrolysis of torrefied samples containing mainly C-C networks occurs at higher temperature which can be accompanied by a higher activation energy [20]. During torrefaction of TEFB 240 and TDEFB 240, the hemicellulosic component decomposes and results in the cleavage of C-O and C-H bonds. The removal of oxygen atoms occurs also during this stage either in the form of water or carbon monoxide.

Conclusion

In this work, the effects of sequential pre-treatment consist of demineralization and torrefaction on oil palm empty fruit bunch have been studied. This pre-treatment effectively enhances the quality of bio fuel feedstock by increasing the carbon content, eliminating the AAEMs content and reducing oxygen content. It was found that, this sequential pre-treatment improves the energy density of the selected torrefied samples by decreasing oxygen content which contributed in increasing the calorific value. This results also can also be supported by decreasing the intensity of hydroxy and carbonyl peaks that was observed in FTIR spectra of the samples. According to thermogravimetric analysis, both samples showed the increasing in C-C networks that represent the removal of hemicellulose and rich in cellulose content. These results can be supported by increasing the activation energy of TEFB 240 and TDEFB 240 using Coats-Redfern method of kinetic study. Thus, this sequential pre-treatment is suggested to be the most effective approach for upgrading the quality of solid fuel feedstock for further thermal conversion process for the future.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] N. N. Kasim, K. Ismail, M. A. M. Ishak, R. Ahmad, A. R. Mohamed, and W. I. Nawawi, "Demineralization of oil palm empty fruit bunch (EFB) intended as a high quality bio-oil feedstock," 4th IET Clean Energy Technol. Conf. (CEAT 2016), 107 (6.), 2016. https://doi:10.1049/cp.2016.1364.

[2] B. R. Singh and O. Singh, "Global Trends of Fossil Fuel Reserves and Climate Change in the 21st Century," Foss. Fuel Environ., 167–192, 2012. https://doi:10.5772/38655.

[3] M. V. Kok and E. Özgür, "Thermal analysis and kinetics of biomass samples," Fuel Process. Technol.,106, 739–743, 2013. https://doi:10.1016/j.fuproc.2012.10.010.

[4] N. Abdullah and F. Sulaiman, "The oil palm wastes in Malaysia," in Biomass Now – Sustainable Growth and Use, 75–100, 2013. https://doi:10.5772/55302.

[5] A. R. Mohamed and Z. Hamzah, "An alternative approach for the screening of catalytic empty fruit bunch (EFB) pyrolysis using the values of activation energy from a thermogravimetric study," React. Kinet. Mech. Catal.,114, 529–545, 2014. https://doi:10.1007/s11144-014-0798-8.

[6] X. Y. Lim and J. M. Andrésen, "Pyro-catalytic deoxygenated bio-oil from palm oil empty fruit bunch and fronds with boric oxide in a fixed-bed reactor," Fuel Process. Technol., 92(9), 1796–1804, 2011. https://doi:10.1016/j.fuproc.2011.04.033.

[7] M. a. Aziz, K. M. Sabil, Y. Uemura, and L. Ismail, "A study on torrefaction of oil palm biomass," Journal of Applied Sciences, 12(11). 1130–1135, 2012. https://doi:10.3923/jas.2012.1130.1135.

[8] J. Mabrouki, M. A. Abbassi, K. Guedri, A. Omri, and M. Jeguirim, "Simulation of biofuel production via fast pyrolysis of palm oil residues," Fuel,159,819–827, 2015. https://doi:10.1016/j.fuel.2015.07.043.

[9] Y. Uemura, W. N. Omar, T. Tsutsui, and S. B. Yusup, "Torrefaction of oil palm wastes," Fuel, 90(8), 2585–2591, 2011. https://doi:10.1016/j.fuel.2011.03.021.

[10] S. Matali, N. A. Rahman, S. S. Idris, N. Yaacob, and A. B. Alias, "Lignocellulosic Biomass Solid Fuel Properties Enhancement via Torrefaction," Procedia Eng., 148, 671–678, 2016.

https://doi:10.1016/j.proeng.2016.06.550.

[11] T. Wigley, A. C. K. Yip, and S. Pang, "The use of demineralisation and torrefaction to improve the properties of biomass intended as a feedstock for fast pyrolysis," J. Anal. Appl. Pyrolysis, 113, 296–306, 2015. https://doi:10.1016/j.jaap.2015.02.007.

[12] T. Wigley, A. C. K. Yip, and S. Pang, "Pretreating biomass via demineralisation and torrefaction to improve the quality of crude pyrolysis oil," Energy, 109, 481–494, 2016. https://doi:10.1016/j.energy.2016.04.096.

[13] N. Abdullah, F. Sulaiman, and H. Gerhauser, "Characterisation of oil palm empty fruit bunches for fuel application," J. Phys. Sci., 22(1), 1–24, 2011.

[14] K. Kaminaka, Y. Matsumura, W. Noaman Omar, and Y. Uemura, "Process evaluation for torrefaction of empty fruit bunch in malaysia," J. Japan Pet. Inst., 57(2), 88–93, 2014.

[15] Y. Uemura, W. Omar, N. A. Othman, S. Yusup, and T. Tsutsui, "Torrefaction of oil palm EFB in the presence of oxygen," Fuel, 103, 156–160, 2013. https://doi:10.1016/j.fuel.2011.11.018.

[16] A. Shariff, N. S. M. Aziz, and N. Abdullah, "Slow Pyrolysis of Oil Palm Empty Fruit Bunches for Biochar Production and Characterisation," J. Phys. Sci., 25(2), 97–112, 2014.

[17] S. Zhang, Q. Dong, L. Zhang, Y. Xiong, X. Liu, and S. Zhu, "Effects of water washing and torrefaction pretreatments on rice husk pyrolysis by microwave heating.," Bioresour. Technol., 193, 442–448, 2015. https://doi:10.1016/j.biortech.2015.06.142.

[18] J. M. Reckamp, R. A. Garrido, and J. A. Satrio, "Selective pyrolysis of paper mill sludge by using pretreatment processes to enhance the quality of bio-oil and biochar products," Biomass and Bioenergy, 71, 235–244, 2014. https://doi:10.1016/j.biombioe.2014.10.003

[19] D. Chen, J. Mei, H. Li, Y. Li, M. Lu, and Z. Ma, "Combined pretreatment with torrefaction and washing using torrefaction liquid products to yield upgraded biomass and pyrolysis products," Bioresour. Technol., 228, 62–68, 2017. https://doi:10.1016/j.biortech.2016.12.088.

[20] L. Cao, X. Yuan, L. Jiang, C. Li, Z. Xiao, Z. Huang, X. Chen, G. Zeng, and H. Li, "Thermogravimetric characteristics and kinetics analysis of oil cake and torrefied biomass blends," Fuel, 175, 129–136, 2016. https://doi:10.1016/j.fuel.2016.01.089.

[21] M. Kopczyński, A. Plis, and J. Zuwała, "Thermogravimetric and kinetic analysis of raw and torrefied biomass combustion," Chem. Process Eng. - Inz. Chem. i Proces., 36(2), 209–223, 2015. https://doi:10.1515/cpe-2015-0014.

[22] P. Ninduangdee, V. I. Kuprianov, E. Y. Cha, R. Kaewrath, P. Youngyuen, and W. Atthawethworawuth, Thermogravimetric Studies of Oil Palm Empty Fruit Bunch and Palm Kernel Shell: TG/DTG Analysis and Modeling, 79, 2015. https://doi:10.1016/j.egypro.2015.11.518.

[23] G. Wang, J. Zhang, J. Shao, Z. Liu, G. Zhang, T. Xu, J. Guo, H. Wang, R. Xu, and H. Lin, "Thermal behavior and kinetic analysis of co-combustion of waste biomass/low rank coal blends," Energy Convers. Manag., 124, 414–426, 2016. https://doi:10.1016/j.enconman.2016.07.045.

[24] S. D. Stefanidis, E. Heracleous, D. T. Patiaka, K. G. Kalogiannis, C. M. Michailof, and A. A. Lappas, "Optimization of bio-oil yields by demineralization of low quality biomass," Biomass and Bioenergy, 83, 105–115, 2015. https://doi:10.1016/j.biombioe.2015.09.004.

[25] P. Basu and P. Basu, "Chapter 13 – Analytical Techniques," in Biomass Gasification, Pyrolysis and Torrefaction, 439–455, 2013. https://doi:10.1016/B978-0-12-396488-5.00013-7.

[26] S. S. Idris, N. A. Rahman, K. Ismail, A. B. Alias, Z. A. Rashid, and M. J. Aris, "Investigation on thermochemical behaviour of low rank Malaysian coal, oil palm biomass and their blends during pyrolysis via thermogravimetric analysis (TGA)," Bioresour. Technol., 101(12), 4584–4592, 2010. https://doi:10.1016/j.biortech.2010.01.059.

[27] K. Singh and J. Zondlo, "Characterization of fuel properties for coal and torrefied biomass mixtures," J. Energy Inst., 2016. https://doi:10.1016/j.joei.2016.05.012.

[28] T. G. Bridgeman, J. M. Jones, A. Williams, and D. J. Waldron, "An investigation of the grindability of two torrefied energy crops," Fuel, 89,(12), 3911–3918, 2010. https://doi:10.1016/j.fuel.2010.06.043.

[29] B. Bevan, A. Ahmad, A. Johari, T. Amran, and T. Abdullah, "Torrefaction of Pelletized Oil Palm Empty Fruit Bunches," in The 21st International Symposium on Alcohol Fuels – 21st ISAF, 15–19, 2015.

[30] M. A. Islam, M. Asif, and B. H. Hameed, "Pyrolysis kinetics of raw and hydrothermally carbonized Kar (Pongamia pinnata) fruit hulls via thermogravimetric analysis," Bioresour. Technol., 179, 227–233, 2015. https://doi:10.1016/j.biortech.2014.11.115.

[31] A. R. Mohamed, Z. Hamzah, M. Z. M. Daud, and Z. Zakaria, "The Effects of Holding Time and the Sweeping Nitrogen Gas Flowrates on the Pyrolysis of EFB using a Fixed–Bed Reactor," Procedia Eng., 53, 185–191, 2013.

https://doi:10.1016/j.proeng.2013.02.024.

[32] S. Ren, H. Lei, L. Wang, Q. Bu, S. Chen, and J. Wu, "Thermal behaviour and kinetic study for woody biomass torrefaction and torrefied biomass pyrolysis by TGA," Biosyst. Eng., 116(4), 420–426, 2013. https://doi:10.1016/j.biosystemseng.2013.10.003.

[33] J. Meng, J. Park, D. Tilotta, and S. Park, "The effect of torrefaction on the chemistry of fast-pyrolysis bio-oil.," Bioresour. Technol., 111, 439–446, 2012. https://doi:10.1016/j.biortech.2012.01.159.

[34] Y. Yue, H. Singh, B. Singh, and S. Mani, "Torrefaction of Sorghum Biomass to Improve Fuel Properties," Bioresour. Technol., 232, 372–379, 2017. https://doi:10.1016/j.biortech.2017.02.060.

[35] W.-H. Chen, J. Peng, and X. T. Bi, "A state-of-the-art review of biomass torrefaction, densification and applications," Renew. Sustain. Energy Rev., 44, 847–866, 2015. https://doi:10.1016/j.rser.2014.12.039.

[36] A. Saddawi, J. M. Jones, and A. Williams, "Influence of alkali metals on the kinetics of the thermal decomposition of biomass," Fuel Process. Technol., 104, 189–197, 2012. https://doi:10.1016/j.fuproc.2012.05.014.

[37] P. Basu, Torrefaction. Elsevier Inc., 2013. https://doi:10.1016/B978-0-12-396488-5.00004-6.

[38] S. Zhang, T. Chen, W. Li, Q. Dong, and Y. Xiong, "Bioresource Technology Physicochemical properties and combustion behavior of duckweed during wet torrefaction", 218, 1157–1162, 2016.http://10.1016/j.enconman.2016.07.045

[39] W.-H. Chen, S.-W. Du, C.-H. Tsai, and Z.-Y. Wang, "Torrefied biomasses in a drop tube furnace to evaluate their utility in blast furnaces," Bioresour. Technol., 111, 433–438, 2012. https://doi:10.1016/j.biortech.2012.01.163.

[40] A. Zheng, Z. Zhao, Z. Huang, K. Zhao, G. Wei, X. Wang, F. He, and H. Li, "Catalytic fast pyrolysis of biomass pretreated by torrefaction with varying severity," Energy & Fuels, 28, 5804–5811, 2014. https://doi:10.1021/ef500892k.

[41] S. Wang, G. Dai, B. Ru, Y. Zhao, X. Wang, J. Zhou, Z. Luo, and K. Cen, "Effects of torrefaction on hemicellulose structural characteristics and pyrolysis behaviors," Bioresour. Technol., 218, 1106–1114, 2016. https://doi:10.1016/j.biortech.2017.02.120.

[42] Y. Joshi, M. Di Marcello, and W. de Jong, "Torrefaction: Mechanistic study of constituent transformations in herbaceous biomass," J. Anal. Appl. Pyrolysis, 115, 353–361, 2015. https://doi:10.1016/j.jaap.2016.07.075.

[43] D. Chen, K. Cen, X. Jing, J. Gao, C. Li, and Z. Ma, "An approach for upgrading biomass and pyrolysis product quality using a combination of aqueous phase bio-oil washing and torrefaction pretreatment," Bioresour. Technol., 233, 150–158, 2017.

[44] N. N. Kasim, K. Ismail, M. A. M. Ishak, R. Ahmad, A. R. Mohamed, and W. I. Nawawi, "Palm deto : resourceful , sustainable and favorable solid biofuel," 2, 391–397, 2017.

[45] R. Ahmad, M. A. M. Ishak, K. Ismail, and N. N. Kasim, " Co- palm: versatile , clean and enriched energy fuel," 2, 327–331, 2017.

[46] S. S. Mohtar, T. N. Z. Tengku Malim Busu, A. M. Md Noor, N. Shaari, and H. Mat, "An ionic liquid treatment and fractionation of cellulose, hemicellulose and lignin from oil palm empty fruit bunch," Carbohydr. Polym., 166, 291–299, 2017. https://doi:10.1016/j.carbpol.2017.02.102.

[47] H. Yang, R. Yan, H. Chen, D. H. Lee, and C. Zheng, "Characteristics of hemicellulose, cellulose and lignin pyrolysis," Fuel,86(12–13), 1781–1788, 2007. https://doi:10.1016/j.fuel.2006.12.013.

A S T E S

NonLinear Control via Input-Output Feedback Linearization of a Robot Manipulator

Wafa Ghozlane*, Jilani Knani

Laboratory of Automatic Research LA.R.A, Department of Electrical Engineering, National School of Engineers of Tunis ENIT, University of Tunis ElManar, 1002 Tunis, Tunisia.

A B S T R A C T

This paper presents the input-output feedback linearization and decoupling algorithm for control of nonlinear Multi-input Multi-output MIMO systems. The studied analysis was motivated through its application to a robot manipulator with six degrees of freedom. The nonlinear MIMO system was transformed into six independent single-input single-output SISO linear local systems. We added PD linear controller to each subsystem for purposes of stabilization and tracking reference trajectories, the obtained results in different simulations shown that this technique has been successfully implemented.

1. Introduction

In recent years, Feedback linearization has been attracted a great deal of interesting research. It's an approach designed to the nonlinear control systems, which based on the idea of transforming nonlinear dynamics into a linear form. The base idea of this technique is to algebraically transform a nonlinear dynamics system into a totally or partially linear one, so that linear control techniques can be applied. This notion can be used for both stabilization and tracking control objectives of SISO or MIMO systems, and has been successfully applied to a number of practical nonlinear control problems such as [1-4].

In fact, this technique has been successfully implemented in several faisable applications of control, such as industrial robots, high performance aircraft, helicopters and biomedical dispositifs, more tasks used the methodology are being now well advanced in industry [5-6].

In this case, we applied this technique to lead the control for each joint of a robot manipulator that is has six degrees of freedom, which the equations of motion form a nonlinear, complex dynamic and multivariable system, then, we elaborated a PD linear controller for each decoupled linear subsystem to control

the angular position of each joint of this robot arm for stabilization and tracking purposes. The obtained results in different simulations shown the efficiency of the derived approach [7].

This paper is organized as follows: It is divided into five sections. In Section 2, a description of the input-output feedback linearization approach is detailed. In Section 3, a simplified dynamic model of a robot manipulator with six degrees of freedom is presented, the input-output feedback linearization method is applicated to the above robot and the construction of linear PD controller is derived. In Section 4, the simulation results are presented. Finally, the conclusion was elaborated in Section 5.

2. Input-Output feedback linearization for MIMO nonlinear system.

In this section, we discussed the approach of input-output feedback linearization of nonlinear systems, the central goal of feedback linearization is to design a nonlinear control law as assumed that the inner loop control is, in the most suitable case, precisely linearizes the nonlinear system after appropriate state space modification of coordinates [1]. The developer can then build an outer loop control in the new coordinates to obtain a linear relation between the output Y and the input V and to satisfy the traditional control design specifications such as tracking, disturbance rejection, as shown in Figure 1.

*Corresponding Author: Wafa Ghozlane, Email: ghozlanewafa@yahoo.fr
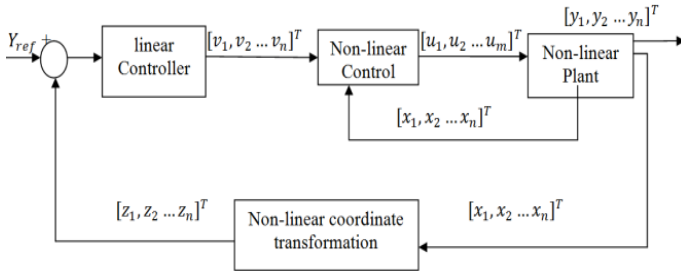
Figure 1. Structure of Input-Output Feedback Linearization Approach.

The basic condition for using feedback linearization method is nonlinear dynamic MIMO of n -order with p number of inputs and outputs described in the affine form;

$$\begin{cases} \dot{X}(t) = f\big(X(t)\big) + \sum_{i=1}^{p} g_i\big(X(t)\big)U_i(t) \\ Y_i(t) = h_i\big(X(t)\big) \\ i = 1,2,\dots p \end{cases} \quad (1)$$

Where,

$X = [x_1, x_2 \dots x_n]^T \in R^n$: is the state vector.

$U = [u_1, u_2 \dots u_p]^T \in R^p$: is the control input vector.

$Y = [y_1, y_2 \dots y_p]^T \in R^p$: is the output vector, $f(X)$, and $g_i(X)$: are n-dimentional smooth vector fields.

$h_i(X)$: is smooth nonlinear functions, with i=1,2...n.

Theorem1:

Let f: $\mathcal{R}^n \to \mathcal{R}^n$ represent a smooth vector field on $\mathcal{R}^n$ and let

h : $\mathcal{R}^n \to \mathcal{R}^n$ represent a scalar function. The Lie Derivative of h, with respect to f, denoted $L_f h$, is defined as [1-2].

$$L_f h = \frac{\partial h}{\partial x} f(x) = \sum_{i=1}^{p} \frac{\partial h}{\partial x_i} f_i(x) \quad (2)$$

The Lie derivative is the directional derivative of h in the direction of f(x), in an equivalent way, the inner product of the gradient of h and f. We defined by $L_f^2 h$ the Lie Derivative of $L_f h$ with respect to f:

$$L_f^2 h = L_f (L_f h) \quad (3)$$

In general we define:

$$L_f^k h = L_f(L_f^{k-1} h) \quad for\ k = 1,\dots,p \quad (4)$$

with $L_f^0 h = h$

Theorem2:

The function $\Phi: \mathcal{R}^n \to \mathcal{R}^n$ defined in a region $\Omega \subset \mathcal{R}^n$ he is called difeomorphisme if it checks the following conditions:

Firstly, a diffeomorphism is a differentiable function whose inverse exists and is also differentiable. Second, we should assume that both the function and its inverse to be infinitely differentiable, such functions are usually referred to as $\mathbb{C}\infty$ diffeomorphisms [3-4].

The diffeomorphism is used to transform one nonlinear system in another nonlinear system by making a change of variables of the form:

$$z = \Phi(x) \quad (5)$$

Where $\Phi(x)$ represents n variables;

$$\Phi(x) = \begin{bmatrix} \Phi_1 \\ \Phi_2 \\ \vdots \\ \Phi_n \end{bmatrix} = \begin{bmatrix} [h_1 \quad L_f h_1 \quad \dots \quad L_f^{r_1-1} h_1]^T \\ \vdots \\ [h_p \quad L_f h_p \quad \dots \quad L_f^{r_p-1} h_p]^T \end{bmatrix}, \quad (6)$$

$$x = [x_1, x_2 \dots x_n]^T$$

The goal is to obtain a linear relation between the inputs and the outputs by differentiating the outputs y_j until the inputs appear. Suppose that r_j is the smallest integer such that fully one of the inputs appears in $y_j^{(r_j)}$ using this expression:

$$y_j^{(r_j)} = L_f^{r_j} h_j(x) + \sum_{i=1}^{p} L_{g_i}(L_f^{(r_j-1)} h_j(x)) u_i \quad (7)$$

$$i, j = 1,2, \dots p$$

Where, $L_f^i h_j$ and $L_g^i h_j$: Are the ith Lie derivatives of $h_j(x)$ respectively in the direction of f and g.

$$L_f h_j(x) = \frac{\partial h_j}{\partial x} f(x), \ L_g h_j(x) = \frac{\partial h_j}{\partial x} g_i(x) \quad (8)$$

r_j: is the relative degree corresponding to the y_j output, it's the number of necessary derivatives so that at least one of the inputs appear in the expression [5].

If expression $L_g h_j(x) = 0$, for all i, then the inputs have not appeared in the derivation and it's necessary to continu the derivation of the output y_j.

The system (1) has the relative degree (r) if it satisfies:

$$\begin{cases} L_{g_i} L_f^k h_j = 0 \quad 0 < k < r_{j-1}, 0 \le i \le n, 0 \le j \le n \quad (9) \\ L_{g_i} L_f^k h_j \ne 0 \quad\quad k = r_{j-1} \end{cases}$$

The total relative degree (r) was considered as the sum of all the relative degrees obtained using (7) and must be less than or equal to the order of the system (10):

$$r = \sum_{j=1}^{n} r_j \le n \quad (10)$$

To find the expression of the nonlinear control law U that allows to make the relationship linear between the input and the output [6], the expression (2) is rewritten in matrix form as:

$$\left[y_1{}^{r_1} \dots y_p{}^{r_p} \right]^T = \propto (x) + \beta(x).U \qquad (11)$$

$$V = \left[v_1\, v_2 \dots v_p \right]^T = \left[y_1{}^{r_1} \dots y_p{}^{r_p} \right]^T \qquad (12)$$

Where:

$$\propto (x) = \begin{bmatrix} L_f{}^{r_1} h_1(x) \\ . \\ . \\ . \\ L_f{}^{r_p} h_p(x) \end{bmatrix} \qquad (13)$$

$$\beta(x) = \begin{bmatrix} L_{g_1}(L_f{}^{(r_1-1)}h_1(x)) & L_{g_2}(L_f{}^{(r_1-1)}h_1(x)) & \dots & L_{g_p}(L_f{}^{(r_1-1)}h_1(x)) \\ L_{g_1}(L_f{}^{(r_2-1)}h_2(x)) & L_{g_2}(L_f{}^{(r_2-1)}h_2(x)) & \dots & L_{g_p}(L_f{}^{(r_2-1)}h_2(x)) \\ . & . & \dots & . \\ . & . & & . \\ . & . & & . \\ L_{g_1}(L_f{}^{(r_p-1)}h_p(x)) & L_{g_2}(L_f{}^{(r_p-1)}h_p(x)) & \dots & L_{g_p}(L_f{}^{(r_p-1)}h_p(x)) \end{bmatrix} \quad (14)$$

If $\beta(x)$ is not singular, then it is possible to define the input transformation "the nonlinear control law " which has this form:

$$U = \beta(x)^{-1}. (- \propto (x) + V) \qquad (15)$$

$$V = \left[v_1\, v_2 \dots v_p \right]^T$$

$$U = \left[u_1\, u_2 \dots u_p \right]^T$$

$$y_j{}^{r_j} = v_j$$

where

V: Is the new input vector. is called a decoupling control law

$\beta(x)$: Is the invertible (pxp) matrix. Is called a decoupling matrix of the system.

2.1. Non-linear coordinate transformation:

$$Z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} = \begin{bmatrix} \Phi_1 \\ \Phi_2 \\ \vdots \\ \Phi_n \end{bmatrix} = \begin{bmatrix} \left[h_1 & L_f h_1 & \dots & L_f{}^{r_1-1}h_1 \right]^T \\ \vdots \\ \left[h_p & L_f h_p & \dots & L_f{}^{r_p-1}h_p \right]^T \end{bmatrix} \qquad (16)$$

By applying the linearizing law to the system, we can transforme the nonlinear system into linear form [7-8]:

$$\begin{cases} \dot{Z} = Az + BV \\ \quad Y = CZ \end{cases} \qquad (17)$$

With,

$$A = \begin{bmatrix} A_{r_1} & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & A_{r_p} \end{bmatrix}, B = \begin{bmatrix} B_{r_1} & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & B_{r_p} \end{bmatrix},$$

$$C = \begin{bmatrix} C_{r_1} & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & C_{r_p} \end{bmatrix}$$

And,

$$A_{r_i} = \begin{bmatrix} 0 & 1 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 0 \end{bmatrix} \in \mathcal{R}^{r_i \times r_i}; \; B_{r_i} = \begin{bmatrix} 0 \\ \vdots \\ 1 \end{bmatrix} \in \mathcal{R}^{r_i};$$

$$C_{r_i} = [1 \quad 0 \quad \dots \quad 0] \in \mathcal{R}^{r_i}$$

2.2. Design of the new control vector V:

The vector v is designed to according the control objectives, for the tracking problem considered, it must satisfy:

$$v_j = y_{d_j}{}^{r_j} + K_{r_{j-1}} \left(y_{d_j}{}^{r_{j-1}} - y_j{}^{r_{j-1}} \right) + \cdots$$
$$+ K_1 \left(y_{d_j} - y_j \right); \qquad (18)$$
$$1 \le j \le p$$

where,

$\left\{ y_{d_j}, y_{d_j}{}^2, \dots, y_{d_j}{}^{r_{j-1}}, y_{d_j}{}^{r_j} \right\}$ denote the imposed reference trajectories for the different outputs. If the K_i are chosen so that the polynomial [9-10];
$s^{r_j} + K_{r_{j-1}} s^{r_j-1} + \cdots + K_2 s + K_1 = 0$ are Hurwitz (has roots with negative real parts). Then it can be shown that the error $e_j(t) = y_{d_j}(t) - y_j(t)$, satisfied $\lim\limits_{t \to \infty} e_j(t) = 0$.

3. Input-Output feedback linearization approach applied to a robot manipulator with six degrees of freedom

3.1. Dynamic modeling of a robot manipulator

In this section, we have applied the proposed approach to a dynamic multivariable system which represent a robot arm with six degrees of freedom "EPSON C4". This is an open chain kinematic manipulator robot consisting of seven rigid bodies interconnected by six revolute joints n = 6 as [11]. So, deriving the motion of robot is a complex task due to the nonlinearities present in this system and the large number of degrees of freedom [12-13]. Then, it is essential to understand exactly the dynamics of this interconnected chain of rigid bodies, to detemine the inverse dynamics model such as relation (19), we analyzed the evolution of motion of this mechanical non linear system by using the Euler-Lagrange equations, which is represented by the equation (20).

$$\Gamma = f(q, \dot{q}, \ddot{q}, f_e) \qquad (19)$$

$$\Gamma_i = \sum_{j=1}^{n} d \frac{\left(\frac{\partial L_j}{\partial \dot{q}_i} \right)}{dt} - \frac{\partial L_j}{\partial q_i} \quad i, j = 1, \dots, n \qquad (20)$$

where Γ, q, $\dot{q}$, $\ddot{q}$ and f_e depicting Torques, articular positions, velocities, accelerations and the external force.

L_j: Defines the lagrangian of the jth link, which is the difference of the kinetic and potential energy, equal to E_j - U_j.

E_j and U_j: Define the kinetic and the potential energies of the jth link.

The calculation of the kinetic energy;

In this section, we have calculated the total kinetic energy of the system which depends on the configuration and joint velocities, such that [13], then, it was described by the equation (21).

$$E = \sum_{j=1}^{n} E_j \qquad (21)$$

E_j : means the kinetic energy of the link C_j, which can be formulated such that (24). Firstly, we have calcuated the linear velocity and the angular velocity using the equations (22) and (23).

$$_j^j V = \,_{j-1}^{j}A(_{j-1}^{j-1}V + \,_{j-1}^{j-1}W \times P_j^{j-1}) + \sigma_j \dot{q}_j a_j \qquad (22)$$

$$_j^j W = \,_{j-1}^{j}A \times \,_{j-1}^{j-1}W + \bar{\sigma}_j \times \dot{q}_j \times a_j \qquad (23)$$

$_{j-1}^{j-1}V$: The linear velocity, it is the derivative of the position vector P_j^{j-1}.
$_{j-1}^{j-1}W$: The angular velocity.
The initial conditions for a robot which the base is fixed, are

$_0^0 V = 0$ and $_0^0 W = 0$.

Then,we have expressed these relations (22) and (23) in Equation (24) as.

$$E_j = \frac{1}{2}\left[_j^j W^T \,_j^j J \,_j^j W + M_j \,_j^j V^T \,_j^j V + 2MS_j^T (_j^j V \wedge \,_j^j W)\right] \qquad (24)$$

where

a_j :is the unit vector along axis z_j.

M_j, $_j^j MS$ et $_j^j J$:are the inertial standard parameters.

M_j : is the mass of link C_j.

$_j^j MS$:design the first moments of inertia of link C_j about the origin of the frame R_j It is equal to $_j^j MS = [\ MX_j\ MY_j\ MZ_j\ ]^T$.

$_j^j J$: is the inertial tensor matrix (3x3) of link C_j with respect to the frame R_j, it is expressed by the matrix (25).

$$_j^j J == \begin{bmatrix} I_{XX_j} & I_{XY_j} & I_{XZ_j} \\ I_{XZ_j} & I_{YY_j} & I_{YZ_j} \\ I_{XZ_j} & I_{YZ_j} & I_{ZZ_j} \end{bmatrix} \qquad (25)$$

I_{XX_j}, I_{XY_j}, I_{XZ_j}, I_{YY_j}, I_{YZ_j} and I_{ZZ_j} represent the elements of the inertial tensor the symetric matrix $_j^j J$ of each link C_j which is expressed by the matrix (25), we have defined all these inertial parameters values in our recent work [14].

The calculation of the potential energy;

In this section, we have represented the potential energy for a manipulator arm [14], which is written by the equation (26).

$$U = \sum_{j=1}^{n} U_j \qquad (26)$$

U_j: Defines the potential energy of the link C_j, which is expressed by the equation (27):

$$U_j = -g_0^T\left(M_j \times \,_j^0 P + \,_j^0 A \times MS_j\right)$$
$$= -[g_0^T \ 0] \times T_{0j} \times \begin{bmatrix} MS_j \\ M_j \end{bmatrix} \qquad (27)$$

Then, we have followed the Euler-Lagrange formalism such that equation (20), we obtained the following relation (28):

$$\Gamma = A(q)\ddot{q} + C(q, \dot{q})\dot{q} + Q(q) \qquad (28)$$

where

A(q): Represents the matrix of kinetic energy (n × n), these elements are calculated as follows:

$-A_{ii}$ is equal to the coefficient of $\frac{\dot{q}_i^2}{2}$ located in the expression of the kinetic energy.

$-A_{ij}$ is equal to the coefficient of $\dot{q}_i \dot{q}_j$

$C(q, \dot{q})\dot{q}$: Defines the vector of coriolis and centrifugal forces/torques (n × 1), these elements are calculated from *the Christoffel symbol $c_{i,jk}$* such as system (29):

$$\begin{cases} c_{ij} = \sum_{k=1}^{n} c_{i,jk}\dot{q}_k \\ c_{i,jk} = \frac{1}{2}\left[\frac{\partial A_{ij}}{\partial q_k} + \frac{\partial A_{ik}}{\partial q_j} - \frac{\partial A_{jk}}{\partial q_i}\right] \end{cases} \qquad (29)$$

Q(q): Represents the vector of torques/forces of gravity, these elements are calculated as: $Q_i = \frac{\partial U}{\partial q_i}$

The elements of A, C and Q are according to the geometric and inertial parameters of the mechanism. The dynamic equations of the robot form a system of n differential equations of the second order, coupled and nonlinear. Then, since the inertia matrix A is invertible for $q \in \mathcal{R}^n$ we may solve for the acceleration $\ddot{q}$ of the manipulator as [14].

$$\ddot{q} = f(q, \dot{q}, \Gamma) \qquad (30)$$
$$\ddot{q} = -A(q)^{-1}[C(q, \dot{q})\dot{q} + Q(q) - \Gamma] \qquad (31)$$

With,
$q = [q_1\ q_2\ q_3\ q_4\ q_5\ q_6]^T$: The angular position vector (6x1).

$\dot{q} = [\dot{q}_1\ \dot{q}_2\ \dot{q}_3\ \dot{q}_4\ \dot{q}_5\ \dot{q}_6]^T$: The angular velocity vector (6x1).
$\ddot{q} = [\ddot{q}_1\ \ddot{q}_2\ \ddot{q}_3\ \ddot{q}_4\ \ddot{q}_5\ \ddot{q}_6]^T$: The angular acceleration vector (6x1).
$\Gamma = [\Gamma_1\ \Gamma_2\ \Gamma_3\ \Gamma_4\ \Gamma_5\ \Gamma_6]^T$: The input torques vector(6x1).

3.2. Application of the input-output feedback linearization approach to a robot manipulator with six degrees of freedom:

In this section, after we dermined the inverse dynamic model of the system [15-16], we used the equation of motion of the six-link of the rigid manipulator robot "EPSONC4" which represented by the relation (31) and we considered the state variables of the system defined in state space as;

$$x_1 = q_1, x_2 = \dot{q}_1, x_3 = q_2, x_4 = \dot{q}_2, x_5 = q_3, x_6 = \dot{q}_3,$$
$$x_7 = q_4, x_8 = \dot{q}_4, x_9 = q_5, x_{10} = \dot{q}_5, x_{11} = q_6, x_{12} = \dot{q}_6,$$

After derivation of the above state variables, we written the obtained system as;

$$
\begin{cases}
\dot{x}_1 = x_2 \\
\dot{x}_2 = \ddot{q}_1 = -A(x_1)^{-1}[C(x_1, x_2)x_2 + Q(x_1) - \Gamma_1] \\
\dot{x}_3 = x_4 \\
\dot{x}_4 = \ddot{q}_2 = -A(x_3)^{-1}[C(x_3, x_4)x_4 + Q(x_3) - \Gamma_2] \\
\dot{x}_5 = x_6 \\
\dot{x}_6 = \ddot{q}_3 = -A(x_5)^{-1}[C(x_5, x_6)x_6 + Q(x_5) - \Gamma_3] \\
\dot{x}_7 = x_8 \\
\dot{x}_8 = \ddot{q}_4 = -A(x_7)^{-1}[C(x_7, x_8)x_8 + Q(x_7) - \Gamma_4] \\
\dot{x}_9 = x_{10} \\
\dot{x}_{10} = \ddot{q}_5 = -A(x_9)^{-1}[C(x_9, x_{10})x_{10} + Q(x_9) - \Gamma_5] \\
\dot{x}_{11} = x_{12} \\
\dot{x}_{12} = \ddot{q}_6 = -A(x_{11})^{-1}[C(x_{11}, x_{12})x_{12} + Q(x_{11}) - \Gamma_6]
\end{cases}
\tag{32}
$$

Then, the affine form of nonlinear, multivariable and dynamic model of the robot manipulator is appeared which given by the following system (33):

$$
\begin{cases}
\dot{X}(t) = f(X(t)) + \sum_{i=1}^{p} g_i(X(t))U_i(t) \\
Y_i(t) = h_i(X(t)) \\
i = 1,2, \dots 6
\end{cases}
\tag{33}
$$

where,

$$
f(x) =
\begin{bmatrix}
x_2 \\
-A(x_1)^{-1}[C(x_1, x_2)x_2 + Q(x_1)] \\
x_4 \\
-A(x_3)^{-1}[C(x_3, x_4)x_4 + Q(x_3)] \\
x_6 \\
-A(x_5)^{-1}[C(x_5, x_6)x_6 + Q(x_5)] \\
x_8 \\
-A(x_7)^{-1}[C(x_7, x_8)x_8 + Q(x_7)] \\
x_{10} \\
-A(x_9)^{-1}[C(x_9, x_{10})x_{10} + Q(x_9)] \\
x_{12} \\
-A(x_{11})^{-1}[C(x_{11}, x_{12})x_{12} + Q(x_{11})]
\end{bmatrix}
;
$$

$$
g_1(x) =
\begin{bmatrix}
0 \\ A(x_1)^{-1} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0
\end{bmatrix}
; g_2(x) =
\begin{bmatrix}
0 \\ 0 \\ 0 \\ A(x_3)^{-1} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0
\end{bmatrix}
; g_3(x) =
\begin{bmatrix}
0 \\ 0 \\ 0 \\ 0 \\ A(x_5)^{-1} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0
\end{bmatrix}
$$

$$
g_4(x) =
\begin{bmatrix}
0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ A(x_7)^{-1} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0
\end{bmatrix}
; g_5(x) =
\begin{bmatrix}
0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ A(x_9)^{-1} \\ 0 \\ 0 \\ 0
\end{bmatrix}
; g_6(x) =
\begin{bmatrix}
0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ A(x_{11})^{-1}
\end{bmatrix}
$$

And,

$$X = [x_1, x_2, x_3, x_4, x_5, x_6 x_7, x_8, x_9, x_{10}, x_{11}, x_{12}]^T;$$
$$\dot{X} = [\dot{x}_1, \dot{x}_2, \dot{x}_3, \dot{x}_4, \dot{x}_5, \dot{x}_6 \dot{x}_7, \dot{x}_8, \dot{x}_9, \dot{x}_{10}, \dot{x}_{11}, \dot{x}_{12}]^T;$$
$$U = [u_1, u_2, u_3, u_4, u_5, u_6]^T = [\Gamma_1, \Gamma_2, \Gamma_3, \Gamma_4, \Gamma_5, \Gamma_6]^T;$$

Links positions
$$
\begin{cases}
y_1 = h_1(x) = x_1 = q_1; \\
y_2 = h_2(x) = x_3 = q_2; \\
y_3 = h_3(x) = x_5 = q_3; \\
y_4 = h_4(x) = x_7 = q_4; \\
y_5 = h_5(x) = x_9 = q_5; \\
y_6 = h_6(x) = x_{11} = q_6;
\end{cases}
\tag{34}
$$

So, we made the derivation of each output y_i intel the inputs appeared in the expression and we computed the relative degrees r_i for each joint of the robot as follows [17-18];

$$
\begin{cases}
y_1 = h_1(x) = x_1 \\
\dot{y}_1 = L_f h_1(x) = \dot{x}_1 = x_2 \\
y_1^{(2)} = L_f^2 h_1(x) + L_g L_f h_1(x)u \\
\quad r_1 = 2 \\
y_2 = h_2(x) = x_3 \\
\dot{y}_2 = L_f h_2(x) = \dot{x}_3 = x_4 \\
y_2^{(2)} = L_f^2 h_2(x) + L_g L_f h_2(x)u \\
\quad r_2 = 2 \\
\vdots \\
y_6 = h_6(x) = x_{11} \\
\dot{y}_6 = L_f h_6(x) \dot{x}_{11} = x_{12} \\
y_6^{(2)} = L_f^2 h_6(x) + L_g L_f h_6(x)u \\
\quad r_6 = 2
\end{cases}
\tag{35}
$$

Therefore, the relative degree of each joint r_i is well defined and is equal to 2, so we computed the nonlinear control law $u_i(t)$ of each joint of the system as this relation (36);

$$
u_i(x(t)) = -\frac{L_f^{r_i} h_i(x(t))}{L_g^{r_i-1} L_f h_i(x(t))} + \frac{v_i(t)}{L_g^{r_i-1} L_f h_i(x(t))}
\tag{36}
$$
$$i = 1 \dots 6$$

By using the nonlinear control law and diffeomorphic transformation given above, the nonlinear dynamic system with six degrees of freedom is converted into the following Brunovesky canonical form and simultaneously output decoupled [19].

$$\begin{cases} \dot{Z} = Az + BV \\ \quad Y = CZ \end{cases} \quad (37)$$

where,

$$Z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_{12} \end{bmatrix} = \begin{bmatrix} h_1 \\ L_f h_1 \\ h_2 \\ L_f h_2 \\ \cdot \\ \vdots \\ h_6 \\ L_f h_6 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{12} \end{bmatrix}; \dot{Z} = \begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \\ \vdots \\ \dot{z}_{12} \end{bmatrix} = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \vdots \\ \dot{x}_{12} \end{bmatrix}; V = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_6 \end{bmatrix}$$

$$A = \begin{bmatrix} 0 & 1 & & & & \\ 0 & 0 & \cdots & & & 0 \\ & \ddots & & \ddots & \cdots & \\ & & & 0 & 1 & \\ \vdots & & \cdots & 0 & 0 & \vdots \\ & & & & & \ddots \\ \vdots & & & & \ddots & \vdots \\ 0 & & \cdots & & & 0 & 1 \\ & & & & & 0 & 0 \end{bmatrix};$$

$$B = \begin{bmatrix} 0 & & \cdots & & 0 \\ 1 & & & & \\ & \ddots & \cdots & & \\ \vdots & \cdots & 0 & & \vdots \\ & & 1 & & \\ \vdots & & & \ddots & \vdots \\ \vdots & \cdots & \cdots & & 0 \\ 0 & & & & 0 & 1 \end{bmatrix}; C = \begin{bmatrix} 1 & 0 & & \cdots & & 0 \\ 0 & & & \ddots & \cdots & \\ & & & & & \\ \vdots & & \cdots & 1 & 0 & \vdots \\ & & & & \ddots & \\ \vdots & & \cdots & & \ddots & \vdots \\ 0 & & & & 0 & 1 & 0 \end{bmatrix}$$

The above matrices A,B and C are of dimension respectively : $(12 \times 12), (12 \times 6) \ and \ (6 \times 12)$.

We note that, the obtained linear system (37) definied by six decoupled and linear subsystems that is has the following form (38), and i=1...6;

$$\begin{cases} \dot{z}_i = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} z_i + \begin{bmatrix} 0 \\ 1 \end{bmatrix} v_i \\ \quad y_i = \begin{bmatrix} 1 & 0 \end{bmatrix} z_i \end{cases} \quad (38)$$

where,

$$z_i = \begin{bmatrix} z_{2i-1} \\ z_{2i} \end{bmatrix}$$

For the objective of linear control, we added to each subsystem a linear PD controller which has represented equation (39) as;

$$v_i(t) = K_p^i \big(1 + K_d^i s\big) e_i(t) \ ; i = 1 \dots 6 \quad (39)$$

$$e_i(t) = y_{d_i}(t) - y_i(t) \quad (40)$$

$$v_i = y_i^{r_i} \quad (41)$$

we obtained the relative degree for each subsystem as : $r_i = 2$

$$v_i = K_d^i \big(y_{d_i}{}^{(r_i-1)} - y_i{}^{(r_i-1)} \big) + K_p^i \big(y_{d_i} - y_i \big) \quad (42)$$

$$v_i(t) = \begin{bmatrix} K_p^i & K_d^i \end{bmatrix} \begin{bmatrix} y_{d_i}(t) - y_i(t) \\ y_{d_i}{}^{(r_i-1)}(t) - y_i{}^{(r_i-1)}(t) \end{bmatrix} \quad (43)$$

Then, we used the diffeomorphic transformation (16) presented above, we considered these relations for each SISO subsystem:

$$\begin{cases} z_1^i = h_i\big(x(t)\big) - y_{d_i} \\ z_2^i = L_f h_i\big(x(t)\big) - \dot{y}_{d_i} \end{cases} \quad (44)$$

That is leads to the non-linear coordinate transformation given by:

$$\begin{cases} \dot{z}_1^i = z_2^i = L_f h_i\big(x(t)\big) - \dot{y}_{d_i} \\ \dot{z}_2^i = L_f{}^2 h_i\big(x(t)\big) + L_g L_f h_i\big(x(t)\big) u_i\big(x(t)\big) - \ddot{y}_{d_i} \end{cases} \quad (45)$$

$$v_i(t) = -\sum_{j=0}^{r_i-1} K_j Z_i = -\sum_{j=0}^{r_i-1} K_j [L_f{}^{(j)} h_i\big(x(t)\big) - y_{d_i}{}^{(j)}(t)] \quad (46)$$

So, we expressed the relation (46) in equation (36),we obtained:

$$u_i\big(x(t)\big) = -\frac{L_f{}^{r_i} h_i\big(x(t)\big)}{L_g{}^{r_i-1} L_f h_i\big(x(t)\big)} + \frac{-\sum_{j=0}^{r_i-1} K_j [L_f{}^{(j)} h_i\big(x(t)\big) - y_{d_i}{}^{(j)}(t)]}{L_g{}^{r_i-1} L_f h_i\big(x(t)\big)} \quad (47)$$

$$i = 1 \dots 6$$

Next, we expressed the equation (47) in the subsystem (45), the non linear subsystem (45) was transformed to a linear subsystem which expressed as relation (48):

$$\begin{bmatrix} \dot{z}_1^i \\ \dot{z}_2^i \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ K_p^i & K_d^i \end{bmatrix} \begin{bmatrix} z_1^i \\ z_2^i \end{bmatrix} \quad (48)$$

$$i = 1 \dots 6$$

The obtained linear subsystem consists of a second order system with linear output, that is can be computed the poles of the above subsystem as;

$$s_{1,2} = -\xi \omega_n \mp \omega_n \sqrt{1 - \xi^2} \quad (49)$$

Where, ξ: means damping ratio, ω_n:means natural frequency, for goal of stability we chosed the parameters of each PD controller applied to each subsystem as [19]:

$$\begin{cases} K_p^j = \omega_n{}^2 \\ K_d^j = \xi \omega_n \end{cases} \quad j = 1 \dots 6 \quad (50)$$

4. Simulation Results

In order to improve the efficiency of the proposed approach, we applied the above method of linearization to a robot manipulator which represent a non linear, decoupled and multivariable system. We imposed the sinusoidal signals as $y_{id}(t), \dot{y}_{id}(t)$ and $y_{id}{}^{r_i}(t)$ which are represented inputs desired trajectories to each joint of the studied system. Therefore, the relative degrees r_i is well

defined, we presented the simulation results depicting the output $y_i^{r_i}(t)$ of each joint i=1...6 as shown in Figures 2,3,4,5,6,7,8-9.
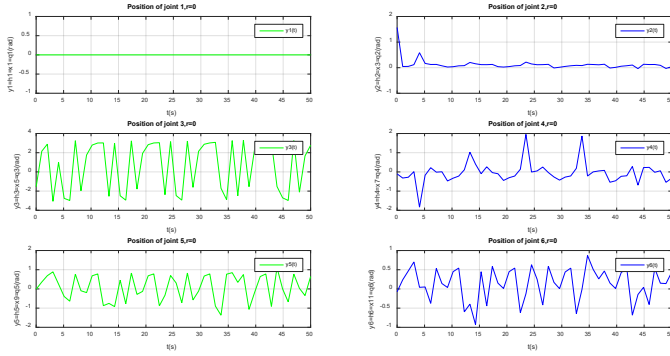
Figure 2. The outputs $y_i^{r_i}(t)$ plot of each joint, the relative degrees $r_i = 0$, i=1...6.

Firstly, we made the simulation of the outputs of each joint of the robot, $y_i(t) = h_i(x(t)) = x_i(t)$, the results are therefore given in Figure 2, we can see the non linearity of each subsystem. Second, we applied the Lie derivation to each output we shown the results presented in Figure 3 as.
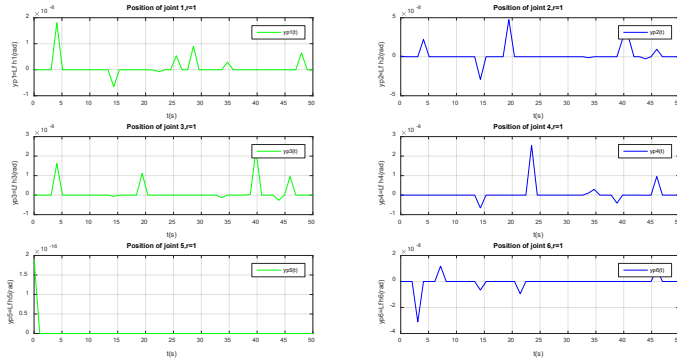
Figure 3. The outputs $y_i^{r_i}(t)$ plot of each joint, the relative degrees $r_i = 1$, i=1...6.

In this case, the relative degrees equal to $r_i = 1$, $L_g L_f h_i(x(t)) = 0$, so no one input has been appeared, as shown in Figure 3, the system still non linear. Finally, we made the derivation again of each recent output $\dot{y}_i$ the inputs appeared in the expression and we computed the final relative degrees $r_i = 2$ for each joint of the robot, $y_i^{(2)} = L_f^2 h_i(x) + L_g L_f h_i(x)u$, $L_g L_f h_i(x) \neq 0$, the results was illustrated in these Figures 4,5,6,7,8-9.
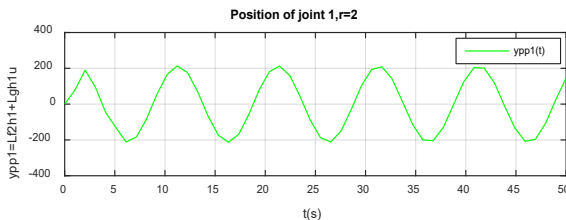
Figure 4. The output $y_i^{r_i}(t)$ plot of the joint 1, the relative degree $r_1 = 2$.
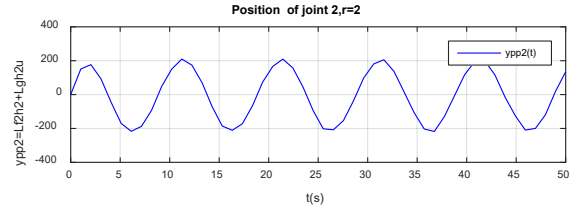
Figure 5. The output $y_i^{r_i}(t)$ plot of the joint 2, the relative degree $r_2 = 2$.
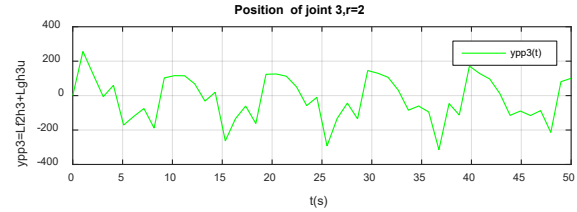
Figure 6. The output $y_i^{r_i}(t)$ plot of the joint 3, the relative degree $r_3 = 2$.
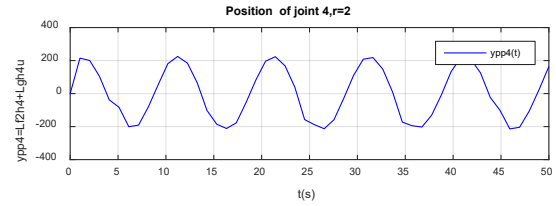
Figure 7. The output $y_i^{r_i}(t)$ plot of the joint 4, the relative degree $r_4 = 2$.
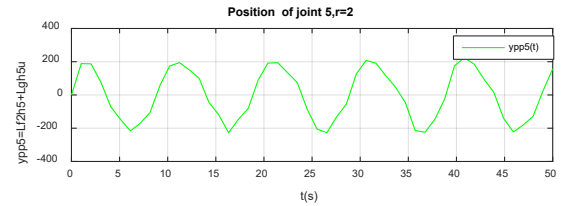
Figure 8. The output $y_i^{r_i}(t)$ plot of the joint 5, the relative degree $r_5 = 2$.
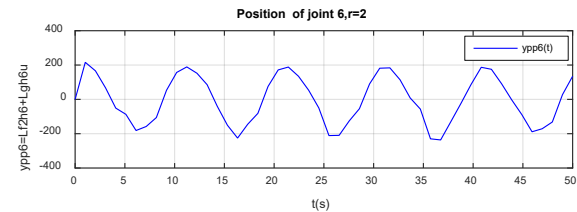
Figure 9. The output $y_i^{r_i}(t)$ plot of the joint 6, the relative degree $r_6 = 2$.

These Figures 4,5,6,7,8-9 shown that the output of each SISO subsystem asymptotically tracks the sinusoidal input trajectory with minimum of oscillation and the efficiency of the studied approach.

5. Conclusion

In this paper, the input-output feedback linearization approach is presented which is a way of algebraically transforming nonlinear, multivariables, complex and dynamics systems into linear ones, so we can be applied linear control, this technique has attracted lots of research in recent years. For that we applied this proposed approach to derive the control of a robot manipulator with six degrees of freedom which followed two major steps. Firstly, by using the above approach and

diffeomorphic transformation, we converted the non linear and decoupled dynamics model of the robot to linear system. Then, we designed a linear PD control law for each decoupled subsystem to control the angular position of each joint of this robot for tracking purposes. Finally, the obtained results in different simulations illustrated the accuracy of the proposed approach.

References

[1] MW. Spong, S.Hutchinson, and M. Vidyasagar. *Robot Dynamics and Control ,* Second Edition. January 28, 2004.

[2] J.-J. Slotine & W. Li, Applied Nonlinear Control, Prentice-Hall, Englewood Cliffs, NJ, 1991.

[3] A. Isidori. Nonlinear Control Systems. Springer Verlag, 1989.

[4] H. Khalil. Nonlinear Systems. Prentice Hall, Upper Saddle River, Second Edition, 1996.

[5] J.Hauser,S.Sastry, and P.Kokolovic,"Non linear control via approximate Input_Output Linearization,The Ball and Beam Exemple". IEEE Transactions onAutomatic Control ,Vol 37,No.3 March 1992.

[6] E.D. Sontag,M. Thoma,A. Isidori and J.H. van Schuppen. Nonlinear and Adaptive Control with Applications, 2008 Springer-Verlag London Limited.

[7] H. Nijmeijer and A. J. van der Schaft, Nonlinear Dynamical Control Systems, Springer-Verlag, New York, 1990.

[8] M. Vidyasagar. Nonlinear Systems Analysis. Prentice Hall,Second Edition, 1993.

[9] K.M. Hangos, J. Bokor et G. Szederkényi, Analysis and Control of Nonlinear Process Systems, Springer-Verlag London Limited, 2004.

[10] J.HAUSER, SH. SASTRY and G. MEYER. "Nonlinear Control Design for Slightly Non minimum Phase Systems: Application to V/STOL Aircraft". International Federation of Automatic Control. Vol. 28, No. 4, pp. 665-679, 1992.

[11] W.Ghozlane and J.Knani."Modelling and Simulation Using Mathematical and CAD Model of a Robot with Six Degrees of Freedom,"CEIT-2016, 4thInternational Conference on Control Engineering & Information Technology on IEEE.16-18 December 2016-Hammamet, Tunisia.

[12] B.Siciliano and O. Khatib, Handbook of Robotics: Springer, 2007.

[13] W. Khalil and E. Dombre, Modeling Identification and Control of Robots, Hermes Penton London, 2002.

[14] Thomas R. Kurfess, Robotics and Automation Handbook, CRC press, 2005.

[15] Y . L . Chen: " Nonlinear Feedback and Computer Control of Robot Arms". D.Sc. Dissertation, Washington University, St. Louis 1984.

[16] V.D. Yurkevich, Design of Nonlinear Control Systems with the Highest Derivative in Feedback,World Scientific Publishing, 2004.

[17] Frank L.Lewis, Robot dynamics and control, in robot Handbook: CRC press, 1999.

[18] L. Sciavicco and B. Siciliano, Modeling and control of robot manipulators, 2nd edition, Springer-Verlag London Limited, 2000.

[19] C.T.Leondes. Control and Dynamic Systems,Advances In Theory and Applications, part 1 of 2,USA,1991.

[20] Zakaria, M. Z., Jamaluddin, H., Ahmad, R. & Loghmanian, S. M. R. (2010). "Multiobjective Evolutionary Algorithm Approach in Modeling Discrete-Time Multivariable Dynamics Systems". Computational Intelligence, Modelling and Simulation (CIMSiM), 2010 Second International Conference on, Bali, Indonesia. 28-30 Sept. 2010. 65-70.

[21] Zakaria, M. Z., Jamaluddin, H., Ahmad, R. & Loghmanian, S. M. R. (2012). "Comparison between Multi-Objective and Single-Objective Optimization for the Modeling of Dynamic Systems". Journal of Systems and Control Engineering, Part I. Proc. Instn. Mech. Engrs., 226 (7): 994-1005.

Virtual Watershed System: A Web-Service-Based Software Package For Environmental Modeling

Rui Wu[1], Connor Scully-Allison[*,2], Moinul Hossain Rifat[2], Jose Painumkal[2], Sergiu Dascalu[2], Jr. Frederick C Harris[2]

[1]*Department of Computer Science, East Carolina University, 27858, USA*

[2]*Computer Science & Engineering Department, University of Nevada, Reno, 89557, USA*

A B S T R A C T

The physically-based environmental model is a crucial tool used in many scientific inquiries. With physical modeling, different models are used to simulate real world phenomena and most environmental scientists use their own devices to execute the models. A complex simulation can be time-consuming with limited computing power. Also, sharing a scientific model with other researchers can difficult, which means the same model is rebuilt multiple times for similar problems. A web-service-based framework to expose models as services is proposed in this paper to address these problems. The main functions of the framework include model executions in cloud environments, NetCDF file format transmission, model resource management with various web services. As proof of concept, a prototype is introduced,implemented and compared against existing similar software packages. Through a feature comparison with equivalent software, we demonstrate that the Virtual Watershed System (VWS) provides superior customization through its APIs. We also indicate that the VWS uniquely provides friendly, usable UIs enabling researchers to execute models.

1 Introduction

Modeling has become an indispensable tool in growing environmental scientists' understanding of how natural systems react to changing conditions. It sheds light on complex environmental mysteries and helps researchers in formulating policies and decisions on future scenarios. Environmental modeling is highly challenging as it involves complex mathematical computations, rigorous data processing, and convoluted correlations between numerous parameters. Three commonly-used and important scientific software quality measures are maintainability, quality, and scalability. Issues like data storage, coupling models, retrieval, and running are hard problems and need to be addressed with extra efforts from software engineering perspective. It is a challenging job to design integrated systems that can address all these issues.

It is essential to build high-quality software tools and design efficient frameworks for scientific research. Abundant scientific model data are generated and collected in recent years. Software engineering can assist this emergence through the creation of distributed software systems and frameworks enabling scientific collaboration with previously disparate data and models. It is challenging to implement software tools for interdisciplinary research because of the problem of communication and team building among different scientific communities. For example, the same terminology can have different meanings in different domains. This increases difficulties in comprehending software requirements when a project involves stakeholders (e.g. researchers) from different fields.

Most work described in this paper is for Watershed Analysis, Visualization, and Exploration (WC-WAVE), which is a NSF EPSCoR-supported project and initiated by jurisdictions of EPSCoR of Nevada, Idaho and New Mexico. The WC-WAVE project includes three principal components: watershed science, data cyberinfrastructure and data visualization [1]. The project main goal is to implement VWS with the collaborations

---

[*]Connor Scully-Allison, 1664 N. Virginia Street, CSE (0171), Reno, NV 89557, (775) 771-1469 & cscully-allison@nevada.unr.edu

between cyberinfrastructure team members and hydrologists. The platform is able to store, share, model and visualize data on-demand through an integrated system. These features are crucial for hydrologic research.

Different hydrologic models, such as ISNOBAL and PRMS, are commonly used by WC-WAVE hydrologists. These models are leveraged to predict or examine hydrologic processes of Lehman Creek in Nevada, Dry Creek and Reynolds Creek in Idaho, and Jemez Creek in New Mexico. We propose a framework for representing model data in a standardized format called the "Network Common Data Format" (NetCDF) and exposing these hydrologic models through web services. This framework is based on our previous work introduced in [2]. To improve on our prior work we have implemented some new Docker APIs to control system components wrapped in docker containers. To simplify data extraction, modification, and storage, NetCDF data format [3] is used in the system for grid-organized scientific data. Most parts of the system framework can be reused for data-intensive purpose because it is designed with the blueprint and template concepts. ISNOBAL and PRMS are physically-based models which can produce very accurate results. However, these two models require abundant computing power. To solve the challenge, we leverage a cluster to execute models in parallel. ISNOBAL and PRMS are used in this paper to demonstrate the ideas and functionality of the proposed framework. Throughout the remainder of this paper, we refer to this prototype system as the VWS.

ISNOBAL is a grid-based DEM (Digital Elevation Model) and created to model the seasonal snow cover melting and development. The model author is Marks et al. [4] and it is initially developed for Utah, California, and Idaho mountain basins. The model determines runoff and snowmelt based on terrain, precipitation, region characteristics, climate, and snow properties [4].

PRMS is short for Precipitation-Runoff Modeling System and is initially written with FORTRAN in 1983. PRMS is prevalent physical process based distributed-parameter hydrologic model and the main function of a PRMS model is to evaluate a watershed response to different climate and land usage cases [5, 6, 7]. It composed of algorithms describing various physical processes as subroutines. The model, now in its fourth version, has become more mature over the years of development. Different hydrology applications, such as measurement of groundwater and surface water interaction, the interaction of climate and atmosphere with surface water, water and natural resource management, have been done with the PRMS model [5, 6, 7].

This paper is organized as follows in its remaining sections: Section 2 introduces background and related work; Section 3 describes the system design; Section 4 describes the prototype system and how the software was built using RESTful APIs; Section 5 compares our work with related tools; and Section 6 contains the papers conclusions and outlines planned future work.

2 Background and Related Work

"How to implement software for interdisciplinary research?" is an interesting question and there exists some successful work on environments and frameworks which seek to answet this question. In this section, relevant, popular earth science applications and frameworks are introduced.

Community Surface Dynamics Modeling System (CSDMS) was a project started in 1999 to conduct expeditious research of earth surface modelers by creating a community driven software platform. CSDMS applies a component-based software engineering approach in the integration of plug-and-play components, as the development of complex scientific modeling system requires the coupling of multiple, independently developed models [8]. CSDMS allows users to write their components in any popular language. Also they can use components created by others in the community for their simulations. CSDMS treats components as pre-compiled units which can be replaced, added to, or deleted from an application at runtime via dynamic linking. Many key requirements drove the design of CSDMS, including the support for multiple operating systems, language interoperability across both procedural and object-oriented programming languages, platform independent graphical user interfaces, use of established software standards, interoperability with other coupling frameworks and use of HPC tools to integrate parallel tools and models into the ecosystem.

A leading hydrologic research organization is CUAHSI. "CUAHSI" , and acronym for "Consortium of Universities for the Advancement of Hydrologic Science Inc.," represents universities and international water science-related organizations. One of most highly esteemed products is HydroShare, which is a hydrologic data and model sharing web application. Hydrologists can easily access different model datasets and share their own data. Besides this, this platform offers many distributed data analysis tools. A model instance can be deployed in a grid, cloud or high-performance computing cluster with HydroShare. Also, a hydrologist is able to publish outcomes of their research, such as a dataset or a model. In this way, scientists use the system as a collaboration platform for sharing information. HydroShare exposes its functionality with Application Programming Interfaces (APIs), which means its web application interface layer and service layer are separated. This enables interoperability with other systems and direct client access [9].

Model as a Service is proposed by Li et al. [10] for Geoscience Modeling. It is a cloud-based solution and [10] has implemented a prototype system to execute high CPU and memory usage models remotely as a service with third party platform, such as AWS (Amazon Web Service) and Microsoft Azure.

The key idea of MaaS is that model executions can be done through a web interface with user inputs. Computer resources are provisioned with a cloud provider, such as Microsoft Azure. The model registration is done in the framework with a virtual machine image

repository. If a model is registered and placed in the repository, it can be shared by other users and multiple model instances can be executed in parallel based on demand.

McGuire and Roberge designed a social network to promote collaboration between watershed scientists. Despite being highly available, the collaboration between the general public, scientists, and citizen has not been leveraged. Also, hydrologic data is not integrated in any system. The main goal of this work is to design a collaborative social network for multiple watershed scientific and hydrologic user groups [11]. However, more efforts need to be done.

The Demeter Framework by Fritzinger et al. [12] represents another attempt to utilize software frameworks as a scientific aid in the area of climate change research. A software framework named "The Demeter Framework" is introduced in the paper and one of the key ideas is a component-based approach to integrate different components into the system for the "model coupling problem." "The model coupling problem" refers to using a model's outputs as another model's inputs to solve a problem.

Walker and Chapra proposed a web-based client-server approach for solving the problem of environmental modeling compared to the traditional desktop-based approach. The authors assert that, with the improvement in modern day web browsers, client-side approaches offer improved user interfaces compared to traditional desktop software. In addition, powerful servers enable users to perform simulations and visualizations within the browser [13].

The University of New Mexico has implemented a data engine named GSToRE (Geographic Storage, Transformation and Retrieval Engine). The engine is designed for earth scientific research and the main functions of the engine are data delivery, documentation, and discovery. It follows the combination of community and open standards and implemented based on service oriented architecture. [14].

2.1 Service Oriented Architecture

Industry has shown more and more interests on Service Oriented Architecture (SOA) to implement software systems. [15]. The main idea of SOA is to have business logic decomposed into different units (or services). These units are self-contained and can be easily deployed with container techniques, such as Docker.

Representational State Transfer Protocol (REST) is primarily an architectural style for distributed hypermedia systems introduced by Fielding, Roy Thomas in his Ph.D. dissertation [16]. REST defines a way for a client-server architecture on how a client and a server should interact, by using a set of principles. REST has been adopted for building the main architecture of the proposed system. Statelessness, uniform Interface, and cache are the main characteristics of a REST client-server architecture [16]. It is for these characteristics that REST is leveraged in our proposed system.

Statelessness Statelessness is the most important property or constraint for a client-server architecture to be RESTful. The communication between the client and the server must be stateless, which means the server is not responsible for keeping the state of the communication. It is the client's responsibility. A request from the client must contain all the necessary information for the server to understand the request [16, 17]. Two subsequent requests to the server will not have any interdependence between each other. Introducing this property on the client-server architecture presents several benefits regarding visibility, reliability, and scalability [16, 17]. For example, as the server is not responsible for keeping the state and two subsequent requests are not interrelated, multiple servers can be distributed across a load-balanced system where different servers can be responsible for responding to different requests by a client.

Uniform Interface Another important property of a RESTful architecture is it provides a uniform interface for the client to interact with a server. Instead of an application's particular implementation, it forces the system to follow a standardized form. For example HTTP 1.1 which is a RESTful protocol provides a set of verbs (e.g., GET, POST, PUT, DELETE, etc.) for the client to communicate with the server. The verbs, such as "GET" and "POST", work as an interface making the client-server communication generic[16].

Cache REST architecture introduces cache constraint to improve network efficiency [16]. A server can allow a client to reuse data by enabling explicitly for labeling some data cacheable or non-cacheable. A server can serve data that will not change in the future as cached content, allowing the client to eliminate partial interaction with that data in a series of requests.

2.2 REST Components

The main components of REST architecture include resources, representations, and resource identifiers.

Resources The resource is the main abstract representation of data in REST architecture [16]. Any piece of data in a server can be represented as a resource to a client. A document, an image, data on today's weather, a social profile, *everything* is considered a resource in the server. Formally, a resource is a temporarily varying function of $M_R(t)$ that maps to a set of entities for time t [16].

Representations A resource is the abstract building block of the data in a web server. For a client to consume the resource it needs to be presented in a way the client can understand. This is called representation. A representation is a presentation format for representing the current state of a resource to a consumer. Some commonly used resource representation format in the current standard are HTML (Hypertext Markup

Language), XML (Extensible Markup Language), JSON (JavaScript Object Notation), etc. A server can expose data content in different representations so that consumer can access the resources through resource identifiers (discussed in Section 2.2) in the desired format.

Resource Identifiers A resource is uniquely identified through a resource identifier in a RESTful architecture. For example, in HTTP Uniform Resource Identifier (URI) is used to identify a resource in a server. A URI can be thought as the address of a resource in the server [18]. A resource identifier is the key for a client to access and manipulate a resource in the server.

2.3 Microservice Architecture

Software as a Service (SaaS), as a new software delivery architecture, has emerged to leverage the widely-used REST standard for processing in addition to data transfer. The main advantage of SaaS is that it does not require local installation and this is a significant IT trend based on industry analysis [19].

Similarly, "microservice" decomposes an application into small components (or services) and these components communicate with each other through APIs. "Microservice" is a solution to monolithic architecture relevant problems. [20]. Because of the "microservice" characteristics, an application can be easily scaled and the deployment risks have been reduced without interrupting other services.

Traditional monolithic applications are built as a single unit using a single language stack and often composed of three parts, a front end client, a back-end database and an application server sitting in the middle that contains the business logic. Here, the application server is a monolith that serves as a single executable. A monolithic application can be scaled horizontally by replicating the application server behind a load balancer to serve the clients at scale. The biggest issue with monolithic architecture is that, as the application grows, the deployment cycle becomes longer as a small change in the codebase requires the whole monolith to be rebuilt and deployed [20]. It leads to higher risk for maintenance as the application grows. These pitfalls have lead to the idea of decomposing the business capabilities of an application into self-contained services.

Being a relatively new idea, researchers have attempted to formalize a definition and characteristics of *Microservices*. [20] [20] has put together a few essential characteristics of a microservice architecture that are described in brief in the following sections.

Componentization via Services The most important characteristic of a microservice architecture is that service functionalities need to be componentized. Instead of thinking of components as libraries that use in-memory function calls for inter-component communication, we can think of components in terms out-of-process services that communicate over the network,

quite often through a web service or a remote procedure call. A service has to be atomic, doing one thing and doing one thing well [21].
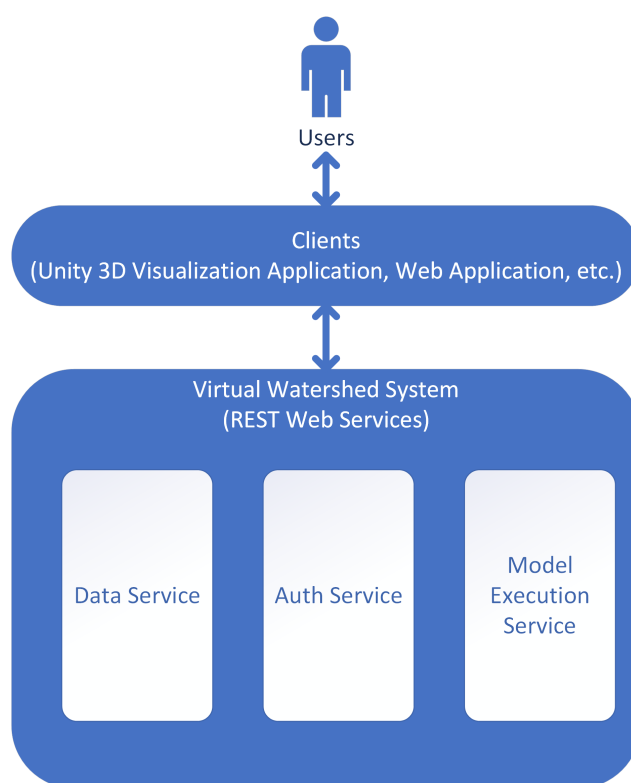
Figure 1: High-level diagram of VWS: clients and VWS communicate through REST Web Services. Possible clients include a script and applications. VWP provide data service, auth service, and model execution service.

Organized Around Business Capabilities Monolithic applications are typically organized around technology layers. For example, a typical multi-tiered application might be split logically into persistence layer, application layer and UI layer and teams are also organized around the technologies. This logical separation creates the need for inter-team communication even for a simple change. In microservices, the product is organized around business capabilities where a service is concentrated on one single business need and owned by a small team of cross-domain members.

Decentralized Governance In a monolithic application, the product is governed in a centralized manner, meaning it restricts the product to a specific platform or language stack. But in microservices, as each service is responsible implementing an independent business capability, it allows for building different services with different technologies. As a result, the team gets to choose the tools that are best suited for each of the services.

[21], in his book Building Microservices [21], has discussed several concrete benefits of using microservices over a monolith. Several important benefits are discussed below in brief:
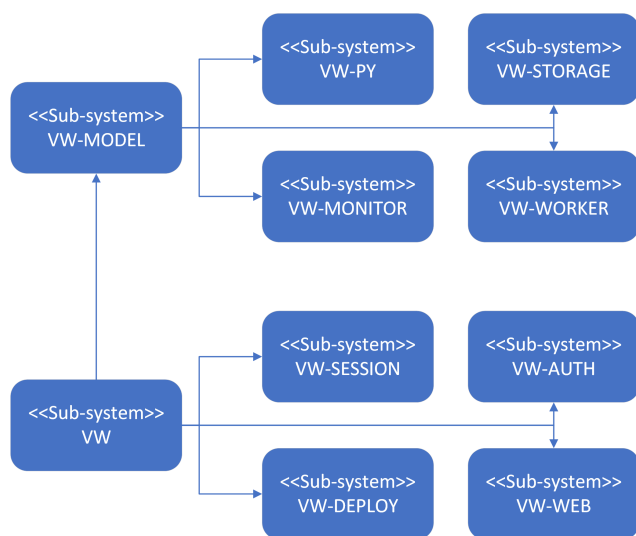
Figure 2: System-level diagram of VWS: two main system components are VW-MODEL and VW. VW-MODEL contains components relative to model execution and VW contains components relative to the platform.

Technology Heterogeneity When independent services are built separately, it allows adoption of different technology stacks for different services. This gives a team multiple advantages: liberty to choose a technology that best suits the need and adopt technology quickly for the needs.

Scaling Monolithic applications are hard to deal with when it comes to scaling. One big problem with monolithic applications is that everything needs to be scaled together as a piece. But microservice allows having control over the scaling application by allowing to scale the services independently.

Ease of Deployment Applying changes for a monolithic application requires the whole application to be re-deployed even for a minor change in the codebase. It poses a high risk as unsuccessful re-deployments even with minor changes can take down the software. Microservices, on the other hand, allow low-risk deployments without interrupting the rest of the services. They also allow having faster development process with small and incremental re-deployments.

3 Proposed Method

Here we introduce detailed design documentation for the VMS. We present the design using many common diagrams used in software engineering, including system diagrams and workflow diagrams. The VWS aims to create a software ecosystem named the *Virtual Watershed* by integrating cyberinfrastructure and visualization tools to advance watershed science research. Fig. 1 shows a general high-level diagram of components of the ecosystem. The envisioned system is centered

around services comprised of data, modeling, and visualization components. A high-level description of each depicted component is provided in Fig. 1.

3.1 Data Service and Modeling Service

To create a scalable and maintainable ecosystem of services, a robust data backend is crucial. The envisioned data service exposes a RESTful web service to allow easy storage and management of watershed modeling data. It allows retrieval of data in various OGC (Open Geospatial Consortium) standards like WCS, WMS to allow OGC compliant clients to retrieve data automatically. This feature is very important for data-intensive hydrologic research and is essential for a scientific modeling tool.

Hydrologists often use different modeling tools to simulate and investigate the change of different hydrologic variables around watersheds. The modeling tools are often complex to setup in local environments and takes up a good amount of time setting up [10]. Besides these modeling tools may require high computational and storage resources that make them hard to run in local environments. The proposed modeling service aims to solve these issues by allowing users to submit model execution tasks through simple RESTful web service API. This approach of allowing model-runs through a generic API solves multiple problems:

- It allows the users to run models on demand without having to worry about setting up environments.

- It allows modelers to accelerate the process of running models with different input parameters by submitting multiple model-runs to be run in parallel which might not be feasible in local environment due to lack of computational and storage resources.

- It opens the door for other services and clients to take advantage of the API to automate the process of running models in their workflow.

The primary goal of this Virtual Watershed system is to allow watershed modelers to share their data resources and execute relevant models through web services without having to install the models locally. The system is designed as a collection of RESTful microservices that communicate internally. The web service sits on top of an extensible backend that allows easy integration of models and scalability over model runs and number of users connected to the service.

The system provides a mechanism for integrating new models by conforming with a simple event driven architecture that allows registering a model in the system by wrapping it with a schema driven adaptor.

As model execution is a CPU intensive process, scaling the server with growing number of parallel model execution is an important issue. The proposed architecture aims to solve this problem by introducing a simple database-oriented job queue that provides options for adding more machines as the system grows.

3.2 System Level Design

Several different submodules comprise The Virtual Watershed system. Each module provides different functionalities to the entire framework. This relationship between modules is described with Figure 2. The following paragraphs will explain each submodule in detail.

VW-PY: Using the VW-PY module, users can define adapters to configure compatibility between different models and the VWS. An adaptor is python code which encapsulates a model and that allows for it to be run programmatically. VW-PY handles many aspects of model "wrapping" with this python code through an interface. Through this API, an user can define the code which handles converting between data formats, executing models, and model-progress-based event triggering. This event triggering system provides model-wrapper developers with the tools to emit model execution progress.

VW-MODEL: The VW-MODEL submodule, through a web service, exposes a RESTful API to the user/client. Users can upload, query, and retrieve model run packages using this API.

VW-WORKER: The VW-WORKER is a service which encapsulates model adaptors in a worker service that is organized in a queue data structure. This component communicates with the VW-MODEL component using a redis data-backend.

VW-STORAGE: The VW-STORAGE module provides an interface for object storage for the VW-MODEL component. Developed as a generic wrapper, sysadmins can configure this interface to work with local or cloud-hosted storage providers.

VW-AUTH: The VW-AUTH module provides authentication services for users/clients connecting to the VWS framework. This service provides clients with a JWT token enabling them to securely utilize the other services described in this section.

VW-SESSION: The VW-SESSION sub module coordinates different components in the VWS to provide a common session backend for a single user. Data about each session is maintained and managed with a Redis data store shared between services.

VW-WEB: This is the web-application front-end which provides users with APIs to interact with the model processing modules described above. In a standard use case of this module, users will get access to a session after logging into this system with the VW-AUTH component. Once given a session, users can use this interface to access resources, run models, track progress and upload/download model run resources.

3.3 Detailed Design

The VWS is comprised of many distinct services and web applications which communicate with one-another to facilitate model runs. We are able to provide secure and centralized communications with the aid of a common authentication gateway. VW-AUTH was developed as a micro service to provide this functionality.

It aggregates functionality for registration, authentication, and authorization for system users.
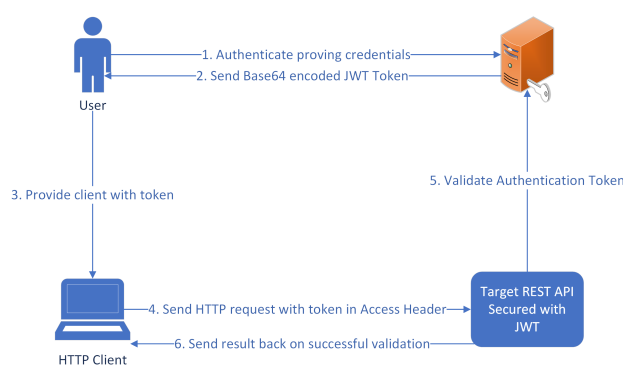
Figure 3: Workflow for accessing secure REST endpoint with JWT token

The Authentication component itself exposes RESTful endpoints that provide user access to different services, in addition to endpoints for authentication and registration. To enable the verification of a client's authentication by other services, a JSON Web Token (JWT) based authorization scheme is utilized. As an RFC standard for exchanging information securely between a client and a server, the using JWT ensures a high level of security in this system. A standard workflow for user authorization is depicted in Figure 3.

Users can manage uploaded models via a RESTful API endpoint provided by the Model Web Service. Using this endpoint, a properly authenticated user is able to request a model run, upload necessary input files needed by the model and start model execution. Model run data is stored by the VWS in a dedicated database located on a server operating out of the University of New Mexico [1].

Available models in the system are self-describing with schema that indicates necessary input files, formats, and execution policies. The schema also shows the mapping between user facing and model adaptor parameters. A user/client can use this schema to understand the necessary resources required to run a given model. The a typical model run from the perspective of the clinent side application involves six steps: 1) retrieve the server-side model schema; 2) instantiate a model run session on the server; 3) upload model inputs; 4) run the model; 5) track model run progress; and finally, 6) download outputs.

Figure 4 shows a simple workflow chart from a user or client's perspective. This workflow offers many advantages this workflow and sever-client architecture compared to traditional approaches to model running:

- Users do not need to know internal specifics of the model. Setup details and dependencies are hidden from the user, which provides a more seamless experience.

- Users don't need to worry about installing model dependencies.

- Users can easily initiate multiple parallel model runs on a server, expediting research.

- Users have ubiquitous access to server-side data via provided RESTful APIs.
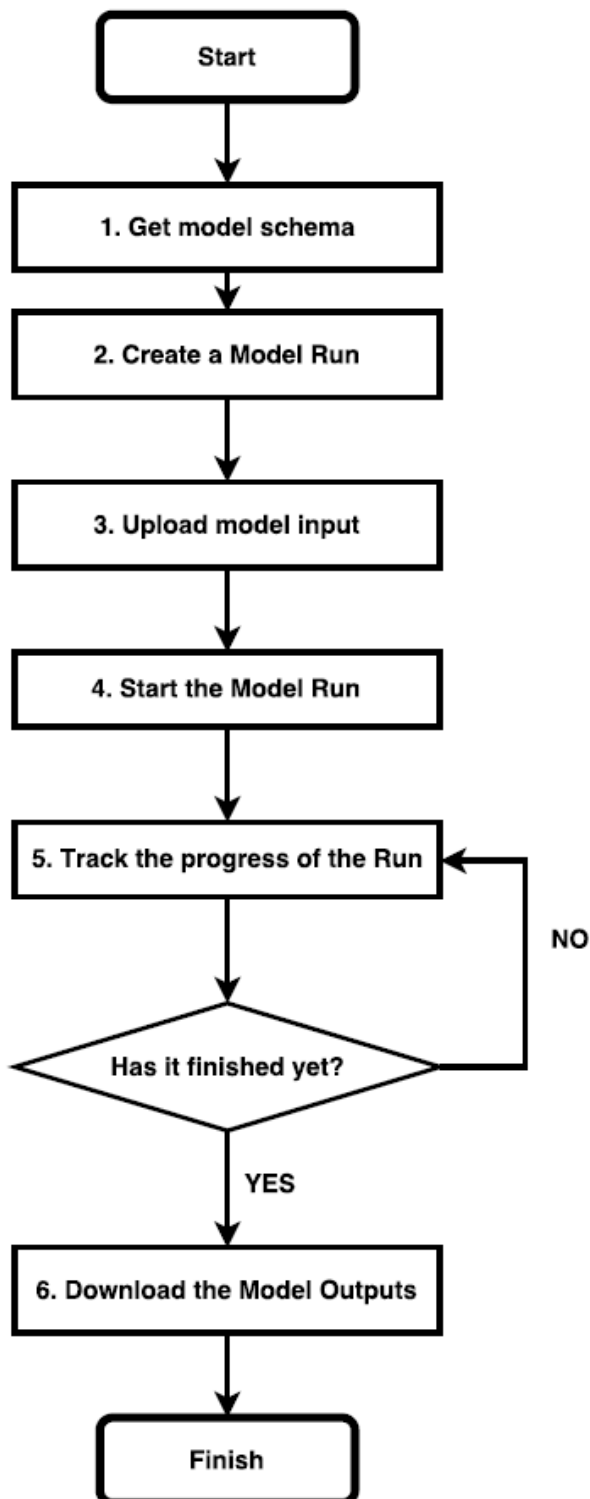
Figure 4: Workflow from user's perspective to run a model

The first step in creating the architecture for exposing "models as services" is enabling the programmatic running of models. A hurdle to accomplishing this is that a model can have many different dependencies required to run. Programmatic running is accomplished in this program by providing developers the tools to create python wrappers around models. These wrappers expose dependencies to model through container images.

In creating this system we had to strongly consider problems of data format heterogeneity with model inputs and outputs. Different models commonly accept and output data in different file formats. To achieve facilitate greater data interoperability between these models, we provided an option for developers to write NetCDF adapters for each models. A model adaptor is a Python program that handles data format conversion, model execution and progress notification. Adaptor developers must provide converters which define the method of conversion and deconversion between the native formats used by the model and netCDF. An execution function must also be implemented for the model. In this method, resource conversion and model execution occurs. The wrapper reports on model execution events via a provided "event emitter." An event listener can catch these events as they are reported by the emitter. This listener provides a bridge between the internal progress of the model and the user with the aid of a REST endpoint.

Through an adaptor, models can be encapsulated for programmatic execution. However, to bridge frontends with actual model execution a process is required from the model worker module. Utilizing a messaging queue, we create a bridge between the client frontend and worker backend. When a run task is submitted through from the client side, it is placed into the queue and assigned a unique id. The consumer/worker process listens to the queue through a common protocol for new jobs.

The worker service is a server-side python module that has access to a model's executable code and installed dependencies. The worker resides in an isolated server instance that contains the dependencies and libraries of the model installed. This module uses Linux containerization to facilitate easy deployment of model workers.

3.4 Deployment Workflow

A Linux-container-based deployment workflow has been devised for the for many different components of VWS. We utilized the containerization software, Docker, to enable our implementation of this workflow [22]. Docker containers have advantages over traditional virtual machines because they use fewer resources. This workflow allows for iterative deployment, simple scaling of the containerized components, and provides a strategy to register new models in the system.

Every VWS component is containerized with Docker. We have a set up a central repository of docker images which contain images for each component in this system. A docker "image" describes a template that provides Docker with the OS, dependencies of an

application and the application itself. Docker uses this template to build a working container. Each repository in the Virtual Watershed has a Dockerfile that describes how the component should be built and deployed into a query-able container. The repositories use webhooks to automate the building of docker containers.

Figure 5: Registration page of Virtual Watershed System

It is easy to register new models in this system using this containerization workflow. Using our system, a user may register a model with the creation of docker image describing the model. With this image users can declare the OS, libraries and other dependencies for a model. A typical registration of a model requires the following steps: 1) create a repository for the model; 2) develop the wrapper; 3) specify dependencies within the dockerfile; and finally, 5) create a docker image in the image repository.

4 System Prototype and Testing

The project re-used some code and similar structures to those introduced in [23, 24]. The web service frontends were built using a Python micro-service framework called Flask, with various extensions [25]. Some key libraries used were: 1) Flask-Restless, used to implement the RESTful API endpoints; 2) SQLAlchemy, to map the python data objects with the database schema [26]; 3) PostgreSQL as the database [27]; and,

4) Flask-Security with Flask-JWT which provides the utilities used for security and authentication. Celery was used to implement the task queue for model workers [28]. Redis worked as a repository for model results output by Celery. A REST specification library called Swagger was used to create the specification for the REST APIs. For the web front end, HTML5, CSS, Bootstrap, and Javascript (with ReactsJS) were used.

Figure 6: Dynamically generated upload form

To efficiently develop the VWS, an MVC design pattern was used to structure the code. The primary repository used to manage and distribute code was Github. Source code for this project is made publicly available through the Virtual Watersheds GitHub repository [28]. Dockerhub [29] is used for image management, which can automatically update images when Github code is changed.

The prototype system is comprised of two main components 1) the authentication module and 2) the modeling module. First, activities related to the VWS are handled by the authentication module. All standard user management functionality is handled by this module: registration, logins, verification, password management, and authentication token generation. Figure 5 shows the registration page of Virtual Watershed system. In addition to this interface, the VWS also provides endpoints for registration and authentication from user constructed scripts.

Modeling is necessary and commonly used in hydrologic research. Modification of the existing model simulations is a complex activity that hydrologists often have to deal with while analyzing complicated environmental scenarios. Modelers must make frequent modifications to underlying input files followed by lengthy re-runs. Programming languages could help significantly with this easily automated and repetitive task. However, it is complex and time consuming for hydrologists to write their own programs to handle file modifications for scenario based studies. To solve this challenge, the modeling module provides an in-

tuitive user interface where users can create, upload, run and delete models. The UI informs users about the progress of the model run with a bar that displays a percentage of completion. The module also includes a dashboard where users can view the models being run, the finished model runs, and also download the model run files. A user can execute multiple PRMS models in parallel by uploading the three input resources needed for PRMS model. The upload interface as shown in Figure 6 is generated dynamically from the model schema defined for each of the models.

Figure 7: Scenario creation interface for PRMS model that uses Modeling REST API to execute model

The modeling interface displays peak utility when users use it to programmatically run a bundle of models in parallel without having to worry about resource management. In the prototype system, the client is used to adjust input parameters for models and for model execution.

Model calibration can be a time consuming process

for many hydrologists. It often requires that the modeler re-run and re-run a model many time with slightly varied input parameters. For many modelers this is a manual process. They will edit input files with a text editor and execute the model from a command line on their local machine.

Understanding that it was crucial to address this issue, the virtual watershed protoype system was developed with a web-based scenario design tool. This tool provides a handy interface which enables for the rapid adjustment and calibration of PRMS model input variables. It also provides the tools to execute the tweaked models via the modeling web service, and visually compare outputs from differently calibrated runs. Figure 7 shows the interface developed which provides visually oriented tools for data manipulation. This interface abstracts away technical model data representations and allows users to model in intuitive ways. Figure 8 shows the output visualization after a the modeling web service has finished executing the model run.
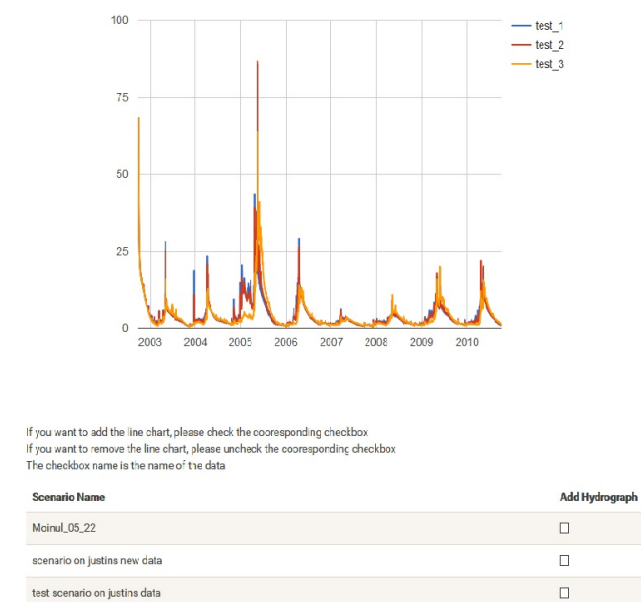
Figure 8: Comparing the results of the scenarios created using the modeling service

We developed multiple IPython [30] notebooks to demonstrate the programmatic approach to running an ensemble of models in parallel by using the modeling API by Matthew Tuner, who is a PhD Student in Cognitive and Information Sciences at University of California, Merced. Figure 9, shows an important step of modeling with the PRMS model. By using the modeling API, an user can execute multiple models in parallel, programmatically. Users can use server resources to run a model many, many times without concerning themselves about limited time and CPU resources.

```python
import itertools
fig, ax = plt.subplots()

cax = ax.matshow(nash_sutcliffe_mat, cmap='viridis')
tix = [1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0]
plt.xticks(range(10), tix)
plt.yticks(range(10), tix)


ax.xaxis.set_ticks_position('bottom')
plt.ylabel('jh_coef_hru factor')
plt.xlabel('rad_trncf factor')

for i, j in itertools.product(range(10), range(10)):
    plt.text(j, i, "%.3f" % nash_sutcliffe_mat[i, j],
            horizontalalignment="center",
            color="w" if nash_sutcliffe_mat[i, j] < .54 else "k")

plt.title('Nash-Sutcliffe Matrix')
plt.grid(b=False)
cbar = fig.colorbar(cax)
```
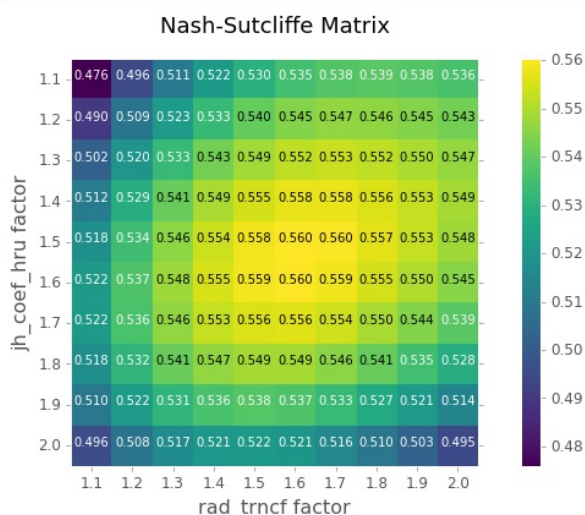

Figure 9: Example of tuning a hundred model through an IPython notebook with different parameters

5 Comparison with Related Tools

A feature based comparison is made between similar software tools focused on hydrologic and environmental modeling in Table 1. The tools discussed in Section 2 are CSDMS [8], Hydroshare [9] and MAAS [10]. Though each of these tools were designed and developed with different goals in mind, all of them are a demonstration of work to assist environmental modeling in general. Each of these tools has their pros and cons from different users perspectives.

CSDMS is a community driver system where modelers can submit their models to CSDMS by implementing model interfaces provided by them. The model gets evaluated and added into the system by the CSDMS committee. CSDMS uses a language interoperability tool called Babel to handle language heterogeneity. From the user's perspective, CSDMS provides a web and desktop based client where users access models and run them using different configurations from within the client. On the server side, CSDMS maintains an HPC cluster where they have all the models and relevant dependencies installed.

Hydroshare [9], on the other hand, is a project started to accommodate data sharing and modeling for hydrologists. The platform is in active development and frequently adds new features. The current version of Hydroshare is more concentrated towards data sharing and discovery for hydrology researchers [31]. Hydroshare also provides programmatic access to its REST API which makes it a good candidate for a data discovery backend for any other similar system.

The MaaS [10] is the most similar project to the Virtual Watershed modeling tool. The main advantage our work has over MaaS is resource management. Dr. Li et al. proposed MaaS, but they use Virtual Machine techniques in their framework. In our system, Docker containers are used, which require fewer resources. Also, we implemented APIs to check and stop a docker container based on model execution status. Existing container orchestration tools, like Docker Swarm, are limited in thier ability to manage containers in this way. Though the approach and implementation of this work differ in various ways, the end goal is similar to what we are trying to achieve. MaaS introduced models as services by creating an infrastructure using cloud platforms. The framework provides users with a web application to submit jobs. The backend that takes care of on-demand provisioning of virtual machines that containing model execution environment setup. MaaS achieves model registration by encapsulating a model as a virtual machine image in an image hub. It provides an FTP based database backend to store the model results.

6 Conclusion and Future Work

A hydrologic model execution web service platform named VWS is described in this paper. The key idea of the proposed platform is to expose model relevant services through python wrapper. This wrapper enables representation of model resources with a common data format (e.g., NetCDF), increasing inseparability. Also, it requires NetCDF format resources, which simplifies model execution on a server node. A user can login once and access all services of the VWS with the authentication/authorization component. Each VWS Component is wrapped in a docker container, which requires less resources than a traditional Virtual Machine. Docker container APIs, including the container termination API, are also implemented to autonomously manage the dynamic resource needs of this system.

This software has significant room to be expanded upon with furutre work. For example, the system currently only provides the option to integrate with GSToRE data backend. This can be improved through implementation of a generic data backend which enables integration with other existing data providers like DataONE, Hydroshare, and CUHASI. This, in turn, could provide better access to data for modeling automation. Developing a pricing component for service usage would be considered for future development. This aspect can help system manager to sustainably

Table 1: Feature Comparison with Related Hydrologic and Environmental Modeling tools

Tool / Feature	CSDMS	HYDRO-SHARE	MAAS	VW-MODEL
Open source	Yes	Yes	No	Yes
Provides REST API	No	Yes	No	Yes
Provides interface for model implementation	Yes	No	No	Yes
Allows model coupling	Yes	No	No	No
Allows model registration	Yes	No	Yes	Yes
Data storage	No	Yes	Yes	Yes
Data sharing and discovery	No	Yes	No	No

maintain the server and hire software developers to implement other useful features.

Conflict of Interest   The authors declare no conflict of interest.

References

[1] Nmepscor.org. The western consortium for water analysis, visualization and exploration — new mexico epscor., 2018. [Accessed on 10 June 2018].

[2] Md Moinul Hossain, Rui Wu, Jose T Painumkal, Mohamed Kettouch, Cristina Luca, Sergiu M Dascalu, and Frederick C Harris. Web-service framework for environmental models. In *Internet Technologies and Applications (ITA), 2017*, pages 104–109. IEEE, 2017.

[3] Russ Rew and Glenn Davis. Netcdf: an interface for scientific data access. *IEEE computer graphics and applications*, 10(4):76–82, 1990.

[4] Danny Marks, James Domingo, Dave Susong, Tim Link, and David Garen. A spatially distributed energy balance snowmelt model for application in mountain basins. *Hydrological processes*, 13(12-13):1935–1959, 1999.

[5] Steven L Markstrom, Robert S Regan, Lauren E Hay, Roland J Viger, Richard M Webb, Robert A Payn, and Jacob H LaFontaine. Prms-iv, the precipitation-runoff modeling system, version 4. Technical report, US Geological Survey, 2015.

[6] Chao Chen, Ajay Kalra, and Sajjad Ahmad. Hydrologic responses to climate change using downscaled gcm data on a watershed scale. *Journal of Water and Climate Change*, 2018.

[7] Chao Chen, Lynn Fenstermaker, Haroon Stephen, and Sajjad Ahmad. Distributed hydrological modeling for a snow dominant watershed using a precipitation and runoff modeling system. In *World Environmental and Water Resources Congress 2015*, pages 2527–2536, 2015.

[8] Scott D Peckham, Eric WH Hutton, and Boyana Norris. A component-based approach to integrated modeling in the geosciences: The design of csdms. *Computers & Geosciences*, 53:3–12, 2013.

[9] David G Tarboton, R Idaszak, JS Horsburgh, D Ames, JL Goodall, LE Band, V Merwade, A Couch, J Arrigo, RP Hooper, et al. Hydroshare: an online, collaborative environment for the sharing of hydrologic data and models. In *AGU Fall Meeting Abstracts*, 2013.

[10] Zhenlong Li, Chaowei Yang, Qunying Huang, Kai Liu, Min Sun, and Jizhe Xia. Building model as a service to support geosciences. *Computers, Environment and Urban Systems*, 61:141–152, 2017.

[11] Michael P McGuire and Martin C Roberge. The design of a collaborative social network for watershed science. In *Geo-Informatics in Resource Management and Sustainable Ecosystem*, pages 95–106. Springer, 2015.

[12] Eric Fritzinger, Sergiu M Dascalu, Daniel P Ames, Karl Benedict, Ivan Gibbs, Michael J McMahon, and Frederick C Harris. *The Demeter framework for model and data interoperability*. PhD thesis, International Environmental Modelling and Software Society (iEMSs), 2012.

[13] Jeffrey D Walker and Steven C Chapra. A client-side web application for interactive environmental simulation modeling. *Environmental Modelling & Software*, 55:49–60, 2014.

[14] Jon Wheeler. Extending data curation service models for academic library and institutional repositories. *Association of College and Research Libraries*, 2017.

[15] Thomas Erl. Service-oriented architecture (soa): concepts, technology, and design. Prentice Hall, 2005.

[16] Roy T Fielding. *Architectural styles and the design of network-based software architectures*, volume 7. University of California, Irvine Doctoral dissertation, 2000.

[17] Alex Rodriguez. Restful web services: The basics. *IBM developerWorks*, 33, 2008.

[18] Leonard Richardson and Sam Ruby. *RESTful web services*. O'Reilly Media, Inc., 2008.

[19] Peter Buxmann, Thomas Hess, and Sonja Lehmann. Software as a service. *Wirtschaftsinformatik*, 50(6):500–503, 2008.

[20] James Lewis and Martin Fowler. Microservices: a definition of this new architectural term. *MartinFowler. com*, 25, 2014.

[21] Sam Newman. *Building Microservices*. O'Reilly Media, Inc., 2015.

[22] Dirk Merkel. Docker: lightweight linux containers for consistent development and deployment. *Linux Journal*, 2014(239):2, 2014.

[23] Md Moinul Hossain. *A Software Environment for Watershed Modeling*. PhD thesis, University of Nevada, Reno, 2016.

[24] Rui Wu. Environment for Large Data Processing and Visualization Using MongoDB. Master's thesis, University of Nevada, Reno, 2015.

[25] Miguel Grinberg. *Flask web development: developing web applications with python*. O'Reilly Media, Inc., 2018.

[26] Rick Copeland. *Essential sqlalchemy.* " O'Reilly Media, Inc.", 2008.

[27] Bruce Momjian. *PostgreSQL: introduction and concepts*, volume 192. Addison-Wesley New York, 2001.

[28] Ask Solem. Celery: Distributed task queue., 2018. [Accessedon 10 June 2018].

[29] Docker. Docker hub, 2018. [Accessed on 10 June 2018].

[30] Fernando Pérez and Brian E Granger. Ipython: a system for interactive scientific computing. *Computing in Science & Engineering*, 9(3), 2007.

[31] Daniel P Ames, Jeffery S Horsburgh, Yang Cao, Jiří Kadlec, Timothy Whiteaker, and David Valentine. Hydrodesktop: Web services-based software for hydrologic data discovery, download, visualization, and analysis. *Environmental Modelling & Software*, 37:146–156, 2012.

Elliptical Printed Dipole Antenna Design using ANN Based on Levenberg–Marquardt Algorithm

Ali I. Hammoodi[1*], Mariofanna Milanova[2], Haider Raad[3]

[1]*Department of Systems Engineering, University of Arkansas at Little Rock, 72204, USA*

[2]*Department of Computer Science, University of Arkansas at Little Rock, 72204, USA*

[3]*Physics Department, Xavier University, 45207, USA*

A R T I C L E I N F O	A B S T R A C T
	A design of elliptical printed dipole antenna based on neural network approach is presented in this paper for recent WiMAX, Bluetooth, WLAN, LTE, and future 5G applications. The dipole patch is printed on top of substrate which has relative permittivity of 2.2 and 0.787mm of thickness. The elliptical dipole antenna is single band and the operational frequency is controlled by the major and minor axes of the ellipse. Neural network composed of one input layer, one hidden layer, and one output layer is implemented using simulated dataset of 24 samples collected using the electromagnetic simulator CST Microwave Studio. The neural network is trained using Levenberg-Marquardt algorithm and we divided the data set into 90% for training, 5% for testing, and 5% for validation. The error, described by the difference between the target data and expected output, is less than 0.005. The neural network model is a promising candidate to implement the design of elliptical dipole antenna to eliminate the simulation time taken by the EM simulator and reduce the cost of using large computing cluster.

1. Introduction

Single band antennas recently become more demanded designs for modern WLAN, WiMAX, and ISM applications due to the small size and uniform radiated gain with minimized side lobes. The light weight feature, low profile, flexibility, conformability, and ease of fabrication increased the demand of printed antennas in integrated circuit of modern wireless systems [1].

There are various tools of electromagnetic simulators have been used to design and analyze variety of antennas and built using computational electromagnetic analysis which aims to solve Maxwell equations for particular design. The three most common software's are: Ansoft HFSS, CST Microwave studio and FEKO. Some of these software's are based on Finite Element Method and some others are based on Finite Difference Time Domain and Finite Difference Frequency Domain as well as Method of Moment. The main disadvantages of these tools are, but not limited to: 1) Complexity of the structure determines the simulation time hence some of the structures could take up to several days , 2) The individual license cost of these tools are extremely high and could go up to thousands of dollars, and 3) requirements of powerful

computing machines to handle the simulation structure of these tools and solve the design in considerable time. As a result of these major drawbacks, the approach we believe is the best candidate to overcome these issues is machine learning. This approach could be utilized to implement microwave circuit and antennas designs despite the fact of long simulation time and by this technique we could obtain cost-effective simulator.

There are different machine learning approaches could be used to implement EM simulators, however, Artificial Neural Network (ANN) is the highly recommend technique due to its capabilities in learning of solving complex and non-linear problems in different aspects. The weight of the neural network adjustment depends on the training algorithm and the given data set. Different training algorithms could be used which give different network performance and depends on the type of the problem we are trying to solve, the size of the training dataset and how we divide them to train the model and number of layers we could use to implement the model [2-4]. The technique of ANN has been implemented on several designs with different functionality. One example of E-shaped slot embedded in microstrip patch antenna where the neural network is used to obtain the effect of feeding point on the antenna performance [5]. Another example of C-Shaped patch antenna

*Corresponding Author: Ali Hammoodi, aihammoodi@ualr.edu

which is implemented using Bayesian Regularization algorithm to calculate the frequency of operation. A compact triple-band PIFA is implemented using ANN for antenna modeling [6].

This paper presents the design of elliptical shape printed dipole antenna using ANN. The ANN constructs of 1 hidden layers of 10 neurons. The Levenberg-Marquardt algorithm based on backpropagation is used to optimize the model. Parametric study and sensitivity analysis have been done on the design of the printed dipole to measure the effect of each design parameters on the frequency of operation. It is noted that the design of the printed dipole antenna is mostly affected by the major and minor axes of the elliptical shape and the gap between the two ellipses. Therefore, the input of the ANN is considered the desire frequency of operation and the output represented by the major axis of the ellipse, minor axis of the ellipse, and the maximum expected gain at the design frequency. The ANN is train using dataset of 24 samples. The dataset was collected using CST Microwave studio, 2017 [7]. The optimized ANN model is achieved when the dataset is split as 90% for training set, and 5% for both validation and testing. The error is less than 0.005 in all the phases as this error representing the difference between the output of the ANN and the expected value. The ANN presented in this paper is a good approach for the design of single band dipole antenna for modern wireless applications.

The follow sections are divided and summarized as: Section two possess antenna geometry and the antenna performance under one case study. Section three describes the neural network model and analysis. Section four concluded the presented work and followed by the used references.

2. Antenna Geometry and Design

The printed dipole antenna is modeled using CST microwave studio, 2017. The scattering parameters (S-parameters) is used to show the performance of the model using EM simulator. Dipole antenna is built on substrate of 2.2 relative permittivity and 0.787 mm of thickness. The dipole is fed at the center by discrete face port. Figure 1 shows the geometry of the printed elliptical dipole.
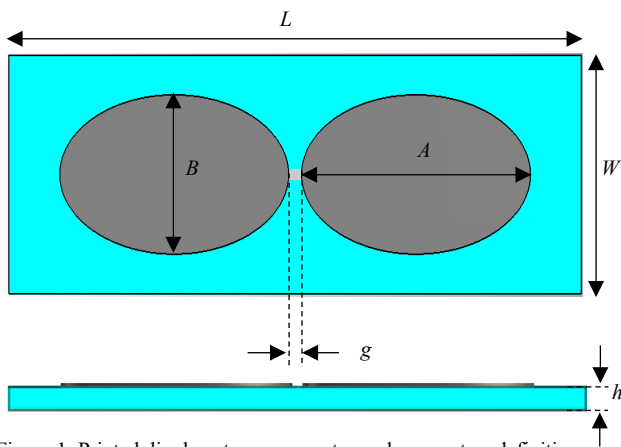
Figure 1. Printed dipole antenna geometry and parameters definition.

The numerical method implemented in the CST EM tool is based on Finit Difference Time Domain which is used to solve Maxwell equations for antennas using time domain analysis. The operational frequency of antenna is represented by the S_{11} and which represemts the ratio of the reflected signal to the incident

signal. By understanding the design philosophy using sensitivity analysis and parameteric study, the operational frequency of the given printed elliptical dipole antenna is controlled by the major and minor axes of the ellipctcal shapes and the gap between the two ellipes is used to achieve impedance matching. One case study of S-parameters is shown in Figure 2 and the design parameters are given in Table 1.

The operational frequency of antenna is given by the point where the S_{11} is $\leq$ -10 dB. This criterion is a must to define valid data set hence the leak of the previous research is addressed by using data set with invalid data points to train the network. As shown in Figure 2, the resonance frequency (f_o) is equal to 6.28 GHz.

Table 1. Design parameters of the case study

Parameter	Value	Parameter	Value
L	27.5 mm	A	11 mm
W	8.25 mm	B	5.5 mm
g	0.624 mm	h	0.787 mm
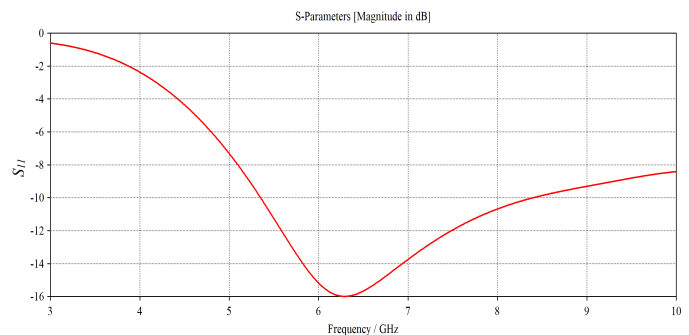
Figure 2. S-parameters of the elliptical shape printed dipole.

3. Neural Network Model

There are several undesired factors encouraged the use of neural network approach for the design of antennas and 3D modeling including, the time the simulator takes to solve the structure, the need of powerful computing unit, and cost of purchasing software's license. As mentioned eariler, machine learning, specifically ANN, is considered as a good candidate to synthesis the design parameters and performance metric of antennas without the need of expesive EM tools. Moreover, implementing designs using 3-D field simulators requiring the understanding of the EM boundaries and requirements for the basic by which the software is built to maintain realistic simulation enviroment.

In this research paper, neural network is used to model the design synthesis of the elliptical printed dipole antenna. The ANN model presented in this paper is built using dataset composed of the S_{11}, realized gain, and antenna's geometry. The implemented network is build of single input layer of one input parameter which is the antenna operating frequency, one optimzied hidden layer of 10 neurons and the output layer is composed of four variables gain (G), major axis (A), minor axis (B), and the gap between the two ellipse (g) as shown in Figure (3).
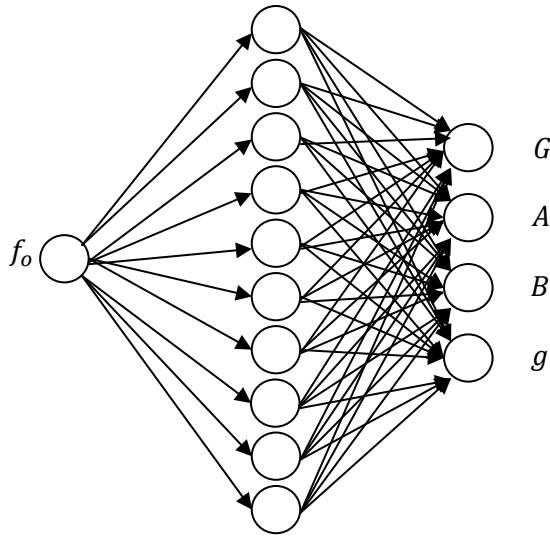
Figure 3. The model of the ANN for the elliptical dipole antenna.

The desired operational frequency is the input to the nerual network and the goal of the ANN is to synthesis the the physical dimenions and the realized gain. The simulated data of 24 samples were collected using the well-known EM simulator CST Microwave Studio, 2017. The major factor to make the ANN model succeed in antenna design is to have valid collected data where the requirement of S-parameters $\leq$ -10 dB is considered. The goal of the implemented ANN is to predict the physical dimenions of the dipole antenna working within the (1-10) GHz as most of the current and modern application work in this frequency band.

The ANN network used is based on the backpropogation technique and the Levenberg-Marquardt algorithm has been used to train the ANN and adjust the weights to predict the parameters of the elliptical printed dipole antenna based on the given operational frequency. In the hidden layer, the neurons compute its trigger value using the activation fucntion given in (1) which shows that each value of basis function for each neuron is a result of multiplying the activation function by the sum of each input multiplies by the weight [2]:

$$\varphi_i(x) = \sigma \sum_k a_{ij} x_k + b_i \qquad (1)$$

where, φ_i is the basis function, σ is the activation function (which is a sigmoid function used in this paper), a_{ik} is the node weight, and x_k is the input value. The system output (y) is described by (2):

$$y = \sum w_i \varphi_i \qquad (2)$$

where, w_i is the orthogonal weight. MATLAB is software utilized to model the network. The dataset is divided into 90% for learning, and 5% for both validation and testing. The output regression of the ANN is shown in Figure 4 and the resultant prediction rate is more than 99%. The regression is described by how the predicted output of the neural network in difference phases are align with the given training data after optimizing the weights. The good regression is when the data are aligned along the slop of the regression (Fit).
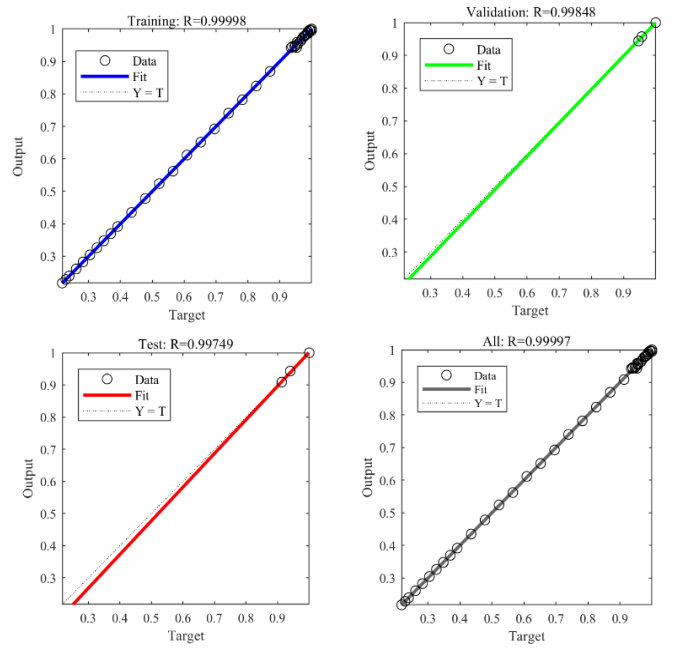
Figure 4. The regression of the neural network in training, validation, and testing phases.

It is noted that the amount of error given by the difference between the output of the ANN and the dataset shown in Figure 5 is below 0.005 over the entire dataset. Therefore, the data set of 24 samples is suffcient to learn the ANN hence the neural network discribed in this paper is a promising candidate for modeling the elliptical printed dipole antenna.
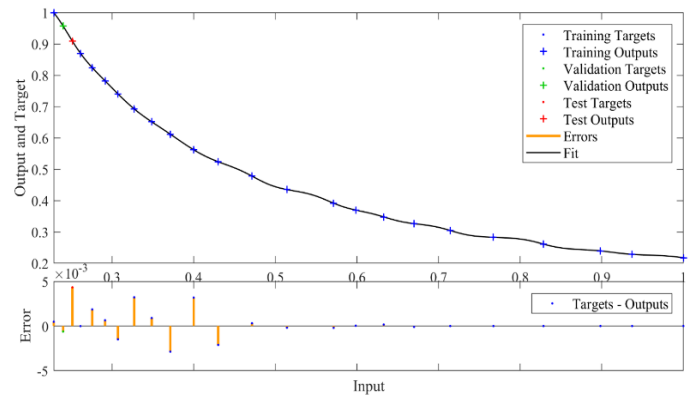
Figure 5. The error between the predicted output and the target output given in the data set.

4. **Conclusion**

This paper presents the design of elliptical printed dipole antenna using machine leanring based neural network. The dipole is printed on top of substrate with 2.2 of relative permittivity and 0.787 mm of thickness. The dipole is fed at the center point by discrete port. Based on sensitivity analysis, the three design parameters control the operation frequency of the dipole; major and minor axis and the gap between the two elliptical poles. The dataset is collected using the CST microwave studio, 2017. The the neural network singe input is the desire frequency of operation and the target output is defined by the physical dimentions and the predicted gain. The traing algorithm is Levenberg-Marquardt

which is based on backpropagation. The data set is divided into 90% for learning, and 5% for both validation, and testing. The error difference between the output of the trained network and the target output is less than 0.005. The ANN is highly recommended candidate for the design of elliptical printed dipole single band antenna to avoid the time consumed using EM tools, and the requirement of powerful computing servers.

References

[1] C. A. Balanis, *Antenna Theory: Analysis and Design*. Wiley, 2016.

[2] Livingstone, D.J. Artificial Neural Networks: Methods and Applications. Humana Press, New York, 2009.

[3] V. S. Kushwah and G. S. Tomar, "Design and Analysis of Microstrip Patch Antennas Using Artifical Neural Network," in Trends in Research on Microstrip Antennas, InTech, 2009, pp. 720–724.

[4] V. V Thakare and P. Singhal, "Microstrip Antenna Design Using Artificial Neural Networks," Int. J. RF Microw. Comput. Eng., pp. 210–224, 2009.

[5] G. Singh and S. Kaur, "Design Analysis of an E-Shaped Slot Loaded MPA and Parameters Estimation using ANN," in *International Conference on Computing, Communication, and Automation (ICCCA2016)*, 2016, pp. 1404–1408.

[6] R. Vanna and J. Ghosh, "Modeling of a compact triple band prF A using Knowledge based neural network," in *2016 IEEE Region 10 Conference (TENCON)*, 2016, pp. 2190–2192.

[7] CST Microwave Studio, 2017.

ASTES

The Effect of Fluence Variations of Nd:YAG Laser Ablation and Sample Condition on Human Tooth

Fatanah Mohamad Suhaimi[*1], Nur Zarifha Zainol Alam[2], Suriani Mat Ariffin[2], Nurul Atiqah Abd. Razak[1], Mohammad Khairul Azhar Abdul Razab[3]

[1]*Craniofacial and Biomaterial Sciences Cluster, Advanced Medical and Dental Institute, Universiti Sains Malaysia, Bertam, 13200 Kepala Batas, Penang, Malaysia*

[2]*School of Physics, Universiti Sains Malaysia, 11800 Gelugor, Penang, Malaysia*

[3]*School of Health Science, Universiti Sains Malaysia, Health Campus, 16150 Kubang Kerian, Kelantan, Malaysia*

A R T I C L E   I N F O

A B S T R A C T

Nd:YAG laser has shown some potential to be used in dental practice replacing the conventional method. In particular, it can be used to modify the tooth surface by the ablation process. The laser provides an ability to accurately deliver a significant amount of energy into a confined region. Thus, alteration of the sample surface and composition may occur during the process. Additionally, the use of a laser in ablation procedure is also associated with heat generation and potential thermal injury that may be experienced by a patient. The wet condition of the samples is expected to reduce the thermal effect. In this paper, the changes of enamel surface and elemental composition following laser irradiation of Nd:YAG laser are discussed. The teeth samples were irradiated at 1.5 Hz of pulse rate, 100 ms of pulse width, and range of fluences of 80-120 J/cm^2. Field Emission Scanning Microscopy (FESEM) and Energy Dispersive X-ray (EDX) were used to analyze the morphology and composition of the teeth samples. Samples were compared before and after laser irradiation. The percentages of carbon increased after laser irradiation, while oxygen decreased for most of the samples. The morphologies of the samples were varied with a more pronounced effect on the sample surface at higher fluence. In addition, the effect of wet sample condition is also investigated and discussed. It is demonstrated that the ablation in wet condition produced less damage to the enamel surface compared to dry sample. However, no remarkable difference between the elemental composition of wet and dry samples.

1. Introduction

Laser has been successfully used in dental applications such as for caries prevention, etching bond strength, restorative procedure, gum recontouring, and surgical procedure [1-3]. Research on medical laser in dental practice has been conducted for several years to investigate the potential and suitability of laser devices and its parameters for replacing the conventional method. However, careful consideration has to be made on selecting parameter settings, types of laser, wavelength, energy, pulse duration, absorption properties and scattering effect. In particular, the interactions of the laser with tissues are pivotal for selecting laser system and operating procedure that will best suit the application in dentistry.

A laser is categorized into four types depending on the medium used, whether it is a liquid, solid-state, gas and semi-conductor. Among all types of laser, Nd:YAG is commonly used in dentistry. The first pulsed Nd:YAG laser was released in 1990 specifically for the dental market [1]. Common laser parameters include fluence, pulse width and pulse rate. Fluence is the optical energy per unit area. Whereas, the pulse width can be defined as the amount of time the laser releases its energy or emits the laser pulse. Pulse rate indicates the number of pulses emitted per second.

In dentistry, Nd:YAG laser has been used for caries prevention [3], germ removal, and modify the bond strength and surface treatment [4, 5]. Nd:YAG laser possesses a suitable wavelength for

[*]Fatanah M. Suhaimi, Craniofacial and Biomaterial Sciences Cluster, Advanced Medical and Dental Institute, Universiti Sains Malaysia, 13200 Kepala Batas, Penang, Malaysia. +6045622561, fatanah.suhaimi@usm.my

high diffusion through hydroxyapatite and water, and soft tissue due to its behaviour of having high penetration depth [6]. Additionally, pre-treatment by Nd:YAG laser irradiation of etch-and-rinse adhesive appeared to have a positive effect on the adhesive-dentin bonding [5]. An in-vitro study by Beketova et al. suggested that Nd:YAG laser has potential in inducing bioactivity of dental ceramic [7].

Eventhough many research had shown the positive effect of using Nd:YAG laser, the impact toward the exposed tissues need to be considered to avoid damage and deterioration of the tissue. Issues related to the ablation process of dental tissue include irregular surface modification, charring, and peripheral cracks [8]. These issues are also associated with dry ablation. An alternative method by lasers for dental ablation characterize the risk involved such as thermal damage experience by patient compare to traditional cavity preparation methods. Moreover, heat generation is crucial as it may lead to painful sensations and potential thermal injury [8-11].

Laser ablation requires sufficient energy to remove material from the tooth by ejecting the electrons from its atom. A study by Cecchini et al. reported that dimineralization of enamel occurs after laser irradiation. Change in morphology is indicated with cracks and crater presence after irradiation [12]. In addition, a research conducted by Khatavkar and Hegde found that Er:YAG laser cause an etching effect on a surface that is comparable to conventional acid etching [13]. Increasing the energy parameters produced a difference surface morphology of enamel from roughening to an etching-like micro-roughen pattern based on a Field Emission Scanning Electron Microscope (FESEM) evaluation.

Histological evaluation on the efficiency of employing ultrashort pulsed laser in dentin and enamel caries removal provided the positive result as no morphological changes and no side effect indicated in the healthy areas of dentin and enamel [8-9]. On the other hand, morphological studies on enamel and dentin irradiated with Nd:YAG laser with different energy densities shows that enamel exhibits shallow crater than dentin due to the fact of varying composition between them. Enamel tissues are found to be more calcified, tight and hard which promotes fissures due to sudden heat distribution throughout the surface. However, the high number of shots resulted in significant changes in morphology with total eradication of smear layer [14].

Besides, the elemental composition of the enamel may also be altered due to the laser irradiation subjected to the sample. According to Al-Hadeethi et al., 2016, radiation by high energy laser melts the enamel surface, which then increased energy density, and thus increases carbon element [15]. Moreover, photothermal evaporation during laser irradiation caused a reduction of water molecules [16]. The changes in water molecules may be assessed by examining the oxygen composition. Interestingly, calcium, phosphorus, and the Ca/P ratio in the sample indicate the hardness of the tooth.

The condition of the sample whether dry or wet plays an important role towards the surface produced after the ablation process. A study by Bartoli et al. on stone cleaning application indicates that ablation in wet condition is more efficient due to vaporization process of water molecules [17]. This statement is also supported by a study conducted on a wet film that suggests the wet condition has less resistant to the laser irradiation [18]. Thus, thermal conductivity, heat transfer and vaporization process that occur during the ablation process may result in different morphological changes of the tooth samples.

This research was done to investigate the effect of Nd:YAG laser fluences on the enamel of the human tooth. Other laser parameters such as pulse width and pulse rate were kept constant. Surface morphology and elemental composition of the enamel were analysed before and after laser irradiation. Additionally, the condition of the sample whether wet or dry is also examined to identify the differences outcome of the laser ablation process. This paper is an extension of work originally presented in the 39th International Conference of IEEE Engineering in Medicine and Biology Society [19].

2. Methods

2.1. Sample Preparation

In this study, seven premolar human teeth were used as the sample. Teeth were collected from the dental clinic of the Advanced Medical and Dental Institute (AMDI), Universiti Sains Malaysia (USM) in Penang, Malaysia. The collected samples were stored individually in a plastic container containing normal saline. Samples were carefully selected, and only healthy teeth were used for the study. Ethical approval for this study has been obtained from the Human Research Ethics Committee of Universiti Sains Malaysia.

Debris on the teeth was removed using a brush and cleaned using water. Then, the cleaned tooth was cut into two using an IsoMet Low-Speed Saw, (Buehler, US). After that, the tooth was mounted in a designated container, and the surface of the mounting was covered with a mixture of resin and hardener. The amount of the resin and the hardener were measured carefully by volume. The sample surface was then polished using a grinding and polishing machine, MetaServ 250 Grinder Polisher (Buehler, Germany). During the grinding procedure, a rotating disk of abrasive paper was used, and it was flushed simultaneously with a coolant to remove debris and minimize the heat.

2.2. Laser Irradiation

Seven premolar teeth were used in this study. Samples were irradiated with Nd:YAG laser (Cynosure, US) at a wavelength of 1064 nm. A 5 mm laser tip was used. The teeth were irradiated with a fluence of 80, 90, 100, 110, and 120 J/cm^2, a pulse rate of 1.5 Hz, and a pulse width of 100 ms. One sample was irradiated for each fluence. Thus, a total of five samples were irradiated in dry condition. The distance of the laser tip was kept constant during the irradiation procedure to ensure a consistent spot size on the enamel surface.

2.3. Sample Analysis

Electron Dispersive X-ray Spectrometer (EDX) and FESEM were used to analyse the samples. The image of the sample was captured, and the morphological analysis was made using the FESEM at several magnifications. The elemental composition in terms of atomic percentage and weight percentage was recorded using the EDX system. The morphology and data on the elemental

composition of the samples were compared before and after laser irradiation.

2.4. Sample Condition

Wet samples were also tested to investigate whether the samples produce different morphology compared to the dry samples. Water spray was applied onto the sample surface before the irradiation process. Two samples were irradiated in the wet condition.

3. Results and Discussion

3.1. Effect of Fluence Variations

The morphology of the enamel surface had been captured before and after laser irradiation with different fluence energies applied to each sample. Figure 1 shows the surface morphology of tooth sample before and after laser irradiation with 80 J/cm^2 at 5000x magnification. The surface appeared rough and non-uniform with large bubbles present after irradiation as compared to before irradiation.

Table 1 presents the identified elements by atomic and weight percentages before and after laser irradiation at 80 J/cm^2. In this table, C is carbon, O is oxygen, Na is sodium, P is phosphorus, Cl is chlorine, Ca is calcium, and Si is silicon. Carbon had the highest atomic percentage after irradiation (54.38%) followed by oxygen (28.20%). Similarly, the weight percentages were 37.78% and 26.10%, respectively. Interestingly, oxygen was the highest element before irradiation. Elements such as Na, P, and Cl were not largely affected after laser irradiation since the atomic and weight percentages were muchly smaller compared to other elements.

Table 1: Identified elements by atomic percentage and weight percentage, before and after laser irradiation at 80 J/cm^2.

Element	Before Irradiation		After Irradiation	
	Atomic	Weight	Atomic	Weight
C	13.31	7.24	54.38	37.78
O	53.87	39.02	28.20	26.10
Na	0.65	0.68	0.40	0.54
P	12.48	17.51	6.09	10.92
Cl	0.70	1.13	0.62	1.25
Ca	18.97	34.42	9.61	22.27
Si	-	-	0.70	1.14
Total	100.00	100.00	100.00	100.00

Figure 2 shows the surface morphology of the tooth sample before and after laser irradiation with 90 J/cm^2 at 5000x magnification. The image consists of large indentations with less bubbles evenly disperse throughout the surface. However, more bubbles were present, and the surface appeared rough after being irradiated by laser at 90 J/cm^2, indicating that the resulted surface had more damages as compared to the sample in Figure 1. However, the surface of the sample before irradiation as shown in

the left panel appeared scalier compared to the first sample in Figure 1.
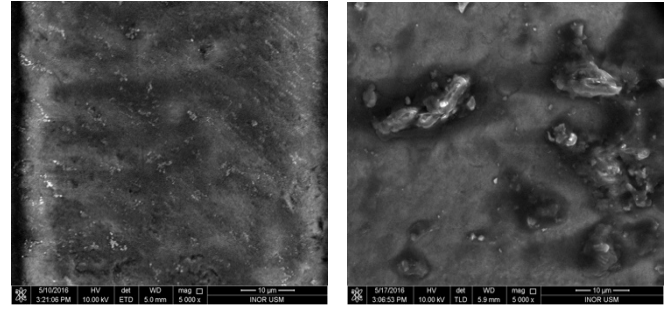
Figure 1: Surface morphology of tooth sample before (left) and after laser irradiation (right) of 80 J/cm^2 at 5000x magnification.
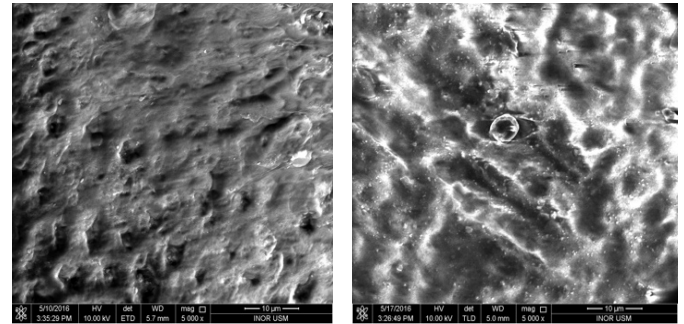
Figure 2: Surface morphology of tooth sample before (left) and after laser irradiation (right) of 90 J/cm^2 at 5000x magnification.

Table 2 presents the identified elements by atomic and weight percentages before and after laser irradiation at 90 J/cm^2. Oxygen, phosphorus, and calcium decreased in both atomic and weight percentages. The percentage difference of weight changes were less than 12% individually for oxygen, phosphorus, and calcium. The highest element after irradiation was oxygen, eventhough oxygen slightly decreased after irradiation. Nevertheless, carbon increased after the irradiation, from 21.33% to 38.85% of atomic percentage. Chlorine and silicon were not extremely modified by the irradiation.

The surface morphology of tooth sample before and after laser irradiation with 100 J/cm^2 at 5000x magnification is shown in Figure 3. Enamel surface appears smooth and shiny with evenly choppy texture. Nonetheless, the appearance of the enamel changed after laser irradiation resulting in a rough, and an uneven bubble distribution on the surface of the sample.

Table 3 presents the identified elements by atomic and weight percentages before and after laser irradiation at 100 J/cm^2. In this table, Al is aluminium. Oxygen, phosphorus and calcium decreased significantly while carbon increased by almost triple the value before irradiation. The atomic percentage of oxygen was largely reduced, which was from 52.26% to 36.68%. Other elements were not significantly changed.

The tooth sample before and after laser irradiation with 110 J/cm^2 at 5000x magnification is shown in Figure 4. The surface appears uniform and rough with fewer indentations before laser irradiation. However, cracks and non-uniform surface were presented after exposing the sample to laser irradiation.

Table 2: Identified elements by atomic percentage and weight percentage, before and after laser irradiation at 90 J/cm^2.

Element	Before Irradiation		After Irradiation	
	Atomic	Weight	Atomic	Weight
C	21.33	12.44	38.85	25.13
O	51.89	40.31	40.01	34.48
Na	-	-	0.74	0.92
P	10.21	15.36	7.57	12.63
Cl	0.42	0.71	0.53	1.01
Ca	15.70	30.56	11.20	24.17
Si	0.45	0.62	1.10	1.66
Total	100.00	100.00	100.00	100.00

Table 3: Identified elements by atomic percentage and weight percentage, before and after laser irradiation at 100 J/cm^2.

Element	Before Irradiation		After Irradiation	
	Atomic	Weight	Atomic	Weight
C	18.66	10.63	44.37	30.19
O	52.26	39.63	36.68	33.25
Na	0.63	0.68	0.55	0.71
P	11.57	16.99	7.66	13.45
Cl	-	-	0.48	0.96
Ca	16.88	32.07	7.59	17.24
Si	-	-	1.91	3.04
Al	-	-	0.76	1.16
Total	100.00	100.00	100.00	100.00
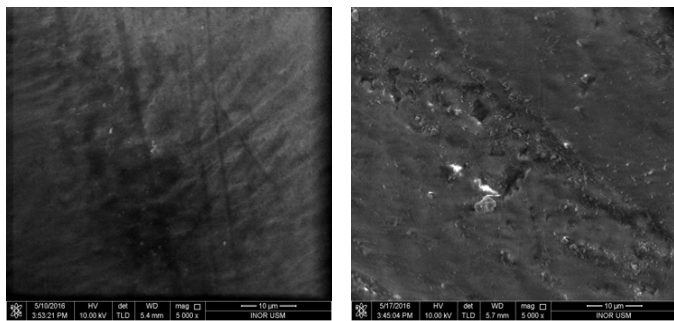
Figure 3: Surface morphology of tooth sample before (left) and after laser irradiation (right) of 100 J/cm^2 at 5000x magnification.

Table 4 presents the identified elements by atomic and weight percentages of the tooth sample before and after laser irradiation at 110 J/cm^2. The percentages of carbon, oxygen, and sodium reduced after laser irradiation. However, the change in oxygen was very light compared to the changes in carbon. Both atomic and weight percentages of phosphorus and calcium were increased. Yet, the changes in calcium were profound compared to

phosphorus. Chlorine and silicon were not detected before irradiation but appeared after irradiation with a very small percentage.
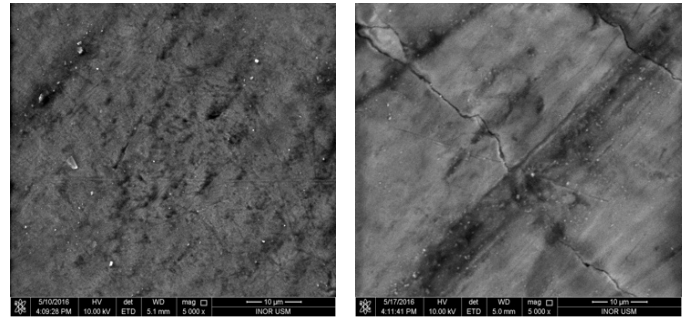
Figure 4: Surface morphology of tooth sample before (left) and after laser irradiation (right) of 110 J/cm^2 at 5000x magnification.
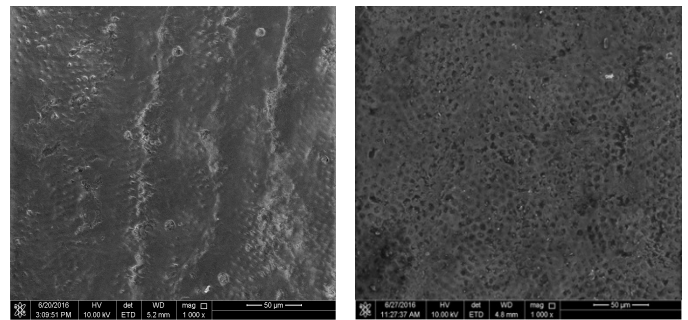
Figure 5: Surface morphology of tooth sample before (left) and after laser irradiation (right) of 120 J/cm^2 at 1000x magnification.

Figure 5 shows the surface morphology of a tooth sample before after laser irradiation at 120 J/cm^2 at 1000x magnification and 500x magnification, respectively. The surface of the sample before irradiation was scaly and uneven. Nevertheless, the surface appeared even with tiny hollows after laser irradiation.

Table 4: Identified elements by atomic percentage and weight percentage, before and after laser irradiation at 110 J/cm^2.

Element	Before Irradiation		After Irradiation	
	Atomic	Weight	Atomic	Weight
C	35.76	23.68	26.18	15.21
O	45.66	40.28	45.26	35.01
Na	0.55	0.70	0.41	0.45
P	8.94	15.26	11.19	16.76
Cl	-	-	0.32	0.56
Ca	9.09	20.08	16.23	31.45
Si	-	-	0.41	0.56
Total	100.00	100.00	100.00	100.00

Table 5 compares the elemental compositions of a tooth sample before laser irradiation and after laser irradiation of 120 J/cm^2. In this table, Mg is magnesium. Before irradiation, oxygen has the highest atomic percentage with 38.29%, followed by calcium with 28.15%, phosphorus with 16.38%, and carbon with 16.24%. The

highest weight percentage is calcium with 45.68%, followed by oxygen, phosphorus, and carbon. Carbon and oxygen increased after laser irradiation. However, other elements decreased after laser irradiation. The percentages of silicon, sodium, magnesium and chlorine were considered not significant compared to other elements in this sample.

Table 5: Identified elements by atomic percentage and weight percentage, before and after laser irradiation at 120 J/cm^2.

Element	Before Irradiation		After Irradiation	
	Atomic	Weight	Atomic	Weight
C	16.24	7.90	35.70	22.32
O	38.29	24.81	40.74	33.92
Na	0.54	0.50	0.30	0.36
P	16.38	20.54	8.52	13.72
Cl	0.40	0.57	0.37	0.69
Ca	28.15	45.68	12.84	26.79
Si	-	-	1.33	1.95
Mg	-	-	0.20	0.25
Total	100.00	100.00	100.00	100.00

3.2. Effect of Wet Sample Condition

Figure 6 shows the surface morphology of the tooth sample before and after laser irradiation of 80 J/cm^2 at a wet condition. The surface after laser irradiation at a wet condition appeared more uniform with tiny holes presented. Additionally, Table 6 compares the identified elements by atomic percentage and the weight percentage of the tooth sample before and after laser irradiation of 80 J/cm^2 at the wet condition. In this table, C is carbon, O is oxygen, Na is sodium, P is phosphorus, Ca is calcium, Si is silicon, and Mg is magnesium.

Oxygen was the highest element before irradiation, followed by calcium and phosphorus. After laser irradiation, the atomic percentage of oxygen reduced significantly while slight changes can be observed in the weight percentage. On the other hand, carbon increased massively to almost six times of its atomic percentage and nine times of its weight percentage, becoming the highest element exist in the sample. Phosphorus and calcium were also decreased after laser irradiation. Silicon was not detected before irradiation, but appeared after irradiation with a very small percentage. Nevertheless, sodium and magnesium were not found in the sample after irradiation.

Figure 7 presents the surface morphology of a tooth sample before and after laser irradiation of 110 J/cm^2 at a wet condition with a magnification of 1000x, and 500x, respectively. The sample appears uniform, smooth and without crack formation. Additionally, Table 7 shows the identified elements by atomic percentage and weight percentage before irradiation and after laser irradiation of 110 J/cm^2 at the wet condition.

Oxygen was the highest element in the sample before irradiation and after laser irradiation, eventhough it was largely

decreased after irradiation. Additionally, carbon also decreased after laser irradiation. Other elements include sodium, phosphorus, and calcium were increased after laser irradiation. However, the increased in calcium was significant compared to the changes in phosphorus and sodium. Chlorine and silicon were not detected before irradiation but appeared after irradiation with a very small percentage.
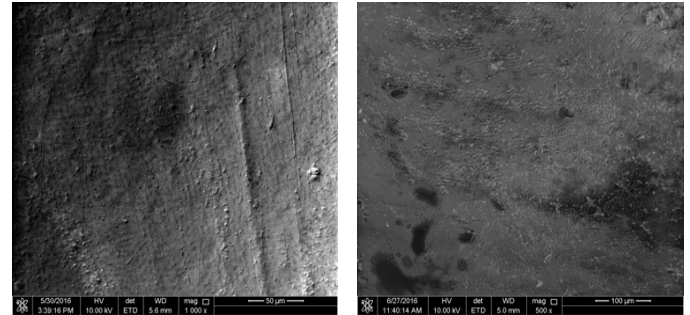
Figure 6: Surface morphology of the tooth sample before laser irradiation (left) at 1000x magnification and after laser irradiation with 80 J/cm^2 at wet condition (right) at 500x magnification.
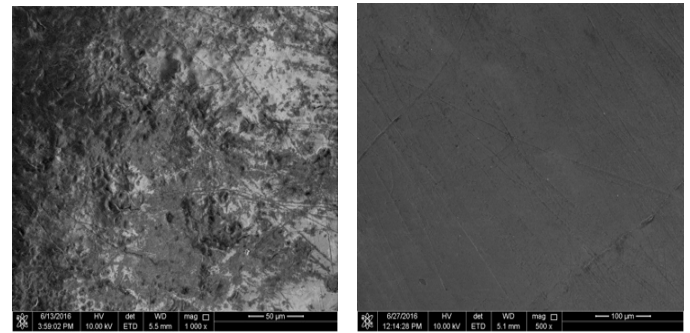
Figure 7: Surface morphology of the tooth sample before laser irradiation (left) at 1000x magnification and after irradiation with 110 J/cm^2 at wet condition (right) at 500x magnification.

Table 6: Identified elements by atomic percentage and weight percentage, before and after laser irradiation of 80 J/cm^2 with wet condition.

Element	Before Irradiation		After Irradiation	
	Atomic	Weight	Atomic	Weight
C	8.66	4.87	49.12	38.46
O	62.60	46.88	43.59	45.46
Na	0.54	0.59	-	-
P	11.90	17.26	4.28	8.63
Ca	16.06	30.14	2.47	6.47
Si	-	-	0.54	0.98
Mg	0.24	0.26	-	
Total	100.00	100.00	100.00	100.00

Figure 8 presents the bar charts on atomic and weight percentage of C, O, P and Ca before irradiation on the top panel and after laser irradiation of 80 J/cm^2, 90 J/cm^2, 100 J/cm^2, 110 J/cm^2, and 120 J/cm^2 on the bottom panel. These four elements

were selected for further comparison across samples as they have more significant values, highly associated with the ablation process, and important element for tooth sample. In this figure, C is carbon, O is oxygen, P is phosphorus, and Ca is calcium.

Table 7: Identified elements by atomic percentage and weight percentage, before and after laser irradiation of 110 J/cm^2 with wet condition.

Element	Before Irradiation		After Irradiation	
	Atomic	Weight	Atomic	Weight
C	22.59	14.93	18.57	10.18
O	60.93	53.63	48.44	35.35
Na	0.50	0.63	0.51	0.54
P	8.84	15.07	12.31	17.40
Cl	-	-	0.55	0.89
Ca	7.14	15.74	19.20	35.11
Si	-	-	0.42	0.53
Total	100.00	100.00	100.0	100.00

There is an abundance of oxygen concerning the atomic and weight percentages before irradiation as observed in Figure 8, followed by carbon, calcium, and phosphorus. However, the percentages of carbon and calcium were varied among these samples, whereas the percentages of phosphorus are about the same. The weight percentage of carbon following irradiation increased tremendously, except for the sample irradiated with a fluence of 110 J/cm^2. After laser irradiation, the percentages of carbon were significantly increased for most of the samples while calcium and oxygen reduced profoundly. Additionally, phosphorus is not majorly changed by the laser irradiation.

In general, the increased and decreased pattern of the four elements with regards to the atomic percentage is about the same across samples. Yet, the percentage differences are highly variable from one sample to another. Only tooth sample that has been irradiated with 110 J/cm^2 has an opposite pattern. These patterns of element changing are almost similar for all the samples presented regardless of the fluence used during the ablation process.

Figure 9 presents the bar charts on atomic and weight percentages of C, O, P and Ca before irradiation and after irradiation with 80 J/cm^2 and 110 J/cm^2 at wet condition. The trend of tooth sample irradiated with 80 J/cm^2 is similar to the overall trend seen in Figure 8. Nevertheless, the trend of tooth sample irradiated with 110 J/cm^2 is contrary.

Comparing the same fluence used at 80 J/cm^2 but at different sample conditions, wet and dry, the percentage of carbon increased almost similar. However, the oxygen content of the dry sample is relatively higher compared to the dry sample. However, the changes in oxygen before and after irradiation for both conditions are about the same. In contrast, the percentage difference of calcium before and after irradiation was less for the wet sample. Thus, calcium is not muchly modified by the laser irradiation in the wet condition.

The samples irradiated with 110 J/cm^2 resulted in a lower percentage of carbon either at the wet or dry condition. The percentage of oxygen also decreased eventhough the wet sample has a larger percentage difference before and after laser irradiation. The percentages of phosphorus and calcium are not much different with regards to the wet or dry condition of the sample.

Theoretically, the higher the energy applied to the samples, more interactions occur giving more energy to the atom, and thus resulted into modifications in both morphology and mineral composition. Formation of cracks, shallow pit and fissures can be seen after higher fluence was applied. Nevertheless, wet samples appear uniform, flat, smooth and less crack formation as shown in Figures 6-7. Shallow indentation appearance on dry sample provides information on the rough surface caused by the laser with huge bubbles presented.

Inconsistent results of the tooth elements modifications show that the effect of the laser varies. Technically, it relies on the laser parameters include fluence energy, pulse width and pulse rate. However, the teeth samples used also affect the outcome. Despite the difference fluence used during ablation, the surface morphology and elemental composition of the collected teeth samples were also varied. Thus, the outcome is also expected to be varied, and the difference in the percentages of the elemental composition are assessed to conclude.

Carbon undergoes percentage increment after laser interaction with the samples in almost all fluence energies applied except irradiation at 110 J/cm^2. This explains the undesired carbonization phenomena within the tissues. As the temperature increases, the tissues tend to experience dehydration and eventually burned. Based on the ablation methodology, the absorption of laser beam energy process occurs on the tooth surface makes the surfaces become more carbonize [15, 20].

In contrast, carbon composition of the sample irradiated with 110 J/cm^2 experienced reduction, which is unexpected. This might due to the natural condition of the tooth sample. The carbon content of a tooth sample is correlated to several factors include enamel maturity, diets intake and caries susceptibility [21]. Increase caries susceptibility has been associated with the increase in carbon element [22]. Additionally, hypoplastic enamel contains more carbon compared to well-calcified enamel. Thus, the collected samples may vary in terms of enamel density, the number of enamel layers and the weight of enamel layers. In particular, the variation in enamel density results into the variation of optical property and laser absorption of the enamel layer. Thus, the contradict changes in carbon percentage is multifactorial and possibly due to the optical property of the sample itself.

In the case of the sample that has been exposed to a fluence of 80 J/cm^2, the carbon content appears to be the highest, even after the lowest fluence exposure. Photothermal interactions of the laser with tissues include absorption, penetration, scattering and reflection. The undesired carbonization occurs due to the higher interactions with biochemical compounds in the tissue sample. Similarly, this is potentially due to the enamel density and optical property of the sample.

Although the elemental composition might not largely affect, the surface morphology seems to have better form when the
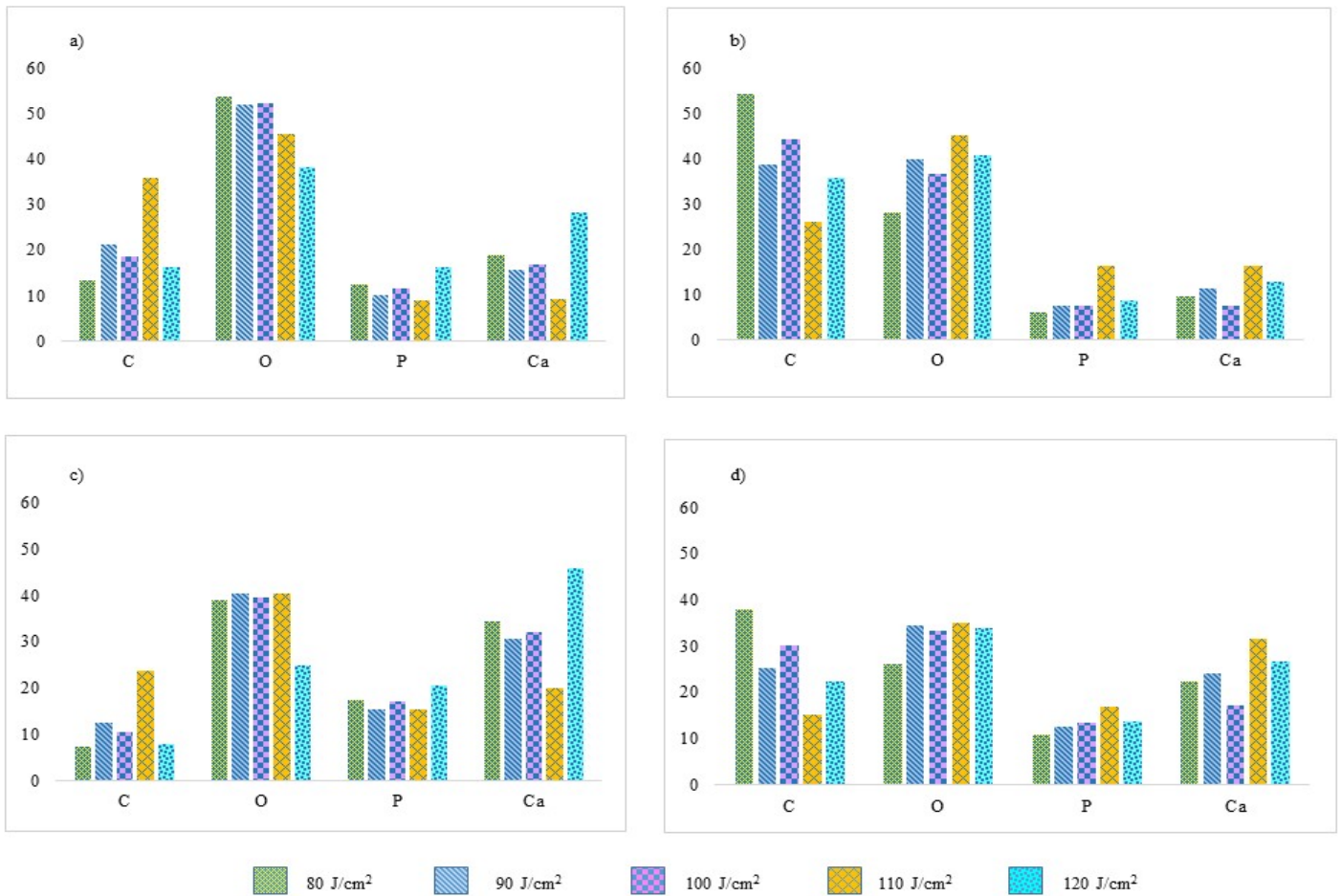
surface is in the wet condition during laser irradiation. The wet condition of the samples changes the morphological appearance of a tooth surface as shown in Figures 6-7. Rough texture resulted from the dry sample leads to heat generation and thus charring pattern on the tooth surface. At a higher laser fluence, tooth surface may appear with a burning effect and peripheral cracks.

The application of water spray has been associated with low laser-ablation efficiency, as reported in several studies. The use of water spray also ensures the rehydration of mineral contains in the tissue [23]. This is particularly important for Nd:YAG laser ablation in which the influence of the water layer on maintaining the mineral content while preventing stalling and excessive peripheral damage. However, for dental application, the use of water spray is necessary to avoid heat sensation to the patients. A fluence up to 100 J/cm^2 seems acceptable to be used in addition to the wet condition of the sample, as seen from this study.

There are some limitations in this study. It is almost impossible to gather samples that have the similar amount of elemental composition. The compositions are highly variable among all the samples, and some other confounding factors such as enamel density, maturity and calcification may contribute toward the variation. Moreover, morphological appearance of one sample to another was also varied. Some of the samples may have

been through a dental treatment which changed the morphological and elemental composition. Thus, comparison of the morphology and elemental composition has to be assessed based on the changes before and after laser irradiation.

As a suggestion, a larger number of samples can provide more statistically significant of the reduction and increment of the elemental composition. Selecting teeth sample from the same subject may reduce the variation. Additionally, the thickness of the sample can be standardized to minimize enamel layer variation, and topography analysis have to be carried out prior to the sample selection. Thus, by considering the sample selection criteria, almost similar or identical enamel characteristics can be selected.

4. Conclusions

Laser irradiation affects dental hard tissues physically and chemically as seen from the percentage difference of elemental compositions and the resulting surface morphology of the samples in this study. This study shows that the irradiation of 100 J/cm^2 obtained less damage on enamel surface compared to the results of higher fluence irradiation. In general, the percentage of carbon increased as an effect of the ablation process, whereas the percentage of oxygen decreased. Phosphorus is not significantly
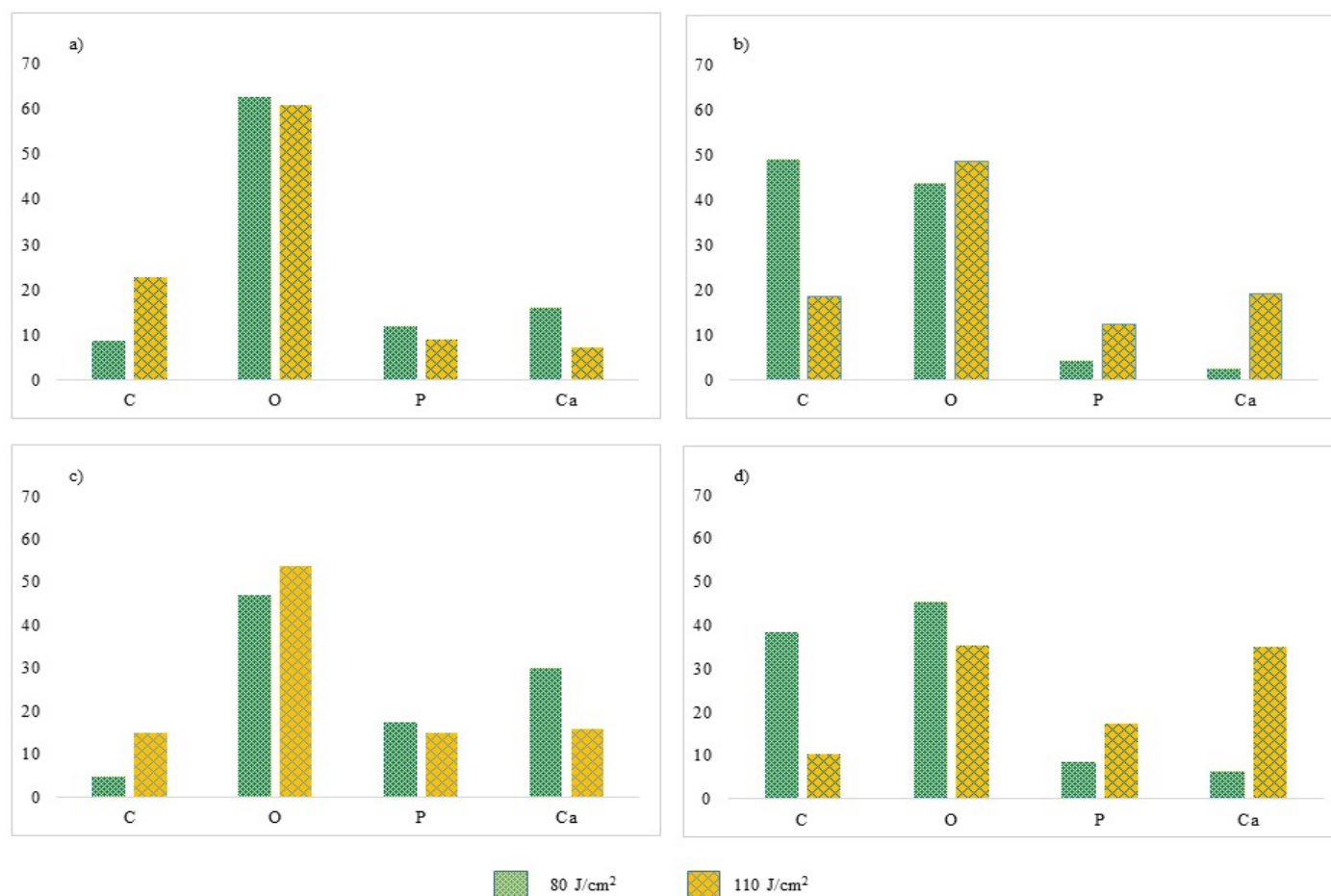
Figure 9: Bar charts of C, O, P and Ca before laser irradiation in terms of a) atomic percentage and b) weight percentage. Bar charts of C, O, P and Ca after laser irradiation of 80, and 110 J/cm^2 at wet condition in terms of c) atomic percentage and d) weight percentage.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

References

[1] S. K. Verma, S. Maheshwari, R. K. Singh, P. K. Chaudhari, "Laser in dentistry: An innovative tool in modern dental practice" Natl J Maxillofac Surg, 3(2): 124-132, 2012. https://doi.org/10.4103/0975-5950.111342

[2] M. Migliario, M. Rizzi, A. G. Lucchina, F. Renò, "Diode laser clinical efficacy and mini-invasivity in surgical exposure of impacted teeth" J Craniofac Surg., 27(8), e779-e784, 2016. https://doi.org/10.1097/SCS.0000000000003128

[3] Y. Rezaei, H. Bagheri, M. Esmaeilzadeh, "Effects of laser irradiation on caries prevention" J Lasers Med Sci, 2(4):159-64, 2011. http://dx.doi.org/10.22037/2010.v2i4.2710

[4] S. Kasraei, L. Rezaei-Soufi, E. Yarmohamadi, A. Shabani, "Effect of CO$_2$ and Nd:YAG lasers on shear bond strength of resin cement to zirconia ceramic" J Dent (Tehran), 12(9), 686–694, 2015.

[5] J. Gan, S. Liu, L. Zhou, Y. Wang, J. Guo, C. Huang, "Effect of Nd:YAG laser irradiation pretreatment on the long-term bond strength of etch-and-rinse adhesive to dentin" Oper Dent., 42(1), 62-72, 2017. https://doi.org/10.2341/15-268-L

[6] N. Gutknecht, "State of the Art in Lasers for Dentistry" J. Laser Health Academy, 2008(3), 1-5, 2008.

[7] A. Beketova, N. Poulakis, A. Bakopoulou, T. Zorba, L. Papadopoulou, D. Christofilos, N. Kantiranis, G. A. Zachariadis, E. Kontonasaki, G. A. Kourouklis, at al., "Inducing bioactivity of dental ceramic/bioactive glass composites by Nd:YAG laser" Dent Mater., 32(11), e284-e296, 2016. 10.1016/j.dental.2016.09.029\

[8] A. Braun, R. F. Krillke, M. Frentzen, C. Bourauel, H. Stark, F. Schelle, "Heat generation caused by ablation of dental hard tissues with an ultrashort pulse laser (USPL) system" Lasers Med Sci, 30(2), 475–481, 2015. https://doi.org/10.1007/s10103-013-1344-z

[9] A. Braun, R. J. Wehry, O. Brede, C. Dehn, M. Frentzen, F. Schelle "Heat generation caused by ablation of restorative materials with an ultrashort pulse laser (UPSL) system" Lasers Med Sci 27(2), 297-303, 2012. https://doi.org/10.1007/s10103-010-0875-9

[10] L. Kuščer, J. Diaci, "Measurement of ebium laser-ablation efficiency n hard dental tissues under different water cooling conditions " J Biomed Opt., 18(10), 108002, 2013. https://doi.org/10.1117/1.JBO.18.10.108002

[11] D. Fried, J. Ragadio, M. Akrivou, J. D. Featherstone, M.W. Murray, K. M. Dickenson, "Dental hard tissue modification and removal using sealed transverse excited atmospheric-pressure lasers operating at lambda=9.6 and 0.6 microm" J Biomed Opt, 6(2), 231-238, 2001. https://doi.org/10.1117/1.1344192

[12] R. C. Cecchini, D. M. Zezell, E. de Oliveira, P. M. de Freitas, P. Eduardo Cde, "Effect of Er:YAG laser on enamel acid resistance: morphlogical and atomic spectrometry analysis." Lasers Surg Med, 37(5), 366-72, 2005. https://doi.org/10.1002/lsm.20247

[13] R. Khatavkar, V. Hegde, "Surface analysis of Erbium:YAG laser etching v/s acid etched surface ESEM observations in vitro study" Laser, 2(1), 2010.

[14] S. Al-Jedani, Y. Al-Hadeethi, M. S. Ansari, M. A. N. Razvi, "Dental hard tisssue ablation with laser irradiation" Austin Dental Sciences, 1(1), 1-10, 2016

[15] Y. Al-Hadeethi, S. Al-Jedani, M. A. Razvi, A. Saeed, A. M. Abdel-Daiem, M. S. Ansari, S. S. Babkair, N. A. Salah, A. Al-Mujtaba, "Data fitting to study ablated hard dental tissues by nanosecond laser irradiation" PLoS One, 11(5), e0156093, 2016. https://doi.org/10.1371/journal.pone.0156093

[16] L. E. Rodríguez-Vilchis, R. Contreras-Bulnes, I. Sánchez-Flores, E. C. Samano, "Acid resistance and structural changes of human dental enamel treated with Er:YAG laser" Photomed Laser Surg., 28(2), 207-11, 2010. https://doi.org/10.1089/pho.2008.2454

[17] L. Bartoli, P. Pouli, C. Fotakis, S. Siano, R. Salimbeni, "Charaerization of stone cleaning by Nd:YAG lasers with different pulse duration" Laser Chemistry, 2006 (8175), 1-6, 2006. https://doi.org/10.1155/2006/81750

[18] I. A. Watson, R. K. Wang, I. Peden, G. D. Ward, D. E. S. Stewart-Tull, A. C. Wardlaw, "Effect of laser and environmental parameters on reducing microbial contamination of stainless steel surfaces with Nd:YAG laser irradiation" Journal of Applied Microbiology, 99, 934-44, 205. https://doi.org/10.1111/j.1365-2672.2005.02665.x

[19] F. M. Suhaimi, N. Z. Zainol Alam, S. Mat Ariffin, N. A. Abd. Razak, M. K. A. A. Razab, "Surface modifications of human tooth using Nd:YAG laser for dental applications" 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Seogwipo, 2017, 4537-4540. https://doi: 10.1109/EMBC.2017.8037865

[20] W. Raucci Neto, C. P. Lepri, J. J. Faraoni Romano, F. S. Fernandes, de L. M. Castro Raucci, L. Bachmann, R. G. Dibb, "Chemical and Morphological Changes of Primary Teeth Irradiated with Nd:YAG Laser: An Ex Vivo Long-Term Analysis" Photomed Laser Surg., 33(5), 266-73, 2015. https://doi.org/10.1089/pho.2014.3876

[21] M. Sydney-Zax, I. Mayer, D. Deutsch, "Carbonate content in developing human and bovine enamel" J Dent Res., 70(5), 913-6, 1991. https://doi.org/10.1177/00220345910700051001

[22] M. F. Little, F. Brudevold, "A study of the inorganic carbon dioxide in intact human enamel" J Dent Res., 37(6), 991-1000, 1958. https://doi.org/10.1177/00220345580370062001

[23] H. W. Kang, I. Rizoiu, A. J. Welch, "Hard tissue ablation with a spray-assisted mid-IR laser" Phys Med Biol, 52(24), 7243-59, 2007. https://doi.org/10.1088/0031-9155/52/24/004

An Approach for Determining Rules used to Select Viable Junction Design Alternatives Based on Multiple Objectives

Erwin Bezembinder[*,1], Luc Wismans[2], Eric van Berkum[2]

[1] *Windesheim University of Applied Sciences, Zwolle, The Netherlands*

[2] *University of Twente, Enschede, The Netherlands*

A R T I C L E I N F O

A B S T R A C T

Transport planners and engineers frequently face the challenge to determine the best design for a specific junction. Many road design manuals provide guidelines for the design and evaluation of different junction alternatives, however these mostly refer to specialized software in which the performances of design alternatives can be modelled. In the first stage of the design process, such assessments of many alternatives are undesirable due to time and budget constraints. There is a need for quick design rules which need limited input data. Although some of these rules exist, their usability is limited due to inconsistencies in rules and non-transparency in combination with objectives. In this paper, we present an approach by which consistent and transparent junction design rules can be determined. The resulting rules can be used to predict a set of viable junction design alternatives for the first stage of the junction design assessment process. The predicted set is in fact the Pareto optimal set of solutions for multiple objectives, e.g. regarding operational, safety and/or environmental impact. The Pareto optimal set of solutions always contains the best solution, whatever set of weights is used for different objectives in a later stage of the assessment process, thus handling multiple objectives in a straightforward manner. The rules are derived from a dataset by using decision tree data mining techniques. For this, a large dataset is first generated, using performance models, with Pareto optimal sets of junction design alternatives for a large amount of, randomly generated, traffic volumes. The approach is applied and evaluated on cases for two different countries. Results show that for over 90% of the situations the Pareto optimal set can be predicted by the new rules, whereas existing rules hardly reach 33%. The new rules provide junction design alternatives with a better performances.

1 Introduction

This paper is an extension of work originally presented in 2017 IEEE 20th International Conference on Intelligent Transport Systems [1]. Where our previous work was a first evaluation of using traditional decision tree algorithms in order to predict a set of (Pareto optimal) junction design alternatives, we now extended the approach, applied and evaluated it on cases for two different countries and compared the results with existing junction design rules by comparing both accuracy and performance.

Junction design involves both the choice for the main type, such as signalised junction or roundabout, as well as choices concerning the number and configuration of the approach and exit lanes, priority handling, slow traffic crossing facilities, signal control and geometric attributes such as the lane length and the central reservation width. The need to determine the best junction design is not only triggered by the construction of new infrastructure. Due to changes in traffic volumes, travel routes and vehicle types there is a regular need for re-evaluation of the junction design.

[*]Erwin Bezembinder, Windesheim University of Applied Sciences, +31 884698436 & e.bezembinder@windesheim.nl

The junction design assessment process, which is used to select the best alternative, typically involves three stages [2]. In the first stage, viable alternatives are identified based on limited input data, such as the average annual daily traffic volume, using decision rules, in the form of trees, look-up tables or simple calculation methods, provided in design manuals. In the second stage, more detailed input data, such as the peak hour traffic volumes for each turning movement and specific geometric and control attributes, are collected and the operational, safety and environmental performances are determined by using tools ranging from static analytical to dynamic simulation models. In the third stage, the best alternative is selected based on multiple criteria. The performance measures determined in the previous stage are analysed and weighted in combination with other criteria such as specific local constraints and cost, before selecting the best alternative. This three-stage approach is used to avoid time and money consuming assessment of many alternatives, but possesses the risk of omitting the best solution due to a deficient identification of viable alternatives in the first stage of the assessment process [3]. This deficiency is mostly caused by a lack of consciousness concerning which objectives are served by the rules used in this stage.

The availability and quality of the decision rules used to identify the viable alternatives varies by state, region or country. The rules have been developed over a period of many years based on a combination of data collection and expert judgement. Obviously, this is a positive feature, but also comprises some disadvantages. First, the decision rules do not offer a complete and consistent approach for all junction design alternatives. Some rules are only meant for signalised junctions where others only encompass whether a left-turn bay is needed on a major road approach to a priority junction. Some rules are formal warrants where others are informal guidelines. Second, many decision rules have not been updated for many years and thus do not reckon with changes in traffic behaviour and vehicle types and do not include relatively new junction design alternatives. Third, rules are often based on one underlying objective, generally referring to the operational performance, causing junction design alternatives that are better or best for other objectives to be neglected. Other rules implicitly contain a weighing or preferred order for different objectives, which restricts and complicates the assessment of multiple objectives in a later stage of the junction design assessment process.

Surprisingly, only limited efforts have been made to develop new junction design rules. [4] used the HCM 2010 methodologies [5] to distinguish between different junction types based on the major and minor street volumes. They generated a dataset with more than 6,000 scenarios of varying traffic flows for three main junction types. The results were plotted in 2- and 3D diagrams using the major an minor volumes and the average control delay classified by junction type, in order to identify the areas where a junction type has the lowest delay. [6] used the HCM 2000 method-

ologies [7] to determine so-called shape-grammars for junction type choice based on the total traffic volume and the through traffic share for three and four arm junctions resulting in look-up tables. Although these studies provided interesting approaches, they examine a limited number of junction designs and demand volumes and moreover, they conducted a manual analysis of the generated data. [8] generated a dataset with the operational performance of 1,296 scenarios for different junction design types and traffic patterns using VISSIM, determining the total delay, stop delay and the average number of stops per vehicle. A classification tree method was then used to group the scenarios into as many homogeneous classes as possible. Classes were constructed for capacity shortage as well as delay. Based on these classes one can easily discover the average operational performance for one or more junction design alternatives. Still, the resulting classes do not provide a set of viable junction design alternatives and are only based on the operational objective.

In this paper, we present an approach by which new junction design rules can be determined. The resulting rules can be used to predict a set of viable junction design alternatives for the first stage of the junction design assessment process. The predicted set is in fact the Pareto optimal set of solutions for multiple objectives, e.g. regarding operational, safety and/or environmental impact. The Pareto optimal set of solutions always contains the best solution, whatever set of weights is used for different objectives in a later stage of the assessment process, thus handling multiple objectives in a straightforward manner. The rules are derived from a dataset by using decision tree data mining techniques. For this, a large dataset is needed, which is first generated, using performance models, with Pareto optimal sets of junction design alternatives for a large amount of, randomly generated, traffic volumes. The approach is applied and evaluated on cases in two different countries.

In the next section, we will first explain the suggested approach. Subsequently, in section 3, the evaluation framework will be discussed. In section 4 the existing junction design rules and the implementation of the approach for the two cases will be explained. The results are discussed in section 5. The paper ends with conclusions in section 6.

2 Approach

In this section we will explain our approach for determining junction design rules used to select viable junction design alternatives based on multiple objectives. The approach consists of three major steps:

1. Define the scope

2. Generate the dataset

3. Determine the decision tree

In the first step, choices are made concerning the junction design alternatives, the traffic flow ranges and

the objectives to be considered. In the second step, a dataset is generated containing a Pareto optimal set of junction design alternatives for each, randomly generated, traffic flow pattern. Each set is determined after running performance models for each objective and junction design alternative. In the third step, decision tree algorithms are used to derive rules from the dataset. The steps will be explained in more detail in the next paragraphs.

2.1 Define the scope

In the first major step, choices are made concerning the:

- Junction designs
- Traffic flows
- Objectives
- Performance models

First, the junction design alternatives to be considered should be defined. This is a list of all the possible junction design alternatives that should be evaluated. Junction design involves both the choice for the main type, such as signalised junction or roundabout, as well as choices concerning the number and configuration of the approach and exit lanes, priority handling, slow traffic crossing facilities, signal control and geometric attributes such as the lane length and the central reservation width. The level of detail depends upon the requirements for the specific case. An important issue involves the necessity to classify the junction design alternatives by a size category. This attribute was introduced in order to prevent large (or expensive) junctions always to be the preferred design, regardless of the traffic flow volumes on the junction.

Second, it should be decided for which traffic flow ranges the junction design alternatives should be evaluated. The traffic flow at least involves the turning volumes for the motorised vehicles on the junction. Additionally, pedestrian and bicycle volumes can be used as well. Typically, peak hour volumes are used, since these are major input variables for the performance model(s) used. The range, i.e. the total traffic volume on the junction, is case specific and dependent upon the chosen junction design alternatives.

Third, it should be decided based on which objectives the junction design alternatives should be evaluated. Typically, minimising the operational performance, e.g. the average delay, is used. Additionally, safety, environmental and financial performance measures can be used. The choice is strongly influences by the possibilities to model the performance measures. The suggested approach assumes at least two objectives in order to determine Pareto optimal sets of junction design alternatives.

Fourth, in close connection with the other choices, it should be decided which performance model will be used. The model uses the traffic flows and junction design alternative as input. There is a wide range of

models available to determine the performance for a specific objective, ranging from static analytical to dynamic simulation models. The choice depends upon the models and underlying methodologies that are accepted in a specific country, and/or the accepted level of detail of input data or the calculation time of the models. Separate models can be used for separate objectives.

In order to illustrate our approach, we will use a simplified example case. Table 1 shows six four-arm junction design alternatives, with four main junction types, being all-way stop controlled junction (AW), two-way stop controlled junction (TW), signalised junction (SIG) and roundabout (RA). Two size categories (Sc) are differentiated and the lane number and configuration for the approach and exit lanes of the major and minor roads are defined for each alternative. The lane configuration ⇧ means that there is one lane which is shared by left turning, straight and right turning traffic, whereas ⇧ means that there are two lanes; a separate lane for left turning traffic and a shared lane for straight and right turning traffic.

Table 1: Simplified example with the definition of six junction design alternatives

Nr	Id	Type	Sc	Major road		Minor road	
				Approach	Exit	Approach	Exit
1	AW1	AW	1	1 ⇧	1	1 ⇧	1
2	TW1	TW	1	1 ⇧	1	1 ⇧	1
3	SIG1	SIG	1	1 ⇧	1	1 ⇧	1
4	TW2	TW	2	2 ⇧	1	1 ⇧	1
5	SIG2	SIG	2	2 ⇧	1	1 ⇧	1
6	RA2	RA	2	1 ⇧	1	1 ⇧	1

For the traffic flows, only motorised vehicles are used, where the total traffic flow on the junction is assumed to range between 0 and 2000 pcu/h. Two objectives, being operation and safety will be used, respectively expressed by the (volume weighted) average delay (s) and the estimated number of fatal injuries per year. A fictitious performance model is assumed to generate the measures for this example case. Additionally, a delay threshold of 50 seconds is used, in order to exclude junction design alternatives with an undesirable operational performance.

2.2 Generate the dataset

In the second major step, a dataset is generated that contains a set of junction design alternatives for each combination of traffic demand pattern and size category. This set of junction design alternatives for each combination is the Pareto optimal set of solutions for multiple objectives. The Pareto optimal set is determined by determining the performances for each junction design alternative given a (randomly generated) traffic demand pattern. The performances are determined by using a traffic performance model that uses the traffic demand pattern and junction design as input. It is necessary to generate the data by using performance models, since consistent and comprehensive

data is not available from surveys.

Objective	Performance measure	Value
Operation	Average delay (s)	12.31
Safety	Fatal injuries per year	0.42
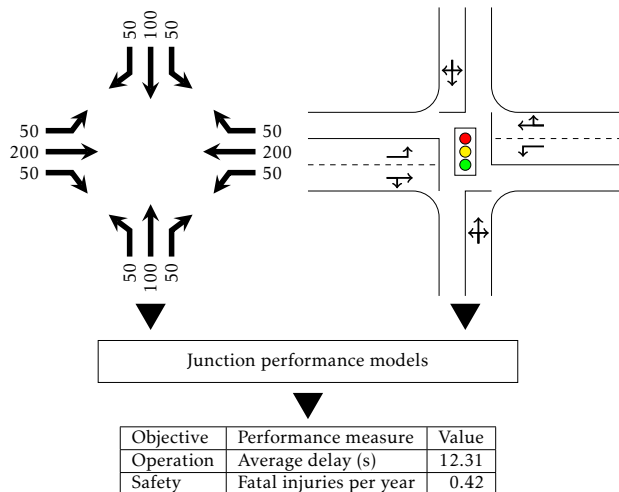
Figure 1: Schematic representation of performance modelling for one specific traffic demand pattern and junction design alternative

Figure 1 shows a schematic representation of the performance modelling for one specific traffic demand pattern (with a total volume of 1000 pcu/h) and a junction design alternative (SIG2 from Table 1). This results in a (volume weighted) average delay of 12.31 seconds and an estimated number of fatal injuries of 0.42 per year. For the given traffic demand pattern, this is repeated for each junction design alternative that has been defined. The operational and safety performances of the six junction design alternatives are shown in Figure 2 for both size categories.
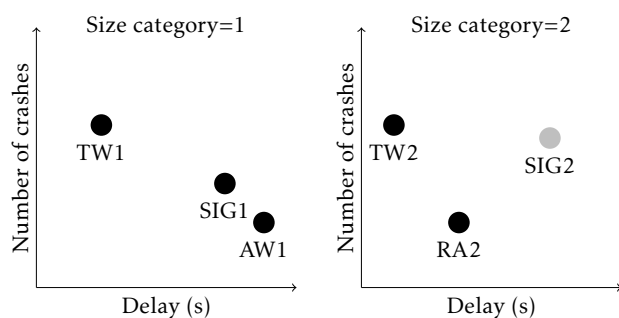
Figure 2: Set of junction design alternative solutions for two different size categories

The main question now is, which junctions design alternatives should be part of the viable set of solutions? The best choice would be the Pareto optimal set of solutions. The Pareto set consists of all solutions for which the corresponding objective values cannot be improved for one objective, without degradation of the other. As said, the Pareto optimal set of solutions always contains the best solution, whatever set of weights are used for the operational or safety performance in a later stage of the assessment process. The Pareto optimal solutions can be determined automatically. In our example, for the first size category, the

Pareto optimal solution would be {AW1,TW1,SIG1}. For the second size category it would be {TW2,RA2}. Additionally, solutions with a performance above a certain threshold value, e.g. a specific average delay, could be removed from the viable solution set, in order to prevent analysis of unrealistic solutions in a later stage of the assessment process.

The described process is repeated for multiple (randomly generated) traffic demand patterns. The number of traffic demand patterns to be generated is case specific and should be determined based on the accuracy of the tree to be predicted in the next main step. The accuracy represents how well the tree predicts the Pareto optimal sets that are used as input. This will be explained in more detail in section 3. In earlier experiments [9] we tested the approach for 10-thousand, 100-thousand and 1-million traffic demand patterns. Although, the predictive accuracy of the trees increased with bigger datasets, the differences between the set sizes were minimal. Table 2 shows an example dataset for only four traffic demand patterns. The dataset contains both the variables representing the base traffic pattern, i.e. the volumes for the twelve turning movements as shown in Figure 1 (v1-v12), as well as multiple combined variables such as the total volume on the major (vMa) or minor road (vMi) or the whole junction (vTot). The combined variables are important for the rules to be determined.

Table 2: Simplified example dataset with four traffic demand patterns

Id	v1	v2	...	v12	vMa	vMi	vTot	Sc	Set
1	25	100		25	300	200	500	1	{AW, TW}
2	25	100		25	300	200	500	2	{TW, RA}
3	50	200		50	600	400	1000	1	{AW, TW, SIG}
4	50	200		50	600	400	1000	2	{TW, RA}
5	75	300		75	900	600	1500	1	{SIG}
6	75	300		75	900	600	1500	2	{SIG, RA}
7	100	400		100	1200	800	2000	1	{OTHER}
8	100	400		100	1200	800	2000	2	{RA}

The table also shows a solution set {OTHER}. For this combination of traffic demand pattern and size category, there was no junction design alternative with a delay below the chosen threshold value (50 seconds). The set is included in the dataset in order to facilitate rules that advise an 'other' solution for the specified combination of traffic demand pattern and size category. The created dataset is the foundation of the training set used to determine the decision tree (and the rules) in the next major step.

2.3 Determine the decision tree

In the third major step, the decision tree is determined. Decision trees use a white box model, are easy to understand and interpret, perform well on large datasets, are robust and offer possibilities to validate the model using statistical techniques [10]. Most important, rules can be read from the resulting trees. Determining a tree which can be used to predict a set of junction design alternatives is basically a multi-labelled decision

tree challenge [11]. A modest number of algorithms have been suggested to construct multi-labelled decision trees [12, 13, 14, 15, 16, 17]. In these algorithms, various functions in the traditional decision tree algorithms are replaced by functions fit for handling multi-labelled data, primarily based on measures for the similarity between one or more sets. Unfortunately, these algorithms have some serious disadvantages, the most important one being the fact that they produce very large and complex trees, which is caused by their lack of pruning methods. To overcome these difficulties, we introduced an alternative approach [1]. In this approach we normalise the training data and use the predicted probabilities of the resulting tree, confronted with a threshold value, to determine multiple target labels. This enables us to predict sets of junction design alternatives with traditional algorithms and thus having the advantage of using profoundly proven and widely available methods with a range of modelling options, such as pruning. This approach consists of three (sub)steps:

- Normalise the data

- Built a decision tree

- Determine predicted sets of solutions

2.3.1 Normalise the data

The first step is to apply the standard data normalisation method [18], which is needed when using single labelled decision tree algorithms. This method transforms the data into single-labelled instances. The example training set from Table 2, would then be transformed to the data set with fourteen instances. One instance for each set element. An instance with a set of three junction design alternatives (id 1 in Table 2) is transformed to three instances with one junction design alternative.

2.3.2 Built a decision tree

In the second step, traditional, single-labelled, decision tree algorithms such as ID3, C4.5, CRT, CHAID and QUEST can be used to build a tree. Most of these decision tree algorithms consist of two conceptual phases: growing and pruning. In the tree growing phase, a tree is constructed by using an iterative procedure, were in each iteration the algorithm considers the partition of the training set by selecting the most appropriate attribute according to some splitting measure(s). The tree growing phase continues until a certain stopping criterion is triggered, e.g. when all instances in nodes belong to one class or when a maximum tree depth has been reached. In the pruning phase, the size of the tree is reduced by removing sections of the tree that provide little power to classify the instances. Pruning reduces the complexity of the final tree and hence improves predictive accuracy by reduction of over-fitting.

A major point of criticism for single-labelled decision tree algorithms, is that the obtained classifier could only predict a single label [13]. However, this is only partly true. Although the decision tree algorithms aim to produce nodes and leaves which contain elements with the same label, leaves more often contain elements with multiple labels. This is particularly true for situations where all the instances in a node have (nearly)the same attributes, as a result of which no attribute and value can be determined to split the instances any further. In our example, the normalised instances coming from one set have the same attributes. When there is a sufficient amount of instances in a leaf, i.e. small tree and large dataset, a leaf contains a probability vector, were the probability is determined by comparing the frequency of the instances with a specific label with the total number of instances in the leaf. This probability vector in combination with the normalisation method, offers the opportunity to use the single labelled methods for multiple labelled data.

Total volume

≥ 1000

k	n	p
AW	0	0.00
TW	0	0.00
SIG	2	0.44
RA	2	0.44
OTHER	1	0.12
Total	4	1.00

{SIG,RA,OTHER}

< 1000

Size category

1   2

k	n	p
AW	2	0.44
TW	2	0.44
SIG	1	0.12
OTHER	0	0.00
Total	0	1.00

{AW,TW,SIG}

k	n	p
AW	0	0.00
TW	2	0.50
SIG	0	0.00
RA	2	0.50
OTHER	0	0.00
Total	4	1.00

{TW,RA}
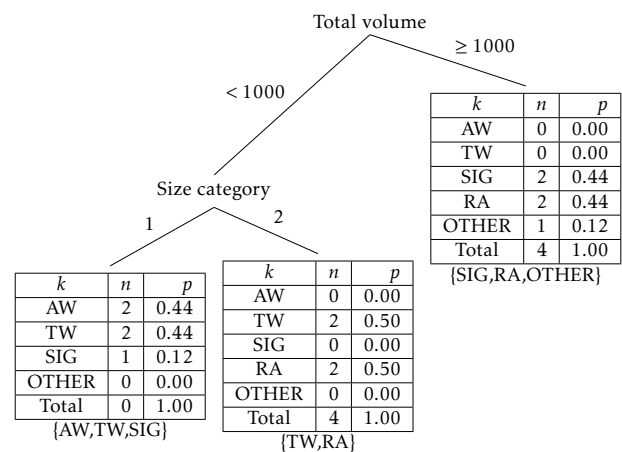
Figure 3: Example decision tree with predicted probabilities

Figure 3 shows the decision tree based on our example training set, but now including the number of instances (count and percentage) for each junction design alternative (k) in each leaf. The percentage represents the predicted probability. Based on preliminary tests we concluded that the CRT decision tree algorithm performed best for our problem. CRT [19] performs binary splits based on a Gini impurity measure and allows pruning.

2.3.3 Determine predicted set of solutions

In the third step, the predicted set of junction design alternatives can be derived from the frequencies and/or predicted probabilities. In our case, a junction design alternative is included in the predicted set if the predicted probability in the leaf is greater than zero. For our example, the predicted set of junction design alternatives is shown at the bottom of each leaf, as can be seen in Figure 3. Alternative threshold values (τ) can be used to determine the predicted sets to reduce the number of junction design alternatives in the sets. Higher values of τ will produce smaller predicted sets of solutions, but with a higher change of missing

junction design alternatives. Lower values of τ will produce larger predicted sets of solutions, but with a higher change of overestimating the number of viable junction design alternatives. The optimal value for τ is to be determined, but can be chosen after the decision tree has been built. Moreover, this variable can either be determined static (1) or dynamic (2):

$$\tau_{static} = a \tag{1}$$

$$\tau_{dynamic} = \frac{1}{(K + b)} \tag{2}$$

Where K is the number of possibly predicted junction design alternatives in the regarding leaf and a and b are constants. In our case K, is dependent upon the size category, since not each size category has the same number of junction design alternatives. In our small example (Figure 3), size category 1 does not include the roundabout, so K would be 4, as can be seen in the leftmost leaf. For size category 2 K is equal to 5, as can be seen in the central leaf. If the size category is not a split attribute, or multiple size categories are combined in one leaf, the highest number for K is presumed, as is the case for the rightmost leaf. Based on preliminary tests we concluded that a dynamic determination of τ with $d = 1$ gave best results.

3 Evaluation framework

We will apply the approach on two cases. The cases will be explained in section 4. In order to determine the success of the approach for the two cases, the new rules resulting from the approach are compared with existing rules. First the predictive accuracy of the rules will be evaluated. This will give insight in how good the rules can reproduce the Pareto optimal set of solutions from the generated dataset as described in section 2.2. Secondly, the performances of the predicted sets of viable junction design alternatives will be evaluated. This will give insight in the average and/or minimal (e.g. operational, safety and environmental) performances of the junction design alternatives that are predicted by the rules. Two-third of the dataset, as described in section 2.2, is used to train the model for the new rules. One-third is used to evaluate both the new and existing rules. The latter is the test set.

3.1 Accuracy

In situations with single-labelled data, the predictive accuracy of a rule is simply the proportion of correctly classified or predicted instances. If we would apply this to our multi-labelled data, this is a very strict measure, being the proportion of instances for which the predicted set of labels is the same as the true set of labels. In our context, predicting a set of viable junction design alternatives for further research, we are also interested in the proportion of instances for which the predicted set at least contains the true set of labels. In this context predicting too much labels instead of

predicting too few is less bad. Therefore, we introduce using multiple measures to evaluate the accuracy of the rules. All accuracy measures discussed are on a per instance basis and the aggregate value is an average over all instances. Let T be the true set of labels (i.e. the Pareto optimal set) and P be the predicted set of labels (i.e. by either the new or existing rules). The measure for the predicted set at least containing the true set, we name sufficiency, which is defined as:

$$Sufficiency = \begin{cases} 0, & \text{if } P \not\subseteq T; \\ 1, & \text{if } P \subseteq T. \end{cases} \tag{3}$$

The measure for the predicted set being equal to the true set, we name equality, which is defined as:

$$Equality = \begin{cases} 0, & \text{if } P \neq T; \\ 1, & \text{if } P = T. \end{cases} \tag{4}$$

Although predicting to much labels is less bad than predicting too few, it is still a disadvantage of the model. Therefore, we also evaluate the number of labels that is wrongfully predicted with the overestimation measure:

$$Overestimation = |P/T| \tag{5}$$

This measure is best evaluated in relation with the average set size.

Additionally, we use a similarity measure which is a more general measure for comparing sets used by various authors [13, 14, 17], also referred to as the Jaccard index or intersection over union:

$$Similarity = \frac{|T \cap P|}{|T \cup P|} \tag{6}$$

The values for sufficiency, equality and similarity range between 0 and 1. Higher values correspond to higher accuracy. The value of overestimation ranges from 0 up to the maximum number of labels. Lower values correspond to higher accuracy.

In order to determine the strengths and weaknesses of the new rules, it is imperative to differentiate the indicators for different size and volume categories.

3.2 Performance

An important issue involves the fact that existing rules are often based on one objective, thereby neglecting junction design alternatives that are important for other objectives. If a rule selects viable junction design alternatives based on minimizing the average delay, the alternative with the lowest number of accidents could be missing from the predicted set. It is important to compare the performances resulting from the new and existing rules in relation to the performances resulting from the Pareto optimal set. Various indicators exist for this purpose. [20] provided an excellent overview in the context of comparing two Pareto sets. Here we use the minimum objective value attained by a set for every objective function. This is a straightforward indicator which uncovers the problem mentioned above. This is done by determining the performances for each

junction design alternative in the predicted set. The performances for each objective are determined by using the same junction performance models used to generate the dataset, as described in section 2.2. For each objective, an average minimum performance $\overline{z}_{min}$ is determined by using:

$$\overline{z}_{min} = \frac{\sum_{i=1}^{n} \min\{z_{i1}, z_{i2}, \ldots, z_{ik}\}}{n} \qquad (7)$$

where $z_i k$ is the performance of a junction design alternative from the predicted set, given a combination of traffic demand pattern and size category i and n is the total number of these combinations to be evaluated. Using the rules, each combination i produces a predicted set of junction design alternatives. Each of these alternatives has a specific performance for the regarding objective. For each combination of i the minimum performance is used. These values are summed for all combinations of i and divided by the total number of combinations n. This is repeated for each objective and for both the Pareto optimal sets and the sets based on the existing and new rules.

A minor complication is the fact that sets can also contain the 'OTHER' junction design alternative. As explained in section 2.2, for this combination of traffic demand pattern and size category, there was no junction design alternative with a delay below the chosen threshold value. Evidently, since the junction design is unknown, no performances can be determined for this alternative. Therefore, if a set contains an 'OTHER' alternative, the regarding combination of traffic demand pattern and size category i is excluded from equation 7.

4   Cases

Junction design rules can be found in design manuals, which are usually issued by government agencies on a provincial, state or national level. For example, junction design rules are provided in design manuals for the following countries: Australia [21], Belgium [22], Germany [23, 24, 25], France [26], The Netherlands [27, 28, 29], UK [30] and USA [31, 32, 7, 2, 5, 33]. The availability and content of the rules is country specific. This is caused by the fact that policies, procedures, junction design alternatives, traffic modes, traffic behaviour and accepted modelling tools are very different by country. For example, in some countries turbo-roundabouts are generally accepted, whereas in other countries they are not (yet) considered as a viable alternative. Another example is that the same roundabout design can have a different operational and safety performance in different countries. It is not useful to determine universal junction design rules, but rather determine them for a specific region or country.

We chose two case countries, being the United States of America (USA) and the Netherlands, mainly due to the availability of documented junction design rules and corresponding performance models. For both cases, first a short description of the existing rules

is given, followed by a brief explanation of the existing rules that haven been chosen for comparison and the application of our approach for the specific case.

4.1   Case 1: United States of America

In this section we will explain the application of our approach as part of a junction design assessment process in the USA.

4.1.1   Existing rules

In the USA, junction design rules can be found in both national and state documents. Here, we will focus on the national level. Junction design rules are scattered over various junction design manuals and technical reports, published by TRB [2, 33], FHWA [34, 32] and AASHTO [31, 35]. A lot of rules require data generated by performance models and lie outside our research scope which aims to determine rules for the first stage of the junction design assessment process. The Manual on Uniform Traffic Control Devices (MUTD) [32] describes eight warrants that define conditions in which a traffic control signal is likely to improve junction safety, operations, or both. If a warrant is met, then signal control may be appropriate. Some warrants require volume based input data, such as eight, four or one hour volumes, while others require performance based data such as delay or the crash history or specific local data such as the area population. Various rules exist for the realisation of specific junction design alternatives and elements, such as volume thresholds for roundabouts or for the number of entry lanes required, e.g. [33]. An overall approach is also provided, combining the earlier mentioned warrants and various other rules in one method [2]. Not all warrants and rules are relevant and/or usable in order to determine a viable set of junction design alternatives. The rules had to be adapted to be used in this research, in order to make a fair comparison between existing and new rules. An extensive description of the existing rules that we used can be found in the Appendix.

4.1.2   Application of the approach

New junction design rules will be determined by using the approach as described in section 2. An important first step is to define the scope, where choices have to be made concerning the junction design alternatives, the traffic flow ranges, the objectives and the performance models to be considered. Table 3 shows the scope definition for this case.

The junction design attributes have been chosen to match the junction design types for the existing rules. Four main types are differentiated. The alternatives are based on the design alternatives mentioned in the underlying existing rules. When there are multiple approach lanes, different lane configurations are possible for the major and minor roads. Multiple variants are then differentiated, leading to 67 separate alternatives to be tested. Seven size categories are differentiated,

which is necessary for the approach to avoid that only the largest junction design types are part of the Pareto optimal set and thus the rules to be determined. An overview of the junction design alternatives can be found in 10 in the Appendix.

Table 3: Scope definition for case 1: USA

Topic	Item	Choices
Junction design	Arms	4
	Main types	AWSC, TWSC, signalised, roundabout
	Alternatives	21
	Variants	67
	Size categories	7
	Attributes	Junction: Main type; Arm: Approach lane configuration, central reservation width, number of exit lanes, sign type, circulating lanes
Traffic flow	Units	Peak hour volumes (pcu/h)
	Modes	Motorised vehicles
	Flows	12 turning flows for motorized vehicles
	Ranges	Total volume: 0-7000 (pcu/h),
Objectives	Objectives	Operation, safety
	Measures	Volume weighted average delay (s), number of fatal-and-injury crashes per year
	Thresholds	Average delay ≤ 50 (s)
Models	Operation	HCM 2010 methodologies
	Safety	HSM 2010 methodologies

Random traffic flow patterns are generated based on a total motorised volume ranging between 1 and 7000 (pcu/h). Two objectives and corresponding performance measures have been defined. Junction design alternatives with an average delay above 50 seconds are excluded from the choice sets.

The models used to determine the performances are static analytical based. The operational performance model is an implementation of the Highway Capacity Manual 2010 (HCM2010) methodologies [5]. For signalised junctions control settings (cycle time, green times) are determined (based on the traffic flow and junction design) which aim to minimise the average delay on the junction. The safety performance model is an implementation of the Highway Safety Manual 2010 (HSM2010) methodologies [35].

4.2 Case 2: The Netherlands

In this section we will explain the application of our approach as part of a junction design assessment process in the Netherlands.

4.2.1 Existing rules

In the Netherlands, junction design rules can be found in various publications of the CROW [36, 27, 28, 29]. Most rules are capacity thresholds, which means that they provide a value for the maximum allowed traffic volumes for a certain time interval, e.g. the total traffic volume on a junction or arm in a peak hour (pcu/h). Various tables are provided for various junction design types. Another, rather general but common rule, prescribes a preferred order in which main junction design types should be considered [28, 29]. The preferred order, being; roundabout, priority junction and signalised junction is based on the sustainable safety principle, in which roundabouts are considered (and

proven to be) the safest junction design types. The rule dictates that a roundabout should be considered, unless it should be excluded for operational reasons. Furthermore, various design rules exist for specific design elements and/or junction design types. For example, a graphically represented rule which can be used to determine whether a single lane roundabout should be considered for a given combination of entry flow, conflicting circulating flow and crossing bicycle flow [36].

There is no overall (set of) junction design rule(s) available applicable for all relevant junction design alternatives and traffic modes. Therefore, we use a combination of rules from two source. The first source [27] provides two capacity thresholds for different types of junctions, ranging from single lane roundabouts, two lane roundabouts to turbo-roundabouts, priority junctions and signalised junctions. The first threshold represents the maximum allowed sum of all entry flows in a peak hour. The second threshold represents the sum of the entry flow and the (potentially) conflicting flow for one arm and is thus evaluated for each arm. If one of the threshold values is exceeded, then this junction design type is considered unsuitable to be taken into account for further research. The second source [29] provides the so-called Slop-criterion which calculates a value based on the (eight-hour) total volume on the major road, the (eight-hour) maximum volume of the minor roads and a set of parameters. If the resulting value exceeds a certain threshold value, a priority junction is no longer a viable solution. Using the (capacity) thresholds does not provide a set of viable junction design alternatives. However, the rule can be used to determine such a set. An extensive description of the existing rules that we used can be found in the Appendix.

Since bicycle flows are a substantial part of traffic flow in the Netherlands, it is desirable to reckon with these in the rules. However, since there are no existing rules for all junction design types, crossing bicycle flows will be reckoned with in the existing rules by simply adding them to the conflicting flow (1 bicycle = 0.2 pcu), which is used in various rules.

4.2.2 Application of the approach

New junction design rules will be determined by using the approach as described in section 2. An important first step is to define the scope, where choices have to be made concerning the junction design alternatives, the traffic flow ranges and the objectives to be considered. Table 4 shows the scope definition for this case.

The junction design attributes have been chosen to match the junction design types for the existing rules. Four main types and 16 junction designs are differentiated. When there are multiple approach lanes, different lane configurations are possible for the major and minor roads. Multiple variants are then differentiated, leading to 42 separate alternatives to be tested. Five size categories are differentiated, which is neces-

sary for the approach to avoid that only the largest junction design types are part of the Pareto optimal set and thus the rules to be determined. An overview of all junction design alternatives and variants can be found in the Appendix in Table 17.

Table 4: Scope definition for case 2: The Netherlands

Topic	Attribute	Choices
Junction design	Arms	4
	Main types	Equal, priority, signalised, roundabout
	Alternatives	16
	Alternatives	42
	Size categories	5
	Attributes	Junction: Main type; Arm: Approach lane configuration, central reservation width, number of exit lanes, sign type, number of opposed circulating lanes
Traffic flow	Units	Peak hour volumes (pcu/h)
	Modes	Motorised vehicles, bicycles
	Flows	12 turning flows for motorised vehicles, 4 crossing flows bicycles
	Ranges	Total motorised volume: 1-7000 (pcu/h), total bicycle volume 0-500 (bicycles/h)
Objectives	Objectives	Operation, safety, environment
	Measures	Volume weighted average delay (s), number of fatal-and-injury crashes per year, NO_x/PM_{10} emission (g)
	Thresholds	Average delay ≤ 50 (s)
Models	Operation Safety Environment	Local static analytical models

Random traffic flow patterns are generated based on a total motorised volume ranging between 1 and 7000 (pcu/h) and a total bicycle volume ranging between 0 and 500 (bicycles/h). Three objectives and corresponding performance measures have been defined. Junction design alternatives with an average delay above 70 seconds are excluded from the choice sets.

The models used to determine the performances are static analytical based and were specially developed and tested for use in the Dutch practice [37]. The operational performance model is similar to the methodologies provided in the American HCM2010 [5] and the German HBS2015 [25], extended with new methods for (turbo)roundabouts and the effects of bicycle traffic. The safety performance model is similar to the American HSM2010 [35] methodologies, where parameters were adopted from [38] for local situations. The environmental performance model determines the emissions based on local emission substance factors multiplied by the queue on each lane [39, 3].

5   Results

In this section we will discuss the results. We will first consider the predictive accuracy of the existing and newly determined junction design rules. Next we will evaluate the performances of the predicted sets of viable junction designs. All results were determined with the test dataset.

5.1   Accuracy

Table 5 shows accuracy values for both the existing and new rules for both cases. The accuracy represents how good the rules can predict the Pareto optimal sets. The table shows values for sufficiency, similarity, equality and overestimation, according to equations 3-5. In addition to the accuracy values, the average set size is presented.

Table 5: Accuracy measures of existing and new rules for both cases

Item	Case 1		Case 2	
	Existing	New	Existing	New
Sufficiency	0.311	0.905	0.845	0.899
Similarity	0.249	0.723	0.391	0.821
Equality	0.129	0.492	0.051	0.745
Overestimation	1.320	0.550	1.596	0.293
Average set size	1.857	1.741	2.606	1.242

The table shows that the new rules can better predict the Pareto optimal sets then the existing rules can. For both cases, the sufficiency values, being the measure for the predicted set at least containing the true (Pareto optimal) set, for the new rules are higher. For case 1 the difference is evident. The existing rules have a sufficiency rate of 31.1% where the new rules accomplish a 90.5%. For case 2, the difference in sufficiency rates is limited, with respectively 84.5% and 89.9% for the existing and new rules. However, there are big differences for the other measures. This is caused by the fact that the existing rules predict large sets, with an average set size of 2.606 and an average overestimation of 1.596. Larger sets have a higher change to at least contain the true set, which is reflected in the sufficiency value. However, equality, which reflects the rate at which the predicted set being equal to the true set, is extremely low. The rules are not discriminating enough, thus causing substantially more work in a later stage of the junction design assessment process. The new rules produce more equal and smaller sets.

The presented accuracy values in Table 5 are average values for the whole test set. In order to determine whether differences exist for different size or volume categories, differentiated results are presented in Table 6 and Table 7.

Table 6: Accuracy measures for the new rules differentiated by size category for both cases

Case	Size category	Sufficiency	Similarity	Equality	Overestimation
1	1	0.951	0.826	0.674	0.404
	2	0.899	0.860	0.807	0.182
	3	0.882	0.807	0.586	0.375
	4	0.882	0.747	0.594	0.604
	5	0.873	0.583	0.234	0.805
	6	0.933	0.578	0.216	0.840
	7	0.913	0.660	0.335	0.642
2	1	0.954	0.902	0.851	0.153
	2	0.949	0.844	0.723	0.285
	3	0.894	0.857	0.802	0.179
	4	0.870	0.760	0.678	0.437
	5	0.830	0.743	0.672	0.410

In Table 6, which shows the accuracy measures differentiated by size categories, it can be seen that sufficiency rates are fairly stable for all size categories. The values range between 87.3% and 95.1% for case 1 and between 83.0% and 95.4% for case 2. For case

2, equality rates are also relatively high. For case 1, these rates are lower for the higher size categories. This means that the new rules predict bigger sets, which also can be seen while looking at the overestimation values. This is not a big problem, since the overestimation is still limited and moreover, the similarity values, which represent an average set comparison, are still reasonably high.

In Table 7, which shows the accuracy measures differentiated by volume categories, a similar stable set of sufficiency rates can be seen for both cases.

Table 7: Accuracy measures for the new rules differentiated by volume category for both cases

Case	Volume category (pcu/h)	Sufficiency	Similarity	Equality	Overestimation
1	<1000	0.961	0.791	0.506	0.527
	1000-1999	0.914	0.741	0.484	0.605
	2000-2999	0.711	0.557	0.331	0.797
	3000-3999	0.851	0.681	0.477	0.555
	4000-4999	0.934	0.743	0.533	0.484
	5000-5999	0.970	0.767	0.551	0.451
	≥6000	0.984	0.773	0.557	0.443
2	<1000	0.931	0.826	0.706	0.292
	1000-1999	0.859	0.826	0.768	0.226
	2000-2999	0.857	0.773	0.666	0.315
	3000-3999	0.931	0.668	0.450	0.743
	4000-4999	0.865	0.798	0.763	0.338
	5000-5999	0.906	0.906	0.906	0.094
	≥6000	0.943	0.943	0.943	0.057

Sufficiency rates are best for the lower and higher volume categories. The middle categories generally have the lowest sufficiency values. This can in part be explained by the fact that there are more junction design alternatives available for these volume categories. Moreover, as was the case for the size categories, the similarity values are again still reasonably high.

Generally, it can be concluded that the new rules have a much better accuracy then the existing rules and that there are no specific weaknesses concerning the accuracy values if they are differentiated for size and volume categories. The Pareto optimal sets can be predicted with a 90% accuracy.

5.2 Performance

Table 8 shows the average minimum performances for all objectives as has been defined in equation 7. The operational performance is expressed as volume weighted average delay (s), the safety performance as the number of fatal-and-injury crashes by year, and the environmental performance as the PM_{10} emissions (g). The table shows the average minimum performances for respectively the Pareto optimal sets, the sets predicted by the existing rules and the sets predicted by the new rules. For a fair comparison, if any of the sets (Pareto, existing, new) only contains the 'OTHER' junction design alternative, the regarding instance is excluded from the analysis, as was mentioned in section 3.2.

Table 8: Performances measures of existing and new rules for both cases

$\overline{z}_{min}$	Case 1			Case 2		
	Pareto	Existing	New	Pareto	Existing	New
Operation	16.415	24.476	18.588	15.579	23.995	31.367
Safety	0.787	0.782	0.776	0.581	0.577	0.574
Environment				0.923	1.869	1.028

For case 1, the new rules show that the average minimum performance for both objectives is better (lower) in comparison with the existing rules. This means that the new rules produce sets of junction design alternatives which contain alternatives with a better (lower) performance for either operation, safety or a combination of one or more of these objectives. This was to be expected, since the sufficiency rate for the existing rules was only 31.1% for this case. On the other hand, the differences are not as big as those for the accuracy rates. The new rules provide an average reduction of 24% for operational and only 1% for safety performances in comparison with the existing rules. This is partly caused by the exclusion of instances for which one of the sets predicts a single 'OTHER' junction design alternative. This was the case for 58% of the instances. A striking point is the fact that the average minimum safety performance for both the existing and the new rules is lower than that for the Pareto optimal set. This seems strange, since an alternative that has the minimum (optimal) performance for one objective is to be expected to be in the Pareto optimal set. This was true, however since we used a threshold value for the operational performance, some alternatives were excluded from the eventual set. When the new rules 'accidentally' predicted these alternatives, the performance measures for these alternatives were yet reckoned with.

For case 2, the average minimum operational performance is better for the existing rules than for the new rules. This is primarily caused by the fact that the existing rules predict large sets, as was mentioned in the previous section, thereby increasing the change that the junction design alternative with the minimum performance is part of the predicted set. Moreover, the existing rules are specifically made to serve the operational objective. On the other hand, the average minimum performance for the new rules is rather high. Further analysis shows that there is a relatively small amount of wrongly predicted junction design alternatives which have very high values for the operational performance. Table 9 shows the number of instances for various delay classes.

Table 9: Number of instances by delay class for case 2

Delay class (s)	Pareto	Existing	New
< 25	5389	5224	5227
≥ 25 and < 50	1157	1183	1185
≥ 50 and < 70	459	498	480
≥ 70	0	70	113

When using median values instead of average, the

operational performances for the Pareto, existing and new sets are respectively 7.145, 7.175 and 7.156. The performance values for the safety and environmental performances are better (lower) for the new rules. The new rules provide an average reduction of 1% for the safety performance and 45% for the environmental performance.

Generally, it can be concluded that the new rules predict junction design alternatives with either equal or better (lower) performances for all objectives.

6 Conclusions

In this study, we demonstrated the application of a new approach to determine junction design rules for use in the first stage of the junction design assessment process. The approach uses decision tree algorithms to determine decision trees based on modelled data. There is a need for quick design rules which need limited input data. Although some of these rules exist, their usability is limited due to inconsistencies in rules and non-transparency in combination with objectives. In this paper, we present an approach by which new and better junction design rules can be determined. The resulting rules can be used to predict a set of viable junction design alternatives for the first stage of the junction design assessment process. Moreover, the predicted set is in fact the Pareto optimal set of solutions for multiple objectives, e.g. regarding operational, safety and/or environmental impact. The Pareto optimal set of solutions always contains the best solution, whatever set of weights is used for different objectives in a later stage of the assessment process, thus handling multiple objectives in a straightforward manner.

The approach was applied and evaluated on cases in two different countries. Results show that for about 90% of the situations the Pareto optimal set can be predicted by the new rules, whereas existing rules hardly reach 35% or are not discriminating enough resulting in large set sizes, creating more work in later stages of the junction design assessment process. The new rules provide better performances for the non-operational objectives (safety and environment). Results for the operational performances were different for the two cases. Generally, the new rules produce better results with smaller predicted sets.

The major contribution of this paper is that it presents an approach to determine consistent and complete junction design rules based on modelled data in a transparent and systematic manner. Although some efforts have been made to determine junction design rules based on modelled data no generic and transferable method had been developed until now.

Appendix

In this appendix we provide a description of the junction design alternatives and existing junction design rules for respectively case 1 and 2. At the end of the appendix, an explanation of the variable notations is provided.

Case 1: United States

Table 10 shows the junction design alternatives used for this case.

Table 10: Junction design alternatives for case 1

| Nr | Id | Sc | Approach lane number and configuration(s) | |
			Major road	Minor road
1	A11	1	1 ⇼	1 ⇼
2	T11	1	1 ⇼	1 ⇼
3	S11	1	1 ⇼	1 ⇼
4	T21	2	2 ⇑,⇙,⇓	1 ⇼
5	S21	2	2 ⇑,⇙,⇓	1 ⇼
6	T22	3	2 ⇑,⇙,⇓	2 ⇑,⇙,⇓
7	T31	3	3 ⇑	1 ⇼
8	S22	3	2 ⇑,⇙,⇓	⇑,⇙,⇓
9	S31	3	3 ⇑	1 ⇼
10	T32	4	3 ⇑	2 ⇑,⇙,⇓
11	S32	4	3 ⇑	2 ⇑,⇙,⇓
12	S41	4	3 ⇑,⇑,⇑	1 ⇼
13	1R11	4	1 ⇼	1 ⇼
14	S33	5	3 ⇑	3 ⇑
15	S42	5	4 ⇑,⇑,⇑	2 ⇑,⇙,⇓
16	S43	6	4 ⇑,⇑,⇑	3 ⇑
17	S44	6	4 ⇑,⇑,⇑	4 ⇑,⇑,⇑
18	2R11	6	1 ⇼	1 ⇼
19	S64	7	6 ⇑	4 ⇑,⇑,⇑
20	2R21	7	2 ⇼	1 ⇼
21	2R22	7	2 ⇼	2 ⇼

A=AWSC-junction,T=TWSC-junction,S=signalised junction,
1R=one-lane roundabout,2R=two-lane roundabout

There are 21 junction design alternatives. Each alternative can be identified by the main type and the number of approach lanes on the major and minor road arms. Junction design alternative number 4, with id 'S21' is a signalised junction with two approach lanes on the major road arms and one approach lane on the minor road arms. A description of the characters uses for the junction types is provided at the bottom of the table. The number of exit lanes is automatically determined, based on the number of approach lanes with a destination to the regarding arm. When there are multiple approach lanes, different lane configurations are possible. Multiple variants are then differentiated, e.g. three (3 times 1) variants for alternative number 4 and nine (3 times 3) variants for alternative number 15. In total 67 separate variants are defined. The junction design alternatives are grouped in seven size categories (Sc), based on the required space. The rules are based on the guidelines provided in [2]. We only use guidelines that refer to a specific junction design alternative. Rules referring to more detailed design elements, such as on-street parking, left-

turn prohibition, lane length, right-turn radius and signal flash mode, are excluded. Obviously, we only use guidelines that need volume-based input. Guidelines that require performance-based input, such as delay or crash-history, are excluded, since we don't want to run a model or perform an extensive survey in the first stage of the junction assessment process. The guidelines regarding signalised junctions in [2] are based on the so-called traffic signal warrants from [40]. These warrants have been updated over the years and therefore we use the most recent version from [32]. Most guidelines regarding stop-controlled junctions are presented as graphs. For these, different (linear, exponential, logarithmic, second-order polynomial) functions have been estimated to fit the lines in the graphs.

Table 11 shows which rules are applied for each size category (Sc). Each size category has a base set of viable junction design alternatives, corresponding to the definitions in Table 10. For each alternative in the base set it is evaluated whether it should be included (I) or excluded (E) from the base set. For each main type (A=AWSC-junction, T=TWSC-junction, S=signalised junction and R=roundabout), this is determined by one ore more rules. Each rule performs a check resulting in a true or false value. If true then that particular type is either included or excluded from the base set. Table 11 shows which rules cause inclusion or exclusion. As a starting point all alternatives are excluded. The rules are executed in the order as shown in Table 11. For example, for size category (Sc) 1, 'S11' is excluded unless one of the rules S1-S4 provides a true value. Alternative 'T11' is excluded unless rule T1 is true, but if one of the rules T2-T4 is true it is always excluded. Table 12 provides the conditions that are checked for each rule.

Table 11: Junction design rules for case 1: Application of rules by size category

Sc	Base set	Rules											
		S1	S2	S3	S4	A1	A2	T1	T2	T3	T4	T5	R1
1	{A11,T11,S11}	I	I	I	I	I	E	I	E	E	E		
2	{T21,S21}	I	I	I	I			I	E			E	
3	{T31,S31,T22,S22}	I	I	I	I			I	E			E	
4	{T32,S32,S41,1R11}	I	I	I	I			I					I
5	{S33,S42}	I	I	I	I								
6	{S43,S44,2R11}	I	I	I	I								I
7	{S64,2R21,2R22}	I	I	I	I								I

Table 12: Junction design rules for case 1: Rules and conditions

Code	Condition	Specifics
S1	$q_{mas8} \geq \lambda_{S1a}$ and $q_{mix8} \geq \lambda_{S1b}$	Signal warrant 1a
S2	$q_{mas8} \geq \lambda_{S2a}$ and $q_{mix8} \geq \lambda_{S2b}$	Signal warrant 1b
S3	$q_{mas4} \geq 0.0003 q_{mix4}^2 - \lambda_{S3a} q_{mix4} + \lambda_{S3b}$	Signal warrant 2
S4	$q_{mas1} \geq 0.0002 q_{mix1}^2 - \lambda_{S4a} q_{mix1} + \lambda_{S4b}$	Signal warrant 3
A1	$q_{mas8} \geq 300$	
A2	$J = signal$	
T1	$q_{mas24} \geq 6000$	
T2	$q_{mil} > \lambda_{T2a} e^{(-0.001 q_{mas1})}$	≥ 2 lanes on minor road
T3	$q_{mao1} > \lambda_{T3a} \ln(q_{ma1}) + \lambda_{T3b}$	left-turn lane on major road
T4	$q_{mar1} > 683.6 e^{(-0.004 q_{mas1})}$	right-turn lane on major road
T5	T3 and T4	
R1	$q_{tot24} \geq 3600 + 9000 l_c (1 + (81/50)) - 94 p_l$	

Tables 13, 14 and 15 provide parameter values for the λ that is used in various rules. Table 16 provides

the volume conversion factors that were used to determine the volumes needed for the rules, based on the peak hour volume flows (pcu/h) that are used as base input. The variable notations are explained at the end of the appendix.

Table 13: Junction design rules for case 1: Parameters for rules S1-S4

l_{ma}	l_{mi}	λ_{S1a}	λ_{S1b}	λ_{S2a}	λ_{S2b}	λ_{S3a}	λ_{S3b}	λ_{S4a}	λ_{S4b}
1	1	500	150	750	75	0.7697	573.8	0.7815	751.44
≥ 2	1	600	150	900	75	0.7689	659.11	0.7339	842.04
≥ 2	≥ 2	600	200	900	100	1.0226	881.44	0.9397	1076.4

Table 14: Junction design rules for case 1: Parameters for rule T2

p_{mir}	λ_{T2a}
≤ 0.35	550.67
> 0.35	641.80

Table 15: Junction design rules for case 1: Parameters for rule T3

p_{mal}	λ_{T3a}	λ_{T3b}
< 0.075	-922.4	6157.0
$0.075 - 0.124$	-951.8	6046.6
$0.125 - 0.174$	-863.7	5369.8
$0.175 - 0.299$	-997.9	5849.3
≥ 0.300	-880.7	5362.9

Table 16: Volume conversion factors for the junction design rules of case 1

Volume	Units	Factor
Peak hour volume	pcu/h	1.00
Peak hour volume	veh/h	0.93
Four-hour volume	veh/h	0.78
Eight-hour volume	veh/h	0.67
Daily volume	veh/day	10.33

Case 2: The Netherlands

Table 17 shows the junction design alternatives used for this case. There are 16 junction design alternatives. As for case 1, each alternative can be identified by the main type and the number of approach lanes on the major and minor road arms. The characters referring to the main junction types are slightly different compared to case 1, and are provided at the bottom of the table. In total 42 separate variants are defined. The junction design alternatives are grouped in five size categories (Sc), based on the required space.

The rules are based on guidelines provided in [27] and [29]. There are three main types of rules. The first rule uses a threshold value for the total amount of traffic on the junction. Different threshold values are used for different junction types. The second rule checks a threshold value regarding the conflicting volume for each arm. If the threshold is exceeded for one

of the arms, a specific type is not viable. The third rule is the so-called Slop-criterion, which is used to check whether a priority junction is still viable. Table 18 shows which rules are applied for each size category (Sc).

Table 17: Junction design alternatives for case 2

Nr	Id	Sc	Approach lane number and configuration(s)	
			Major road	Minor road
1	E11	1	1 ⇧	1 ⇧
2	P11	1	1 ⇧	1 ⇧
3	S11	1	1 ⇧	1 ⇧
4	P21	2	2 ⇧,⇦⇧	1 ⇧
5	P22	2	2 ⇧,⇦⇧	1 2 ⇧,⇦⇧
6	S21	2	2 ⇧,⇦⇧	1 ⇧
7	S22	2	2 ⇧,⇦⇧	2 ⇧,⇦⇧
8	S33	3	3 ⇧,⇧,⇦⇧,⇧,⇦⇧	3 ⇧,⇦⇧
9	1R11	3	1 ⇧	1 ⇧
10	2R11	3	1 ⇧	1 ⇧
11	S44	4	4 ⇧,⇧,⇧,⇧	4 ⇧,⇧,⇧
12	TR22	4	2 ⇦⇧,⇧	2 ⇦⇧
13	2R22	4	2 ⇦⇧	2 ⇦⇧
14	S66	5	6 ⇧⇧⇧⇧	6 ⇧⇧⇧⇧
15	TR23	5	2 ⇦⇧	3 ⇧⇦⇧
16	TR33	5	3 ⇧⇦⇧	3 ⇦⇧

E=equal junction,P=priority junction,S=signalised junction,
1R=one-lane roundabout,2R=two-lane roundabout,TR=turboroundabout

Table 18: Junction design rules for case 2: Application of rules by size category

Sc	Base set	Rules									
		E1	E2	E3	P1	P2	P3	S1	S2	R1	R2
1	{E11,P11,S11}	E	E	E	E	E	E	E	E	E	
2	{P21,P22,S21,S22}					E	E	E	E	E	
3	{S33,1R11,2R11}							E	E	E	E
4	{S44,2R22,TR22}							E	E	E	E
5	{S66,TR23,TR33}							E	E	E	E

In this case, as a starting point all alternatives for a specific size category are included in the final set unless one of the applied rules for that specific alternatives returns a true value. Table 19 provides the conditions that are checked for each rule.

Table 19: Junction design rules for case 2: Rules and conditions

Code	Condition	Specifics
E1	$q_{tot1} > 1500$	total volume threshold
E2	$q_c > 1100$	conflicting volume threshold
E3	$(q_{mix8}/\beta_1) \times (-1 + \sqrt{1 + \beta_2 q_{mas8}/q_{mix8}}) > 1.33$	Slop-criterion
P1	$q_{tot1} > 1500$	total volume threshold
P2	$q_c > 1100$	conflicting volume threshold
P3	$(q_{mix8}/\beta_1) \times (-1 + \sqrt{1 + \beta_2 q_{mas8}/q_{mix8}}) > 1.33$	Slop-criterion
S1	$q_{tot1} > \lambda_{S1}$	total volume threshold
S2	$q_c > \lambda_{S2}$	conflicting volume threshold
R1	$q_{tot1} > \lambda_{R1}$	total volume threshold
R2	$q_c > \lambda_{R2}$	conflicting volume threshold

Table 20: Junction design rules for case 2: Parameters for rules E3,P3

l_{ma}	l_{mi}	β_1	β_2
1	1	300	2.0
2	1	300	2.4
1	2	400	3.2
2	2	400	2.7

Tables 20,21 and 22 provide various parameter values for the rules. In this case only one volume conversion value is used in comparison with the peak hour volume (1.00). In order to obtain the eight-hour volume a factor of 0.63 is used.

Table 21: Junction design rules for case 2: Parameters for rules S1-S2

l_{ma}	λ_{S1}	λ_{S2}
< 4	3500	3800
≥ 4	7500	3800

Table 22: Junction design rules for case 2: Parameters for rules R1-R2

Roundabout alternative	λ_{R1}	λ_{R2}
1R11	2000	1500
2R11	2200	1700
2R22	3500	2400
TR22	3500	2100
TR23	4000	2300
TR33	4500	2800

Notations

Variables:

q_{at}	Vehicular traffic volume for turn(s) a and time unit t
λ_r	Parameter for rule r
p_m	Percentage of traffic volume for movement m
l_c	Number of circulating lanes on a roundabout
l_{ma}	Number of lanes for the major approach
l_{mi}	Number of lanes for the minor approach
J	Junction type

Subscripts (a, t, m):

1	Peak hour volume
4	Four-hour volume
8	Eight-hour volume
24	Daily volume
ma	Major arm approach volume
mao	Major arm opposing volume (right-turn and through)
mar	Major arm right-turn volume
mas	Sum of approach volumes of both major arms
mi	Minor arm approach volume
mir	Minor arm turn-turn volume
mix	Maximum approach volume of both minor arms
tot	Sum of approach volumes of all arms
l	Left-turn movement volume
c	Conflicting volume for an arm

References

[1] E. M. Bezembinder, L. J. J. Wismans, and E. C. van Berkum. Constructing multi-labelled decision trees for junction design using the predicted probabilities. In *Proceedings IEEE 20th International Conference on Intelligent Transportation Systems*, Yokohama, Japan, October16-19 2017.

[2] TRB. Evaluating intersection improvements: An engineering study guide. NCHRP Report 457, Transportation Research Board of the National Academies, Washington D.C., United States, 2001.

[3] E. M. Bezembinder, L. J. J. Wismans, and E. C. van Berkum. Multi-objective assessment framework for intersection alternatives under different demands. In *Proc. Transportation Research Board 95th Annual Meeting*, Washington D.C., United States, January10-14 2016.

[4] L.D. Han, J-M Li, and T. Urbanik. Control-type selection at isolated intersections based on control delay under various demand levels. *Transportation Research Record: Journal of the Transportation Research Board*, 20701:109–116, 2008.

[5] TRB. *Highway Capacity Manual 2010*. Washington D.C., United States, 2010.

[6] B. J. Vitins and K. W. Axhausen. Shape grammars for intersection type choice in road network generation. In *Proc. 12th Swiss Transportation Research Conference*, Monte Verita/Ascona, Switzerland, May2-4 2012.

[7] TRB. *Highway Capacity Manual 2000*. Washington D.C., United States, 2000.

[8] A. Tarko, M. S. Azam, and M. Inerowicz. Operational performance of alternative types of intersections - a systematic comparison for indiana conditions. In *Proc. 4th International Symposium on Highway Geometric Design*, Valencia, Spain, June5-9 2010.

[9] E. M. Bezembinder, L. J. J. Wismans, and E. C. van Berkum. Using decision trees in order to determine intersection design rules. In *Proc. Transportation Research Board 94th Annual Meeting*, Washington D.C., United States, January11-15 2015.

[10] L. Rokach and O. Maimon. *Data mining with decision trees*. World Scientific Publising Co. Pre. Ltd., Singapore, 2008.

[11] S. Kotsiantis. Decision trees: a recent overview. *Artificial Intelligence Review*, 39:261–283, 2013.

[12] A. Clare and R. King. Knowledge discovery in multi-label phenotype data. In *Proceedings of the 5th European Conference on PKDD*, Freiburg, Germany, 2001.

[13] Y.L. Chen, C.L. Hsu, and S.C. Chou. Constructing a multi-valued and multi-labelled decision tree. *Expert Systems with Applications*, 25:199–209, 2003.

[14] S.C. Chou and C.L. Hsu. Mmdt: a multi-valued and multi-labeled decision tree classifier for data mining. *Expert Systems with Applications*, 28:799–812, 2005.

[15] H.Li, R. Zhao, J. Chen, and Y. Chang. *Research on multi-valued and multi-labeled decision trees*, pages 247–254. Springer Berlin Heidelberg, 2006.

[16] W. Yi, M. Lu, and Z. Liu. Multi-valued attribute and multi-labeled data decision tree algorithm. *International Journal of Machine Learning and Cybernetics*, 2:67–74, 2011.

[17] W. Yi, L. Yan, M. Lu, and Z. Liu. A new multi-valued attribute and multi-labeled data decision tree algorithm. *Information*, 16(7A):4771–4780, 2013.

[18] E.F. Codd. *Further normalization of the data base relational model*, volume 6 of *Courant computer science symposium*, pages 33–64. NJ: Prentice-Hall, Englewood Cliffs, 1972.

[19] L. Breiman, J. Friedman, R. Olshen, and C.Stone. *Classification and regression trees*. Wadsworth International Group, Belmont, California, 1984.

[20] T. Brands. *Multi-objective optimisation of multimodal passenger transportation networks*. PhD thesis, University of Twente, Enschede, The Netherlands, 2015.

[21] Austroads. Guide to traffic management part 6: Intersections, interchanges and crossings. Technical report, Austroads, Sydney, Australia, 2013.

[22] AWV. Vadumecum veilige wegen en verkeer. Technical report, Vlaamse Overheid, Agentschap Wegen en Verkeer, Brussels, Belgium, 2009.

[23] FGSV. Richtlinien fr die anlage von stadtstrassen: Rast06. Technical Report 200, Forschungsgesellschaft fr Strassen- und Verkehrswesen, Kln, Germany, 2007.

[24] FGSV. Richtlinien fr die anlage von landstrassen: Ral. Technical Report 201, Forschungsgesellschaft fr Strassen- und Verkehrswesen, Kln, Germany, 2013.

[25] FGSV. Handbuch fr die bemessung von strassenverkehrsanlagen. Technical Report 299, Forschungsgesellschaft fr Strassen- und Verkehrswesen, Kln, Germany, 2015.

[26] SETRA. The design of interurban intersections on major roads. Technical report, Service d'tudes Techniques des Routes et Autoroutes, Bagneux Cedex, France, 2002.

[27] CROW. Turborotondes. Technical Report Publication 257, CROW, Ede, The Netherlands, 2008.

[28] CROW. Asvv 2012. Technical Report Publication 723, CROW, Ede, The Netherlands, 2012.

[29] CROW. Handboek wegontwerp 2013. Technical Report Publications 328-331, CROW, Ede, The Netherlands, 2013.

[30] HE. *Design Manual for Roads and Bridges*. London, England, online version march 2017 edition, 2017.

[31] AASHTO. *A Policy on Geometric Design of Highways and Streets, 6th Edition*. Washington D.C., United States, 2010.

[32] FHWA. Manual on uniform traffic control devices for streets and highways (mutcd). Technical Report 2009 Edition with Revisions 1 and 2, May 2012, Federal Highway Agency of the United States Department of Transportation, Washington D.C., United States, 2012.

[33] TRB. Roundabouts: An informational guide. NCHRP Report 672, Transportation Research Board of the National Academies, Washington D.C., United States, 2010.

[34] FHWA. Alternative intersections/interchanges: Informational report (aiir). Technical Report FHWA-HRT-09-060, FHWA, Washington D.C., United States, 2010.

[35] AASHTO. *Highway Safety Manual*. Washington D.C., United States, 2010.

[36] CROW. Eenheid in rotondes. Technical Report Publication 126, CROW, Ede, The Netherlands, 1998.

[37] E. M. Bezembinder. Junction modelling 2.0. Technical report, Windesheim, Zwolle, The Netherlands, 2018.

[38] A. Dijkstra. Enkele aspecten van verkeersveiligheid. Research report R-2014-21A, Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV, Den Haag, The Netherlands, 2014.

[39] R. H. Gasthuis. Emission modelling on junctions in urban networks. Master's thesis, University of Twente, Enschede, The Netherlands, 2015.

[40] FHWA. Manual on uniform traffic control devices for streets and highways (mutcd). Technical report, Federal Highway Agency of the United States Department of Transportation, Washington D.C., United States, 2000.

Effect of Various Parameters for Temperature Distribution in Human Body: An Analytic Approach

Kabita Luitel[*,2], Dil Bahadur Gurung[1], Kedar Nath Uprety[2]

[1]*Department of Mathematics, School of Science, Kathmandu University, 45210, Nepal*

[2]*Central Department of Mathematics, Tribhuvan University, 44618, Nepal*

A R T I C L E I N F O	A B S T R A C T
	This paper is the extension of the work originally presented in 2nd International Conference on Man and Machine Interfacing (MAMI 2017) and the paper of Yue et al. The present study aims at observing the effect of various parameters on temperature distribution profiles at various environmental temperature, tissue thermal conductivities, metabolic rates, blood perfusion rates, and heat transfer coefficients. The analytic solution of Pennes' bioheat equation in the steady-state case is obtained by using the Modified Bessel's equation incorporating the effect of sweating and non-sweating state of the body. From the study, it was observed that the variation of atmospheric temperature and heat transfer coefficients have a significant effect for the temperature distribution in the body towards the skin surface.

1. Introduction

Heat transfer plays a very important role in the living system. This is the complex process, which includes not only some physical factors—conduction, convection, radiation, evaporation but also the physiological factors—blood flow and metabolism. These physical and physiological factors help maintain the constant human body temperature around 37^0C which is the equilibrium point of the production of heat and loss of heat by the body. The extreme temperatures from the normal body temperature influence the function of biological tissue and the whole system of the body. Death may occur if the body temperature is 27^0C and below and if 42^0C and above. So we should maintain the body temperature around 37^0C. The heat transfer in the blood vessel also helps to maintain uniform body core temperature regardless of changes in environmental temperature. Metabolism, another important source of heat gain is the chemical process that occurs in the living organism to grow and produce, maintain their structure, and respond to their respective environment [1, 2, 3].

Sweat evaporation is one of the effective parameters in the thermoregulatory process, which is the only way to lose heat when the ambient temperature is higher than the normal body (36.1^0C - 37.2^0C) temperature. Due to the sweat evaporation, 22% of heat losses from the body. Evaporative heat exchange also involves the loss of heat through the evaporation of sweat from the skin surface. We generally calculate the rate of sweat evaporation as the weighted mean value of the body core and skin temperature. The reasonable equation for the sweat rate is [4, 5].

$$E = 8.47 \times 10^{-5}\{(0.1 \times T_{sk} + 0.9 \times T_b) - 36.6°C\}Kg/m^2/sec$$

Where $T_{sk} = $ skin surface temperature

and $T_a = 37°C$ (body core temperature)

It has been proved that the transfer of latent heat from the living being to its environment is often estimated by multiplying the loss in weight attributable to evaporation by the latent heat vaporization of water 'L' which decreases from 2501J/g at 0^0C to 2406 J/g at 40^0C[6]. Havenith et al.[7] however, have mentioned the values for 'L' ranging from 2,398 J/g to 2,595J/ g. and finally suggested the latent heat of evaporation is only dependent on temperature giving a number of 2,430J/g at 30^0C.

Acharya et al. [5] used variational finite element (FEM) method to prepare one dimensional heat transfer model for the comparative study of temperature profiles of human male and

*Corresponding author: Kabita Luitel, Department of Mathematics, Bhaktapur Multiple Campus, TU, Nepal, E-mail: kabi123luitel@gmail.com

temperature profiles of female luteal and follicular phases of the menstrual cycle.

Havenith et al. [7] studied experimentally by using thermal manikin to determine the effective cooling power of moisture evaporation. They measured both heat loss and mass loss independently by allowing a direct calculation of an effective latent heat of evaporation.

The knowledge of heat transfer is equally important in the field of biomedical research as well as in the treatment of cancer, now a great threat to the existence of humanity. Bioheat transfer models have therapeutic and clinical importance. These models are helpful for the more effective treatment of cancer than it is now. As per Arkin et al., the effect of hyperthermia treatment depends on the temperature and duration of heating. If a constant temperature could be maintained, the duration of heating would be a reasonable way of expressing thermal dose [8].

Over a hundred years' time, the effects of the blood flow have been examined on the heat transfer of living tissue. After Bernard came up with an experimental study in 1876, physicians, physiologists, and engineers are interested in the mathematical modeling of the complex thermal interaction between the vascular systems of the body.

In 1948 pennes' H.H. [9] proposed the simple linear mathematical model based on experimental observation for describing heat flow within the tissue. Many other researchers, one after another, have developed alternative models for describing the perfusion rate and difference between the arterial blood temperature and the local tissue temperature. But Pennes' model still has an acceptable result to predict the transient temperature due to its simplicity and flexibility [1].

Previous researchers such as Acharya et al. [5], Khandey and Saxena [10], Gurung and Saxena [11], Aijar and Dar [12], Nadel [13] observed the effect of latent heat of sweat evaporation for temperature variation in the human body only by using numerical (Specially FEM) techniques. Numerical methods give the approximation result whereas if the analytic solutions of these equations are attainable, they will give the exact result. Even though the analytical method was used by Yue et al.[2] to solve the model, but this model has not incorporated the effect of Latent heat and sweat evaporation. The analytical approach for the study of such problems is still lacking. So the present study focuses on a mathematical model of the body temperature based on the Pennes' bioheat equation due to the sweat evaporation. The temperature at various physical and physiological conditions in the sweating case in comparison with the non-sweating case will be observed in the study

2. Mathematical Model

The governing differential equation used in the model is given by [9].

$$\rho c \frac{\partial T}{\partial t} = \nabla.(K\nabla T) + M(T_a - T) + S \qquad (1)$$

Where, ρ = density of tissue (kg/m³), c = specific heat (J/Kg. ⁰C), K = thermal conductivity (W/m⁰C), ρ_b = density of blood (Kg/m³), w_b = blood perfusion rate per unit volume (m³/s.m³), c_b = blood specific heat (J/kg.⁰C), M = $\rho_b w_b c_b$ (W/m³.⁰C), S = metabolic

heat generation (W/m³), T_a = temperature of arterial blood (⁰C), T = tissue temperature (⁰C).

2.1. Boundary Conditions

Heat loss takes place from the surface of the human body due to convection, radiation, and evaporation because the human body surface is exposed to the environment. So the boundary conditions used in this study is given by

$$R = 0, \qquad \frac{dT}{dr} = 0 \qquad (2)$$

$$R = r, \qquad -K\frac{dT}{dr} = h_C(T - T_\infty) + LE \qquad (3)$$

Where R is the radius of concerned tissue (m), h_C is the coefficient of heat transfer on the surface of the tissue (W/m².⁰C), T_∞ is atmospheric temperature (⁰C), L is the latent heat (J/Kg) and E denotes the evaporation rate (Kg /m².sec).

3. Analytic Solution

Being human body cylinder in shape, equation (1) has been converted into the cylindrical form. The one dimensional steady–state equation in the radial direction is expressed as,

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{dT}{dr}\right) + \frac{M}{K}(T_a - T) + \frac{S}{K} = 0 \qquad (4)$$

To perform the non-dimensionalization of (4) with boundary conditions (2) and (3), we introduce the following characteristic quantities and dimensionless parameters,

$$\tilde{r} = \frac{r}{R}, \text{ and } \tilde{T} = \frac{T-T_\infty}{T_a-T_\infty}, \qquad (5)$$

$$\widetilde{M} = \frac{MR^2}{K}, \tilde{S} = \frac{SR^2}{K(T_a-T_\infty)}, \qquad (6a)$$

$$\widetilde{h_C} = \frac{h_C R}{K}, \qquad \widetilde{N} = \frac{LER}{K(T_a-T_\infty)} \qquad (6b)$$

Differentiating **(5)** with respect to 'r' then, substituting in (4) we get

$$\frac{1}{\tilde{r}}\frac{d}{d\tilde{r}}\left(\tilde{r}\frac{d\tilde{T}}{d\tilde{r}}\right) + \frac{R^2M(1-\tilde{T})}{K} + \frac{SR^2}{K(T_a-T_\infty)} = 0 \qquad (7)$$

With the use of **(6a)** and **(6b), (7)** reduces after calculating to

$$\frac{1}{\tilde{r}}\frac{d}{d\tilde{r}}\left(\tilde{r}\frac{d\tilde{T}}{d\tilde{r}}\right) - \widetilde{M}\tilde{T} + (\widetilde{M} + \widetilde{N}) = 0 \qquad (8)$$

For computational simplicity, again we put

$$\boldsymbol{M} + \boldsymbol{N} = \text{U}$$

$$\widetilde{\boldsymbol{M}} = \text{V}$$

$$\emptyset = \text{U} - \text{VT}$$

And get the **(9)** as follows

$$\tilde{r}^2 \frac{d^2\emptyset}{d\tilde{r}^2} + \tilde{r}\frac{d\emptyset}{d\tilde{r}} - V\tilde{r}^2\emptyset = 0 \qquad (9)$$

This equation is the Modified Bessel's equation of zero order and comparing it with Modified Bessel equation

$$\left[\begin{array}{c} x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} - (\beta^2x^2 + p^2)y = 0 \\ \textit{Whose solution is given by} \\ y = AI_p(\beta x) + BK_p(\beta x) \end{array}\right]$$

The solution of equation (9) for $\emptyset$ can be written as

$$\emptyset = c_1 I_0(\sqrt{V}\tilde{r}) + c_2 K_0(\sqrt{V}\tilde{r})$$

Where c_1 and c_2 are arbitrary constants.

After differentiating $\emptyset$ with respect to $\tilde{r}$, calculating and substituting the corresponding values with boundary conditions, the solution for $\tilde{T}$ and then the solution for T can be written as in (10) and (11).

$$\tilde{T} = \frac{U}{V} - \frac{1}{V}\left[\left\{\frac{\widehat{h_c}U + \tilde{N}V)I_0(\sqrt{V}\tilde{r})}{\widetilde{h_C}I_0(\sqrt{V}\tilde{r}) + \sqrt{V}I_1(\sqrt{V}\tilde{r})}\right\}\right] \qquad (10)$$

$$T = T_\infty + (T_a - T_\infty)\left[\frac{U}{V} - \frac{1}{V}\left\{\frac{(\widehat{h_C}U + V\tilde{N})I_0((\sqrt{V}\tilde{r}))}{\widetilde{h_C}I_0(\sqrt{V}\tilde{r}) + \sqrt{V}I_1(\sqrt{V}\tilde{r})}\right\}\right] \qquad (11)$$

Table 1: Thermo-physical parameters [1, 14]

Parameters	Symbols	Values	Unit
Thermal conductivity	K	0.48	W/m^{0}C
Blood specific heat	c_b	3850	J/Kg^{0}C
Blood density	ρ_b	1000	Kg/m^3
Perfusion rate	w_b	3	Kg/s.m^3
Metabolism	S	1085	W/m^3
Arterial temperature	T_a	36.98	^{0}C
Tissue thickness	R	0.03	M
Heat transfer coefficient	h_C	10.023	W/m^2.^{0}C
Latent heat	L	2400000	J/Kg
Evaporation rate	E	0.00004	Kg/m^2.sec
Environmental temperature	T_∞	30	^{0}C

The effect at various temperature profiles based on the above parameters values are presented graphically and discussed below.

4. Results and Discussions

The analytical solution of the bioheat equation for the cylindrical body and its respective results of physical properties depends on various factors. So the following values of parameters from table 1 have been used in equation (11) to observe the effect of different atmospheric temperature, thermal conductivities, blood perfusion rates, metabolic heat generation, and heat transfer coefficients.

4.1. Effect of Atmospheric Temperature

The temperature profiles at a various atmospheric temperature below 37^{0}C (normal body temperature) at $T_\infty = 25^0C$, 28^0C, 31^0C and 34^0C, and above 37^{0}C at $T_\infty = 38^0C$, 41^0C, 44^0C and 47^0C, are considered and shown respectively in Figures 1(a), 1(b) and 1(c), 1(d).

Although the graphs in both Figures 1(a) and 1(b) show that the body temperature decreases from the core of the body towards the skin surface if the atmospheric temperature is less than the body core temperature, the graphs in Figure 1(b) show that the body temperature decreases more quickly towards the skin surface than in Figure 1(a). This is due to the evaporation effect. On the other hand, if the atmospheric temperature is higher than the body core temperature, the body temperature increases from the core of the body towards the skin surface, which can be seen in Figures 1(c) and 1(d). Moreover, the graphs in Figure 1(d) show that the temperature increases more slowly towards the skin surface than in the graphs in Figure 1(c) due to sweat evaporation, which helps to cool down the body temperature.
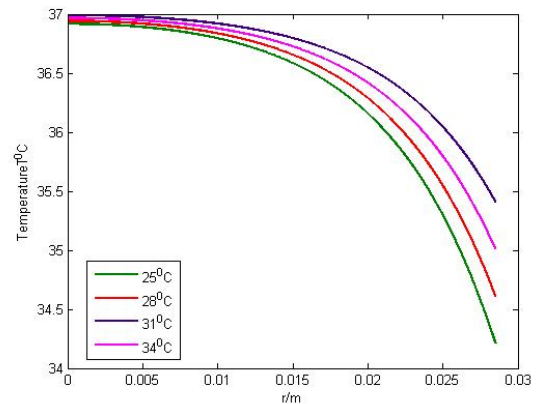
Figure 1(a): Effect of $T_\infty < 37^0$C when $E= 0$
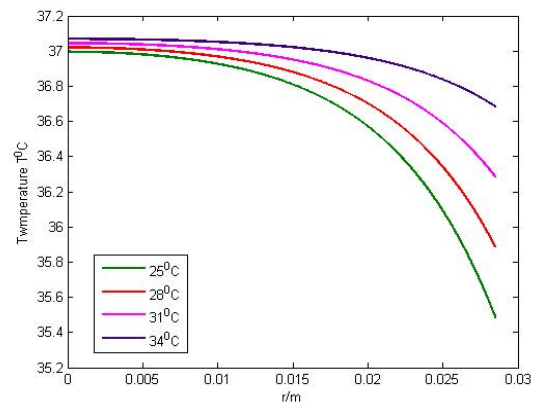
Figure 1(b): Effect of $T_\infty < 37^0$C with LE

Figure 1(c): Effect of $T_\infty > 37^0C$ when $E= 0$

Figure 2(b): Effect of thermal Conductivities with *LE*

4.3. Effect of Blood Perfusion

The various values of blood perfusion are taken as w_b = 1.5 Kg/s.m^3, 2 Kg/s.m^3, 2.5 Kg/s.m^3 and 3 Kg/s.m^3 to observe their effect on the temperature of the body. The graphs in Figures 3(a) and 3(b) indicate that the gradient temperature variation in radial direction decreases with the increase of the blood perfusion. Figure 3(b) shows that the temperature falls more sharply towards the surface of the body than in Figure 3(a) because of the evaporation effect together with the blood perfusion.

Figure 1(d): Effect of $T_\infty > 37^0C$ with *LE*

4.2. Effects of Thermal Conductivities

The temperature profiles at various thermal conductivities K = 0.24W/m^{0}C, 0.48W/m^{0}C, 0.60W/m^{0}C and 0.72W/m^{0}C are shown graphically in Figures 2(a) and 2(b). With the increase in thermal conductivities, the inner part of body temperature increases but the temperature on the body surface decreases sharply as the thermal conductivities increase. It happens due to the conduction process at the body surface.

Figure 3(a): Effect of blood perfusion when $E= 0$

Figure 2(a): Effect of thermal Conductivities when $E= 0$
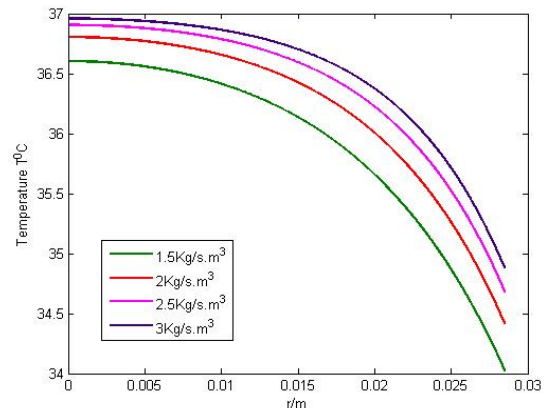
Figure 3(b): Effect of blood perfusion with *LE*

4.4. Effect of Metabolic Heat Generation

The various values of metabolic heat generation q_m = 141.05 W/m^3, 271.25 W/m^3, 542W/m^3 and 1085 W/m^3 are taken. Graphs in Figures 4(a) and 4(b) show that the values of metabolic heat generation have a very small effect on temperature distribution in the human body. The change in metabolic heat from 542W/m^3 to 1085 W/m^3 make the small change in body temperature almost 0.1^{0}C whereas its effect on the skin surface is negligible. This happens due to the negligible concentration of blood vessels towards the skin surface. In the case when sweat evaporation is present (Figure 4(b)), the effect of metabolic heat is almost negligible.

Figure 5(a): Effect of h_C when E= 0

Figure 4(a): Effect of metabolism when $E = 0$
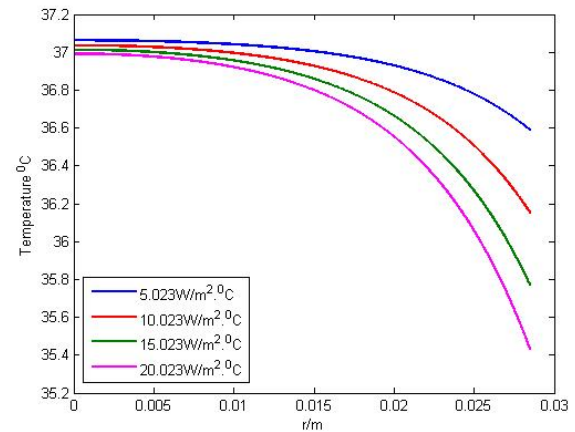
Figure 5(b): Effect of h_C with LE

Figure 4(b): Effect of metabolism with LE

4.5. Effect of Heat Transfer Coefficients

The various values of heat transfer coefficients are considered at h_C = 5.023 W/m^2.^{0}C, 10.023 W/m^2.^{0}C, 15.023 W/m^2.^{0}C and 20.023 W/m^2.^{0}C. The significant effect of the heat transfer coefficient can be seen in Figures 5(a) and 5(b). The curves in these figures indicate that the gradient temperature in the radial direction decreases with the increase of the heat transfer coefficients.

5. Conclusion

In the present study, the analytical solution of the cylindrical form of Pennes' bioheat equation with boundary conditions including the latent heat of sweat evaporation in the one-dimensional steady-state case is obtained. The effect of the temperature changes in various atmospheric temperature, thermal conductivities, metabolic heat generation rates, blood perfusion rates, and heat exchange coefficients has been observed by using the solution obtained from this model. The study reveals that the atmospheric temperature, blood perfusions, and the heat transfer coefficients have a more significant effect on the temperature variation on the skin surface than in the body core. The gradient temperature in the radial direction decreases with the increase of the ambient temperature, the blood perfusion rates, and the heat transfer coefficients. There is a slight variation in body core temperature even an increase in metabolic heat generation. Its effect on the skin surface is negligible. The various parameters used in this model have certainly a more remarkable effect on the temperature distribution in the human body than the case suggested in [1, 2]. Such a model may be useful for the researcher as well as thermal diagnosis and hyperthermia treatment of cancer. The analytic solution obtained in this paper can also be extended in the axial and angular direction as well as unsteady state case.

Conflict of Interest

The authors make it sure that there is no conflict of interest.

References

[1] Kabita Luitel, "Mathematical Model for Temperature Distribution in Cylindrical Human Body" IEEE 2017, 2nd International Conference on Man and Machine Interfacing (MAMI), IEEE 2017.

[2] Yue Kai Zhang, Fan You Xinxin, An Analytic Solution of One – dimensional Steady – state Pennes Bio-heat.Transfer Equation in Cylindrical Coordinates, Journal of Thermal Science, 13(3),52 - 54, 2004.

[3] Ahmed Lakhssassi, Kengne, and Emmanuel Semmaoui, Modified Pennes' equation modeling bio-heat transfer in living tissues: analytical and numerical analysis, Natural Sciences, B, 1375 - 1385, 2010.

[4] Hoppe PR., Heat Balancing modeling, Institute, and Outpatient Clinic for Occupational Medicine, Experientia. 49:741-746, 1993.

[5] Saraswati Acharya, D.B. Gurung, V. P. Saxena, Mathematical Modeling of Sex-Related Differences in the Sensitivity of the Sweating Heat Responses to Change in Body Temperature, British Journal of Mathematics & Computer Science, 12(4), 1 -11, 2016. www.sciencedomain.or.com

[6] Monteith JL., Latent Heat of Vaporization in Thermal Physiology, Nature New Biology 236: 1996, 1972.

[7] George Havenith, Peter Bröde, Emiel den Hartog, Kalev Kuklane, Ingvar Holmer, Rene M. Rossi, Mark Richards, Brian Farnworth, and Xiaoxin Wang, Evaporative cooling: the effective latent heat of evaporation in relation to evaporation distance from the skin, J Appl Physiol, 114, 778 - 785, 2013. doi:10.1152/japplphysiol.01271.2012.

[8] H Arkin., Lx Xu, and K.R Holms, Recent Developments in Modeling, Heat transfer in blood Perfused Tissue, IEEE Transactions on Biomedical Engineering, 41, 97 - 107, 1994.

[9] H.H. Pennes', Analysis of Tissue and Arterial Blood Temperatures in Resting human forearm, Journal of Applied Physiol, 1, 93 - 1948.

[10] M.A. Khandey and V. P. Saxena, FEM Estimation of One-dimensional Unsteady-State Heat Regulation in Human Head exposed to the Cold Environment, Journal of Biological system, 17, 853 -- 863, 2009.

[11] D.B Gurung, V.P Saxena, and P.R Adhikari, Finite Element approach to The one-dimensional Steady-state Temperature Distribution in the Dermal parts with Quadratic Shape Function, J.Appl. Math and information, 27, 301 - 313, 2009.

[12] Mir Aijar and Javid Gani Dar, Mathematical Analysis of Bioheat Equation for the Study of Thermal Stress on Human Brain, Applied Mathematics & Information Sciences and Letters An International Journal, 1 5(1), 33 - 39, 2017.

[13] ER Nadel, RW Bullard, JAJ Stolwijk. Importance of skin temperature in the regulation of sweating. J. Appl. Physiol., 46, 430 - 437, 1979.

[14] Lalif M. Jiji, Heat Conduction, 3rd Edition Springer, 52 – 54, 2009.

[15] W.J Minkowycz and E.M, Sparrow Advance in Numerical Heat Transfer Printed in the United States of America, (III), 2009.

[16] Kabita Luitel, D.B. Gurung, Mathematical Model for temperature Distribution in Cylindrical Human Body, The Nepali Mathematical Sciences Report, 3, (1&2), 19 - 28, 2012.

[17] Mir Aijar and M. A. Khanday, Temperature Distribution and Thermal Damage of Peripheral Tissue in Human Limbs During Heat Stress, A Mathematical Model, Journal of Mechanics in Medicine and Biology, 16(2), 16500--16517, 2016.

[18] Ying He, A Numerical coupling model to analyze the blood flow temperature and oxygen transport in Human Breast Tumor and laser irradiation, *Japanese Journal of Computers in Biology and Medicine,* (2006), 36, P.1336-1350.

[19] D.B Gurung, Saraswoti Acharya, Five Layered Temperature distribution in Human dermal Part. Nepali Mathematical Sciences Report, Central Department of Mathematics, Tribhuvan University, Kirtipur, Nepal, 2011.

[20] D.B Gurung, Kabita Luitel, Development of Bio-Heat Equation and its Application, Proceedings of Nepal Mathematical Society, 85 - 94, 2012.

[21] Mamata Agrawal, Neeru Adlakha, R.K Pardadsani, Modeling and Simulation of the thermal effect of Metastasis Timors in Human Limbs, International Symposium on Devices MEMS, Intelligent system and communication (ISDMISC), 24 - 29, 2011.

[22] Gokul K C, D.B. Gurung, and P. R. Adhikari, Thermal Effect of the eyelid in human eye temperature model, J. App. Math. & informatics 32(5-6), 649 - 663, 2014.

[23] Luisiana X. Cundin, William P. Roach, Nancy Millenbaugh, Empirical Comparison of Pennes' Bio-Heat Equation, Proc. Of SPINE, 7175, 717516-1 -- 717516-9.

Improved DTC Control Strategy of B12 Inverter Fed BLDC Motor Drives Considering Commutation Torque Dips

Rabiaa Mars[*], Badii Bouzidi, Bassem El Badsi, Abderrazak Yangui

University of Sfax, Electrical Department, Sfax Engineering National School, P.O. Box 1173, 3038 Sfax, Tunisia

A B S T R A C T

In most industrial high-power applications, the brushless DC (BLDC) motor is connected to multi-level inverters specially the three-level NPC inverters (B12). Torque ripple is a critical issue in BLDC motor drives. Accordingly, minimizing the ripple produced in the torque is necessary to enhance the BLDC motor drive performances. This paper aims to develop two direct torque control (DTC) schemes. To evaluate the proposed strategies performances, three DTC schemes, namely DTC- 1, DTC- 2, and DTC-3, are treated. The DTC-1 scheme is inspired from the Takahashi's strategy proposed in induction motor drives. This scheme is penalized by the high torque dips. In view of that, the DTC-2 is proposed to address the DTC-1 limitations. Simulation evaluation proves that difference between the slopes of the incoming and outgoing phases currents procreates commutation torque dips penalizing the DTC-2 scheme. This difference is more serious at high speed. To overcome the DTC-2 strategy drawbacks, a DTC-3 scheme, based on the compensation of this difference, is proposed. Taking into account the increase of this difference with the increase of the rotation speed, the vector selection table design of this strategy is synthesized to minimize the torque ripple as far as possible.

1 Introduction

The use of three-level NPC (B12) inverter in high-power applications increases due to their performances [1]- [2]. Indeed, compared to the two-level inverter, the B12 inverter offers the low output current distortion and operates with reduced dv/dt stress [3]-[4]-[5]-[6].

Due to its numerous advantages as its simple structure, low noise, good stability and high efficiency [7]-[8]-[9], BLDC motor is extensively used in numerous industrial applications. Conversely, the torque ripple is a critical issue in BLDC motor drives [10]-[11]-[12]. Accordingly, minimizing the ripple produced in the torque is necessary to enhance the performances of BLDC motor drive system [13]-[14].

The present paper proposes two DTC strategies dedicated to the BLDC motor fed by three-level (B12) inverter. It has been found that commutation torque ripple is a main problem penalizing the first proposed strategy. This commutation phenomenon affects the industrial applications needing high performance [15]-[16]. It leads to a deteriorating efficiency of the BLCD motor. Hence, the second proposed strategy aims to address these drawbacks.
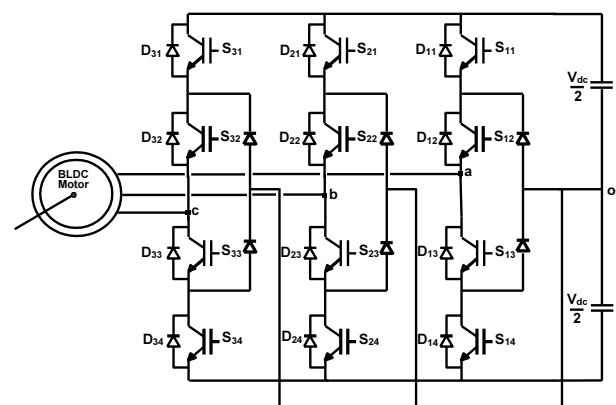
Figure 1: B12 inverter with BLDC motor connected.

[*]Rabiaa Mars, Sfax Engineering National School, marsrabiaa@yahoo.com

2 Analysis of the BLDC motor operation

The B12 inverter is connected with BLDC motor as illustrated in Figure.1

The rectangularly shaped currents(i_a, i_b, i_c) fed BLDC motor are synchronized with the trapezoidal back-EMFs(e_a, e_b, e_c) to accomplish a constant torque with reduced dips as shown in Figure.2.

Figure 2: Back-EMFs, phase currents and the corresponding Hall-effect signals of BLDC motor .

The BLDC motor stator voltages (v_a, v_b, v_c) are given by:

$$\begin{cases} v_a = Ri_a + (L\text{-}M)\frac{di_a}{dt} + e_a; \\ v_b = Ri_b + (L\text{-}M)\frac{di_b}{dt} + e_b; \\ v_c = Ri_c + (L\text{-}M)\frac{di_c}{dt} + e_c. \end{cases} \quad (1)$$

Where:

- Resistance: R (Ω),

- Self inductance: L(H),

- Mutual inductance: M(H).

The produced torque T_{em} is given as:

$$T_{em} = \frac{e_a i_a + e_b i_b + e_c i_c}{\Omega} \quad (2)$$

Where Ω represents the angular speed.

Under two-phase conduction mode, the B12 inverter engenders six Half voltage vectors (V_1 to V_6) and six Intermediate Voltage Vectors (V_7 to V_{12}).

Under three-phase conduction mode, the B12 inverter generates three null vectors (U_{01} to U_{03}), six

Intermediate voltage vectors (U_1 to U_6) and six Full Voltage Vectors (U_7 to U_{12}).

Tables.1 and 2 provide the switch state combinations.

Table 1: Active voltage vectors under the two-phase conduction mode.

(S_a S_b S_c)	V_α	V_β	Vectors
Half Voltage Vectors			
(1100 0000 0000)	$\frac{1}{\sqrt{6}}V_{dc}$	0	V_1
(0000 0000 0011)	$\frac{1}{\sqrt{24}}V_{dc}$	$\frac{1}{\sqrt{8}}U_c$	V_2
(0000 1100 0000)	$-\frac{1}{\sqrt{24}}V_{dc}$	$\frac{1}{\sqrt{8}}U_c$	V_3
(0011 0000 0000)	$-\frac{1}{\sqrt{6}}V_{dc}$	0	V_4
(0000 0000 1100)	$-\frac{1}{\sqrt{24}}V_{dc}$	$-\frac{1}{\sqrt{8}}U_c$	V_5
(0000 0011 0000)	$\frac{1}{\sqrt{24}}V_{dc}$	$-\frac{1}{\sqrt{8}}U_c$	V_6
Intermediate Voltage Vectors			
(1100 0000 0011)	$\sqrt{\frac{3}{8}}V_{dc}$	$\frac{1}{\sqrt{8}}V_{dc}$	V_7
(0000 1100 0011)	0	$\frac{1}{\sqrt{2}}V_{dc}$	V_8
(0011 1100 0000)	$-\sqrt{\frac{3}{8}}V_{dc}$	$\frac{1}{\sqrt{8}}V_{dc}$	V_9
(0011 0000 0011)	$-\sqrt{\frac{3}{8}}V_{dc}$	$-\frac{1}{\sqrt{8}}V_{dc}$	V_{10}
(0000 0011 1100)	0	$-\frac{1}{\sqrt{2}}V_{dc}$	V_{11}
(1100 0011 0000)	$\sqrt{\frac{3}{8}}V_{dc}$	$\frac{1}{\sqrt{8}}V_{dc}$	V_{12}

Figure 3: DTC scheme of the B12 inverter with BLDC motor connected.

3 Control strategies

3.1 DTC-1 strategy

The DTC-1 scheme is inspired from the one proposed by *Takahashi* in [17]. This strategy allows a direct control of the BLDC motor torque.

The DTC-1 scheme block diagram is illustrated in Figure.3. This diagram includes an hysteresis torque estimator which generates two outputs used to determine the vector selection table($c_\tau = +1$:torque increase, $c_\tau = -1$:torque decrease).

The switching table of this strategy is summarized in Table 3.

Table 2: Active voltage vectors under the three-phase conduction mode.

Table 3: The DTC-1 strategy vector selection table.

c_τ	$+1$	-1
Sector I	V_8 (0000 1100 0011)	V_{11} (000000111100)
Sector II	V_9 (001111000000)	V_{12} (110000110000)
Sector III	V_{10} (001100000011)	V_7 (110000000011)
Sector IV	V_{11} (000000111100)	V_8 (0000 1100 0011)
Sector V	V_{12} (110000110000)	V_9 (001111000000)
Sector VI	V_7 (110000000011)	V_{10} (001100000011)

(S_a S_b S_c)	V_α	V_β	Vectors
Intermediate Voltage Vectors			
(1100 0010 0011)	$\sqrt{\frac{3}{8}}V_{dc}$	$\frac{1}{\sqrt{8}}V_{dc}$	U_1
(0100 1100 0011)	0	$\frac{1}{\sqrt{2}}V_{dc}$	U_2
(0011 1100 0010)	$-\sqrt{\frac{3}{8}}V_{dc}$	$\frac{1}{\sqrt{8}}V_{dc}$	U_3
(0011 0100 0011)	$-\sqrt{\frac{3}{8}}V_{dc}$	$-\frac{1}{\sqrt{8}}V_{dc}$	U_4
(0010 0011 1100)	0	$-\frac{1}{\sqrt{2}}V_{dc}$	U_5
(1100 0011 0100)	$\sqrt{\frac{3}{8}}V_{dc}$	$U6$	
Full Voltage Vectors			
(1100 0011 0011)	$\sqrt{\frac{2}{3}}V_{dc}$	0	U_7
(1100 1100 0011)	$\frac{1}{\sqrt{6}}V_{dc}$	$\frac{1}{\sqrt{2}}V_{dc}$	U_8
(0011 1100 0011)	$-\frac{1}{\sqrt{6}}V_{dc}$	$\frac{1}{\sqrt{2}}V_{dc}$	U_9
(0011 1100 1100)	$-\sqrt{\frac{2}{3}}V_{dc}$	0	U_{10}
(0011 0011 1100)	$-\frac{1}{\sqrt{6}}V_{dc}$	$-\frac{1}{\sqrt{2}}V_{dc}$	U_{11}
(1100 0011 1100)	$\frac{1}{\sqrt{6}}V_{dc}$	$-\frac{1}{\sqrt{2}}V_{dc}$	U_{12}

(a)

(b)
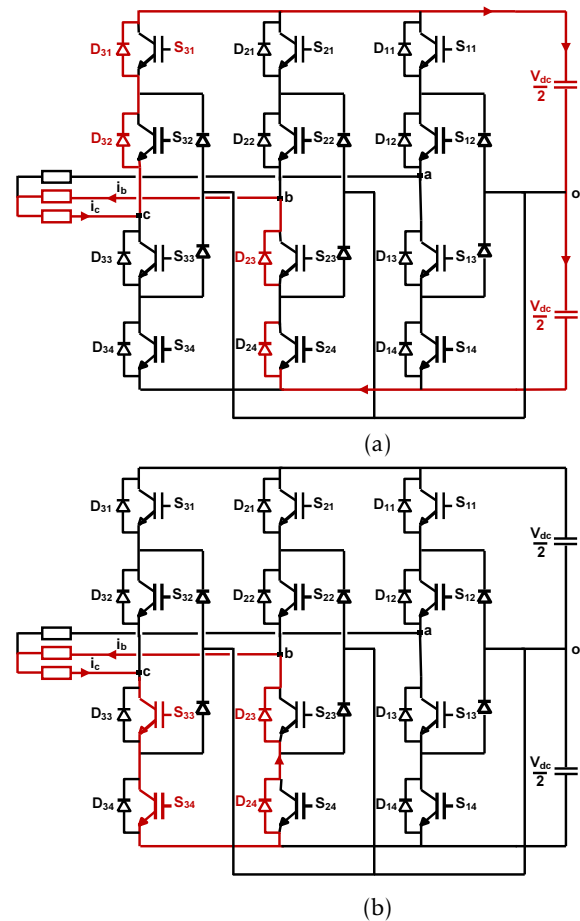
Figure 4: The current flowing during sector I in the case of a torque decrease under **Legend** (a): the DTC-1 scheme, (b): the proposed DTC-2 scheme.

3.2 DTC-2 strategy

It has been found that DTC-1 scheme is penalized by the high torque dips. In view of that, the DTC-2 is proposed to address the DTC-1 limitations. The control scheme of this strategy is similar to the one used in the DTC-1. The only difference is the control rules considered in the vector selection table.

Let's consider sector I, the vector selection table design of this scheme is synthesized as follows:

- To increase the torque, the DTC-2 scheme considers the same vectors applied in DTC-1,

- the vector V_2 achieves the torque decrease. As

illustrated in Figure. 4(b), the vector V_2 is defined by a freewheeling phase involving the two IGBTs S_{33} and S_{34} and the diodes D_{23} and D_{24}. Compared to DTC-2, the DTC-1 scheme considers the vector V_8 which characterized by a regenerative phase. During this phase, the currents i_c and i_b are forced to flow through the supply source Vdc and the diodes (D_{23}, D_{24}, D_{31} and D_{32}) as shown in Figure. 4(a) . The currents slope observed under this phase has an absolute value higher than the one obtained under the freewheeling phase.

Consequently, the resulting DTC-2 strategy vector selection table is given in Table.4.

Table 4: The DTC-2 strategy vector selection table.

c_τ	+1	−1
Sector I	V_8(000011000011)	V_2(000000000011)
Sector II	V_9(001111000000)	V_3(000011000000)
Sector III	V_{10}(001100000011)	V_4(001100000000)
Sector IV	V_{11}(000000111100)	V_5(000000001100)
Sector V	V_{12}(110000110000)	V_6(000000110000)
Sector VI	V_7(110000000011)	V_1(110000000000)

3.3 DTC-3 strategy

It has been found that DTC-2 scheme is penalized by the commutation torque ripple. As illustrated in Figure.5(a), the difference between the slopes of the incoming and outgoing phases currents generates commutation torque dips. In view of that, there is a necessity to make equal the rising time and the falling one of the phase currents as illustrated in Figure.5(b).
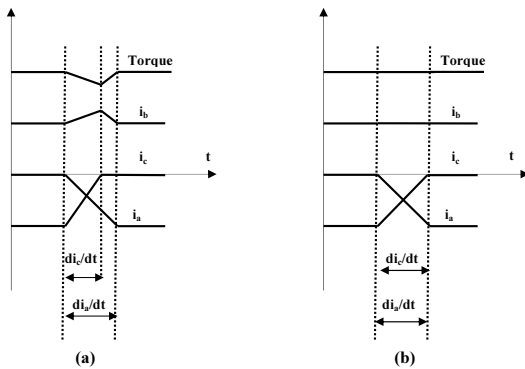
Figure 5: The behavior of the torque during commutation **Legend** (a): unequal slopes of the incoming and outgoing phases currents, (b): equal slopes of the incoming and outgoing phases currents.

The stator voltage equations of the BLDC motor are described by the equation (1). Neglecting the mutual inductance, v_a, v_b and v_c are given by the following system :

$$\begin{cases} v_a - e_a = & Ri_a + L\frac{di_a}{dt} \; ; \\ v_b - e_b = & Ri_b + L\frac{di_b}{dt} ; \\ v_c - e_c = & Ri_c + L\frac{di_c}{dt} . \end{cases} \quad (3)$$

The Laplace Transform method can be used to solve this linear differential equations. Therefore the relation between the phase voltage and the phase current is dealt with using this method, and then the transfer function is similarly written as:

$$H(p) = \frac{v(p) - e(p)}{i(p)} = \frac{1}{Lp - R} \quad (4)$$
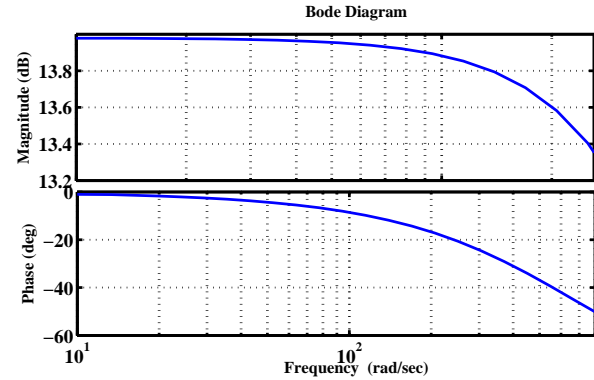
Figure 6: Approximated curve of the frequency response (bode-plot) of the BLDC motor.

Figure6 shows the approximated curve of the frequency response (bode-plot) based on equation (4). The plot displays the magnitude (in dB) and the phase (in degrees) of the system response as a function of frequency. Referring to this figure, the phase delay angle increases with the increase of the rotation speed. Moreover, this delay is more important at high speed. Consequently, the commutation torque dips is very serious at high speed.

To address the DTC-2 limitations, the third DTC scheme, noted DTC-3, is proposed . This strategy is based on the employ of a three-level torque controller. This controller generates 3 outputs as shown in Figure.7 :

- -1 corresponding to the torque decrease,

- +1 corresponding to the torque increase,

- +2 is activated when the torque drops during commutations from sector-to-sector .
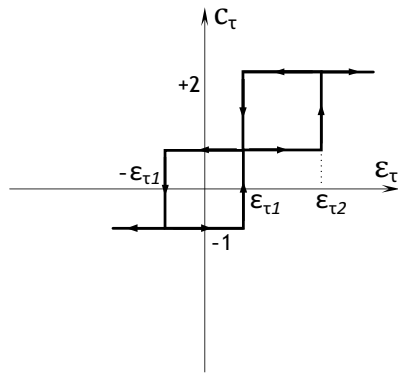
Figure 7: The three-level-torque controller characteristics.

Let's consider sector I, the vector selection table design of this scheme is synthesized as follows:

- c_τ=+1 and c_τ=-1 are achieved by the same vectors used in DTC-2 scheme.
- The suppression of the commutation torque dips (c_τ=+2) is achieved by the application of the vectors corresponding to the three-phase conduction mode. To equalize $|di_a/dt|$ and $|di_c/dt|$,at the beginning of sector II, in the case of a reference speed $\Omega^* \leq \Omega_{limit}$, the current i_c is forced to flow through the IGBT S_{33}, the clamped diode and the supply source $\frac{V_{dc}}{2}$ instead of the uncontrollable diodes (D_{31} and D_{32}) as shown in Figures. 8(b) and 8(c). For a reference speed $\Omega > \Omega_{limit}$, the current i_c is forced to flow through the IGBTs (S_{33} , S_{34}) and the supply source V_{dc} instead of the uncontrollable diodes (D_{31} and D_{32}) as shown in Figures. 8(b) and 8(d).

Consequently,the resulting vector selection is given in Table.7.

4 Evaluation of the proposed DTC schemes effectiveness and performances

Simulation work has been accomplished to highlight the proposed strategies performances. Tables.5 and 6 gives the BLDC motor parameters and ratings.

Table 5: The ratings of BLDC motor

Current = 30A	DC voltage = 24V
Rated torque = 2.5Nm	PM flux-linkage = 14mWb
Power = 600W	Rated speed = 2500rpm

Table 6: BLDC motor parameters

R = 0.2Ω	p = 3
L = 0.3mH	J = 4.1g.m^2
R = 0.2Ω	f = 0.5mN.m.s
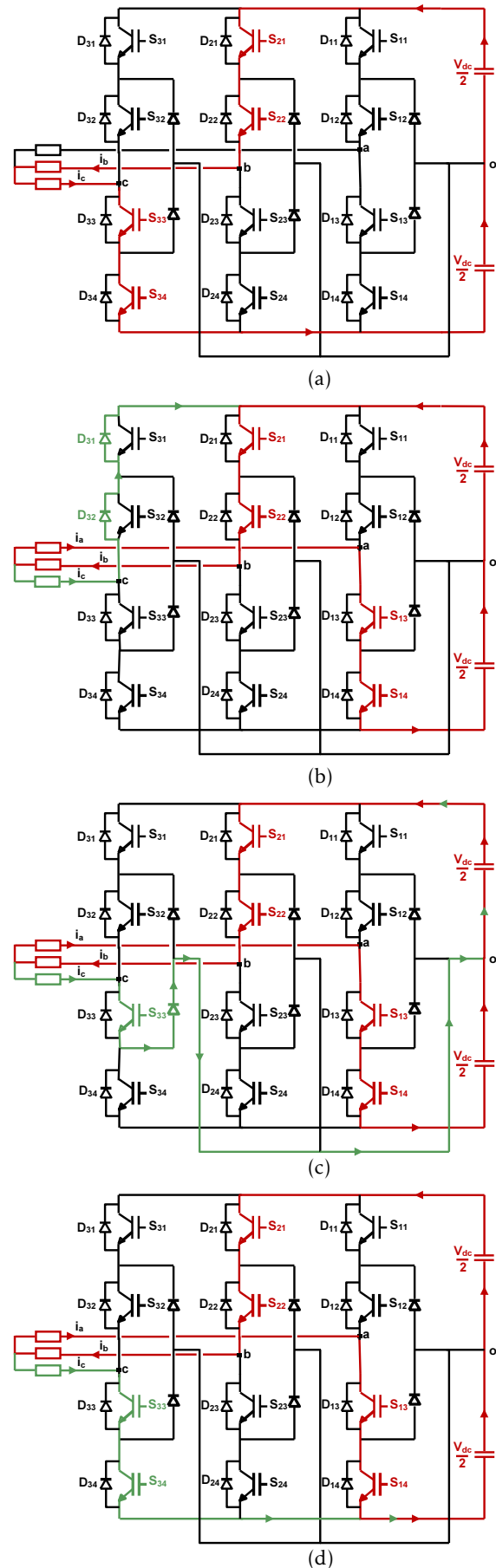
Figure 8: Current flowing during : **Legend** (a): sector I, (b): commutation under DTC-2 scheme, (c): commutation under DTC-3 scheme($\Omega^* \leq \Omega_{limit}$) and (d): commutation under DTC-3 scheme ($\Omega^* > \Omega_{limit}$).

Table 7: The vector selection table of the DTC-3 strategy.

c_τ	+1	−1	c_τ	+2	
				$\Omega^* \leq \Omega_{limit}$	$\Omega^* > \Omega_{limit}$
Sector I	$\mathbf{V}_8(000011000011)$	$\mathbf{V}_2(000000000011)$	Sector VI → Sector I	$\mathbf{U}_2(010011000011)$	$\mathbf{U}_8(110011000011)$
Sector II	$\mathbf{V}_9(001111000000)$	$\mathbf{V}_3(000011000000)$	Sector I → Sector II	$\mathbf{U}_3(001111000010)$	$\mathbf{U}_9(001111000011)$
Sector III	$\mathbf{V}_{10}(001100000011)$	$\mathbf{V}_4(001100000000)$	Sector II → Sector III	$\mathbf{U}_4(001101000011)$	$\mathbf{U}_{10}(001111001100)$
Sector IV	$\mathbf{V}_{11}(000000111100)$	$\mathbf{V}_5(000000001100)$	Sector III → Sector IV	$\mathbf{U}_5(001000111100)$	$\mathbf{U}_{11}(001100111100)$
Sector V	$\mathbf{V}_{12}(110000110000)$	$\mathbf{V}_6(000000110000)$	Sector IV → Sector V	$\mathbf{U}_6(110000110100)$	$\mathbf{U}_{12}(110000111100)$
Sector VI	$\mathbf{V}_7(110000000011)$	$\mathbf{V}_1(110000000000)$	Sector V → Sector VI	$\mathbf{U}_1(110000100011)$	$\mathbf{U}_7(110000110011)$
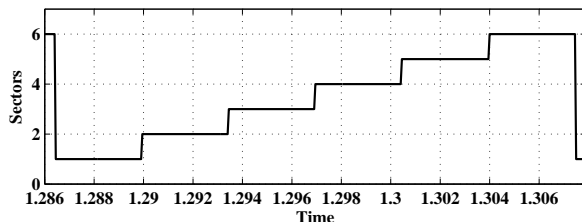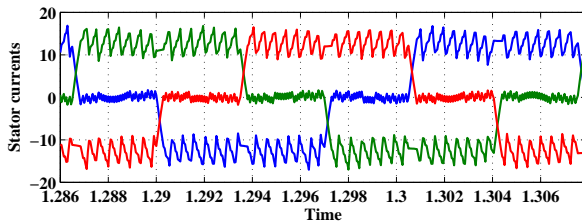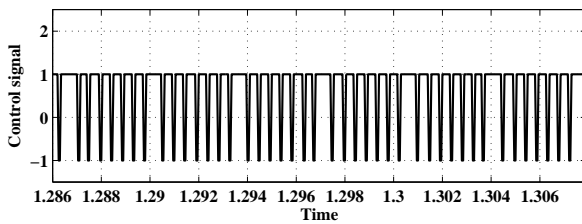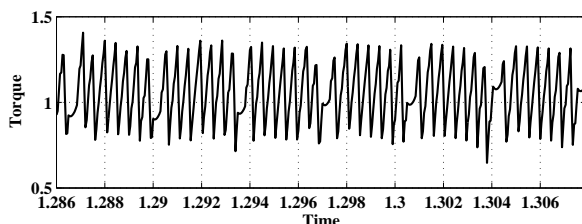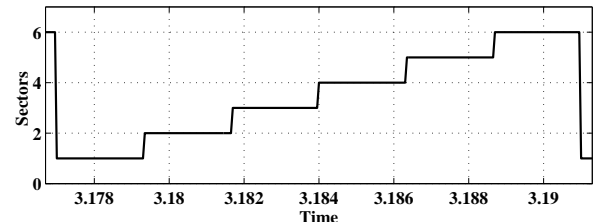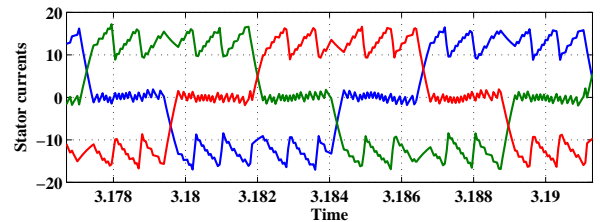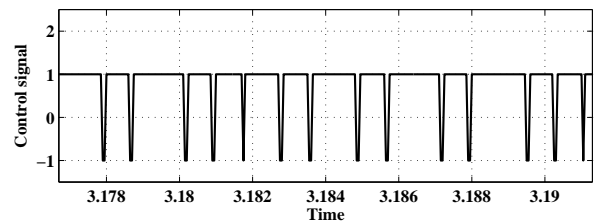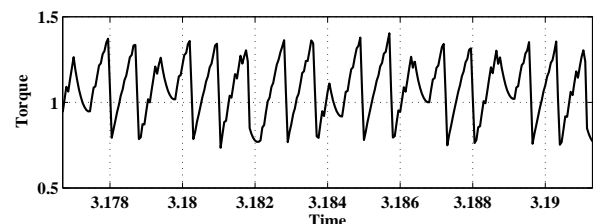
Figure 9: Simulated steady-state variables provided by the classic DTC-1 scheme at $\Omega = +100$rad/s (subscript "1") and $\Omega = +150$rad/s (subscript "2"). **Legend** (a): sector succession, (b): stator currents, (c): control signal c_τ and (d): electromagnetic torque.
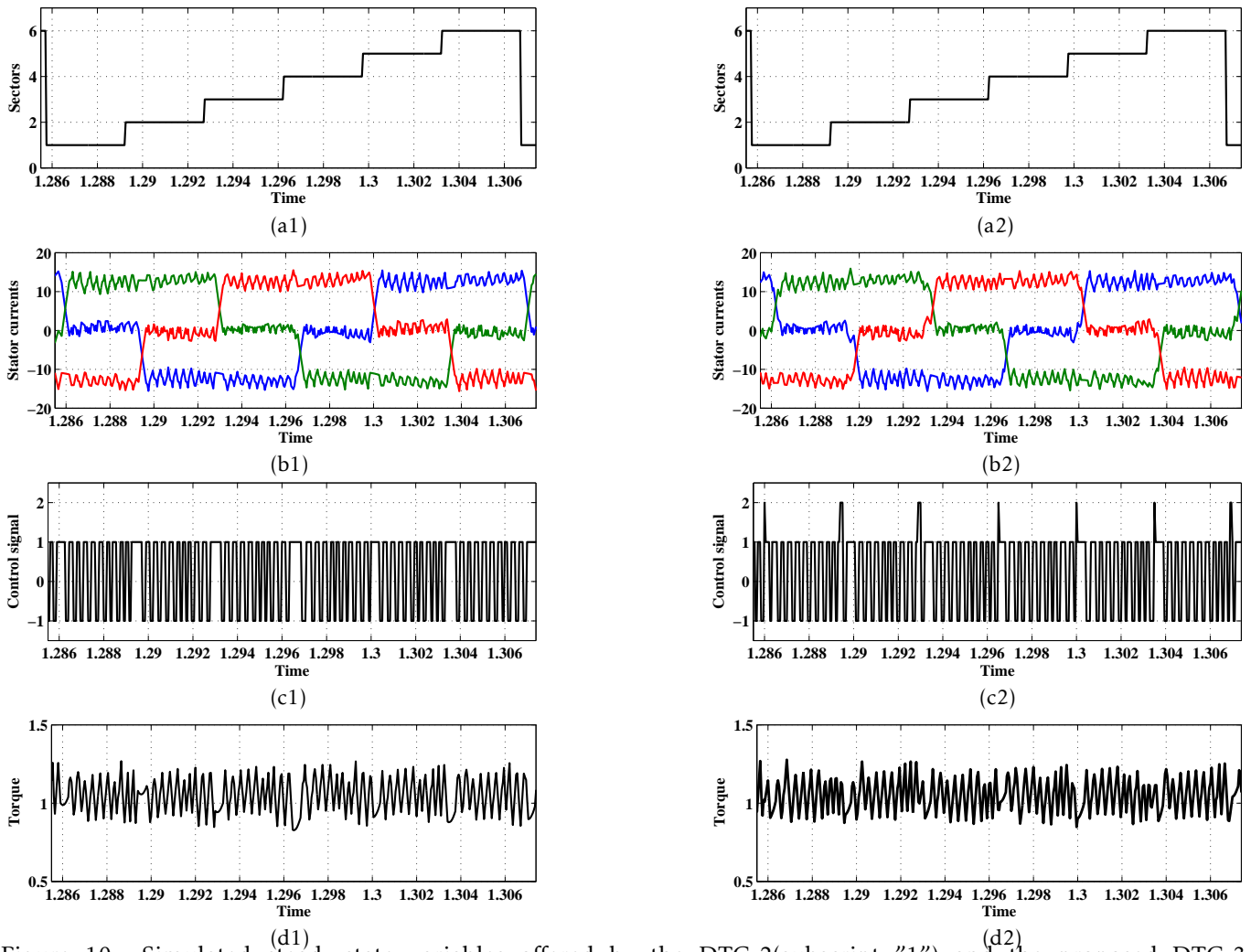
Figure 10: Simulated steady-state variables offered by the DTC-2(subscript "1") and the proposed DTC-3 (subscript "2") schemes at Ω = +100rad/s. **Legend** (a): sector succession, (b): stator currents, (c): control signal c_τ and (d): electromagnetic torque.
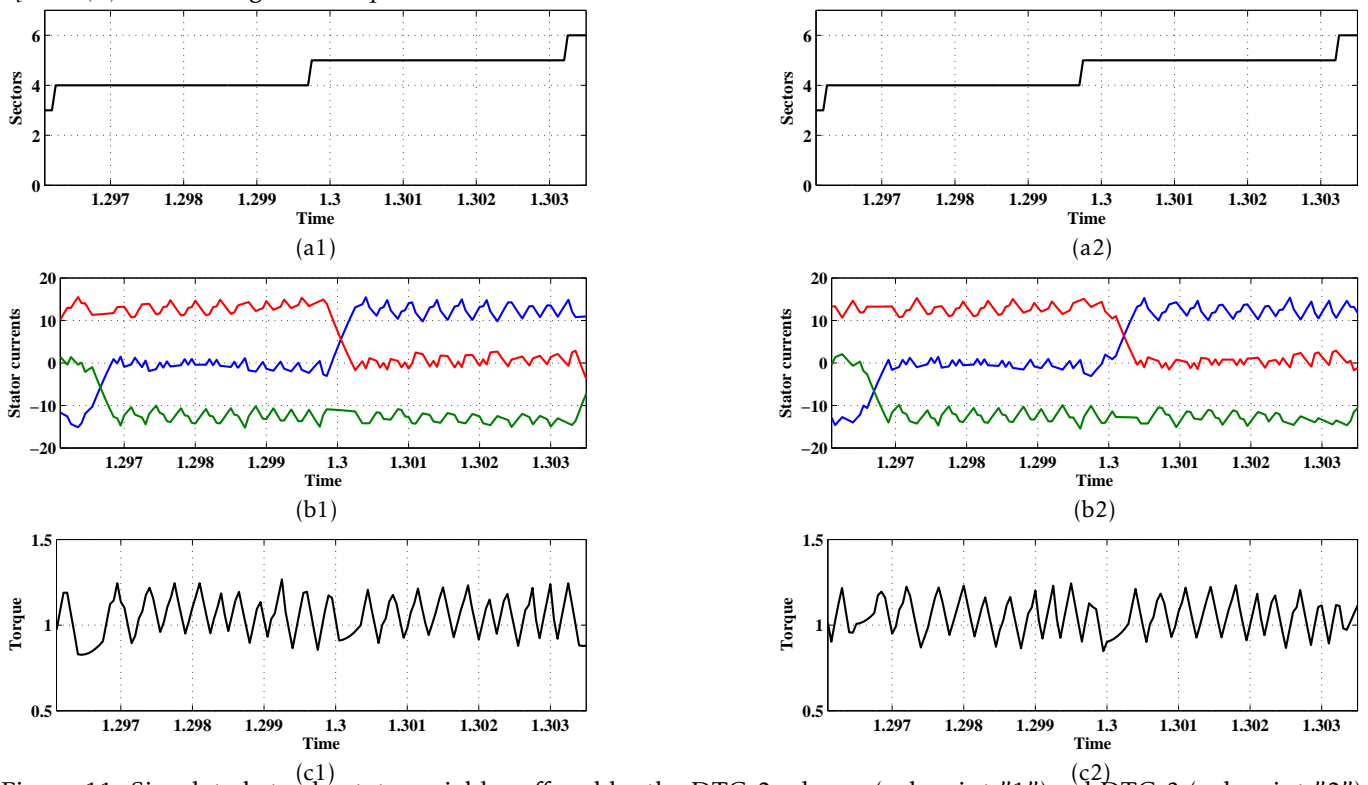
Figure 11: Simulated steady-state variables offered by the DTC-2 scheme (subscript "1")and DTC-3 (subscript "2") strategies for a speed Ω =+ 100rad/s. **Legend** (a): commutation from sector IV to sector V, (b): the stator currents zoom , and (c): the electromagnetic torque zoom.
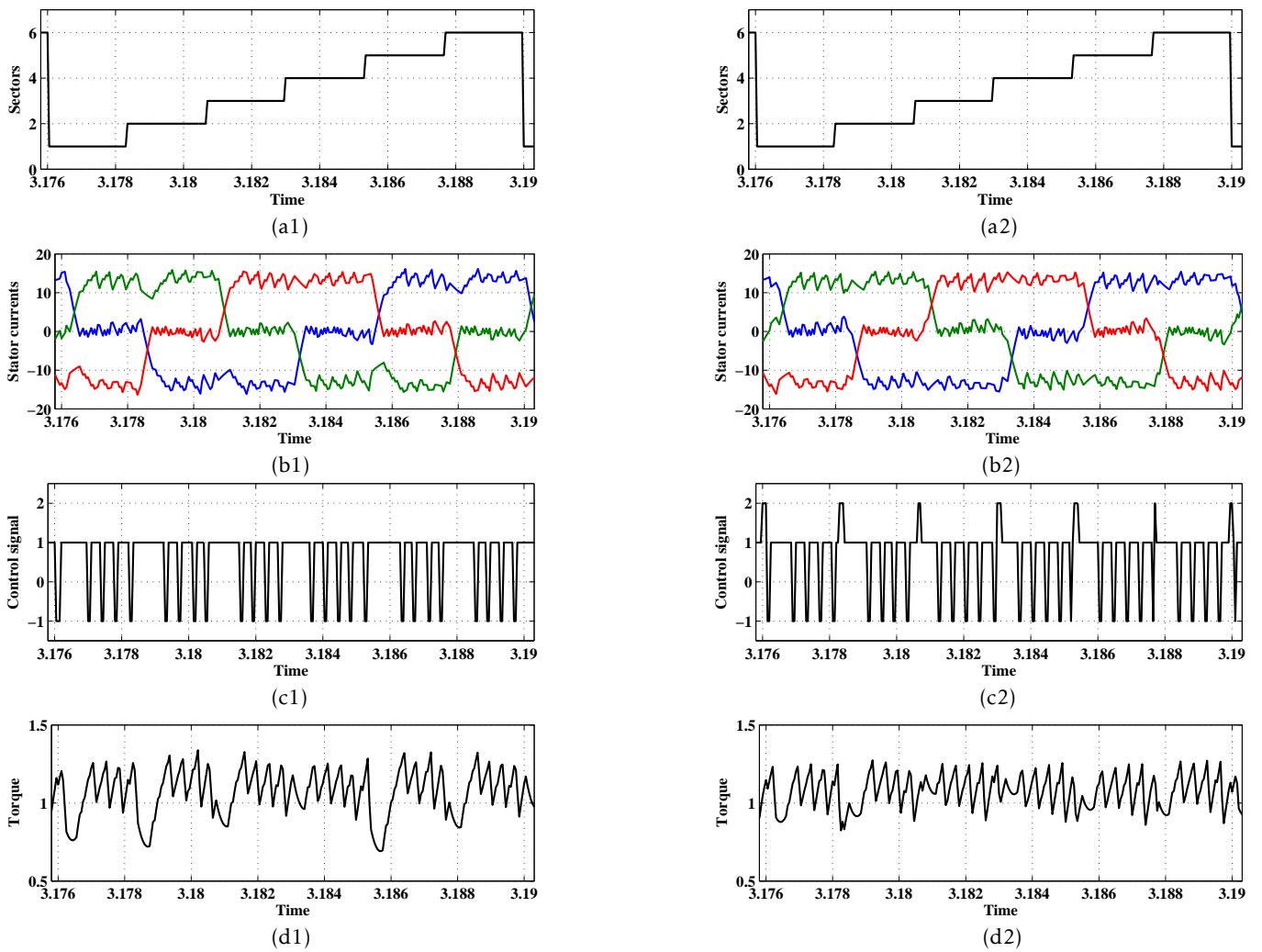
Figure 12: Simulated steady-state variables yielded by the DTC-2 (subscript "1") and the proposed DTC-3 strategies at Ω = +150rad/s. **Legend** (a): sector succession, (b): stator currents, (c): control signal c_τ and (d): electromagnetic torque.
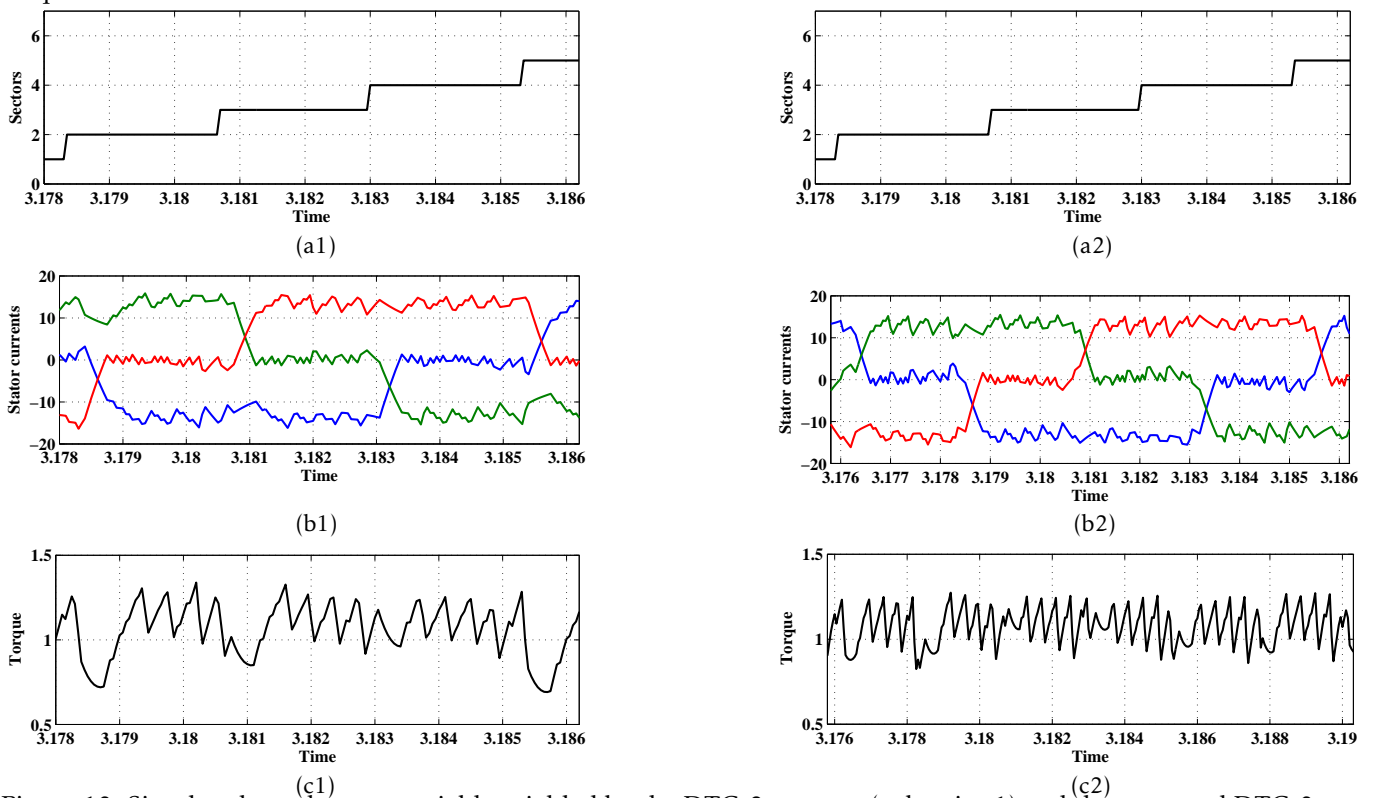
Figure 13: Simulated steady-state variables yielded by the DTC-2 strategy (subscript 1) and the proposed DTC-3 one for a speed Ω =+ 150rad/s . **Legend** (a): commutations from sector II to sector IV, (b): the stator currents zoom, and (c): the electromagnetic torque zoom .
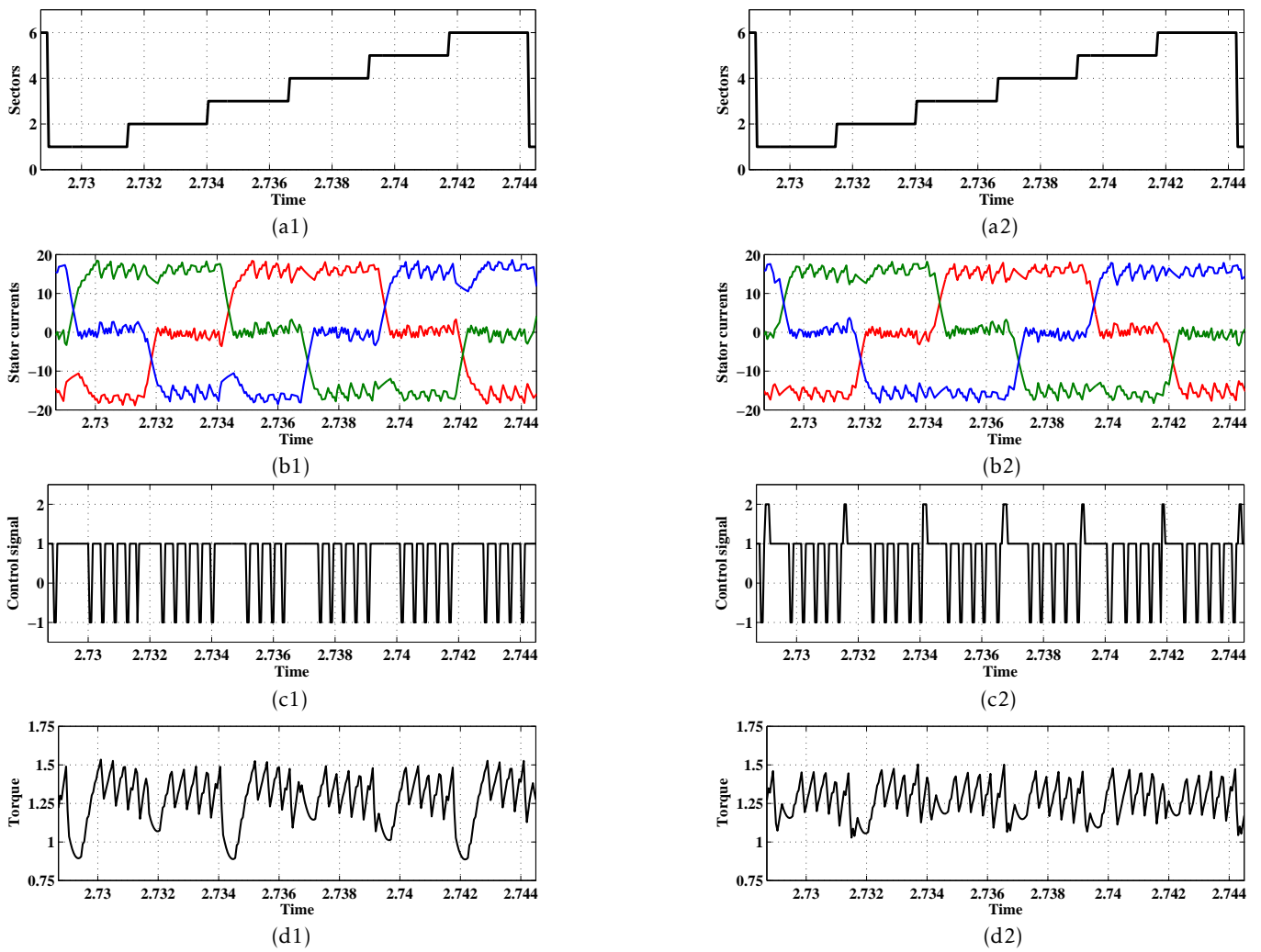
Figure 14: Simulated transient-state variables offered by the DTC-2(subscript "1") and the proposed DTC-3 schemes. **Legend** (a): sector succession, (b): stator currents, (c): control signal c_τ and (d): electromagnetic torque.
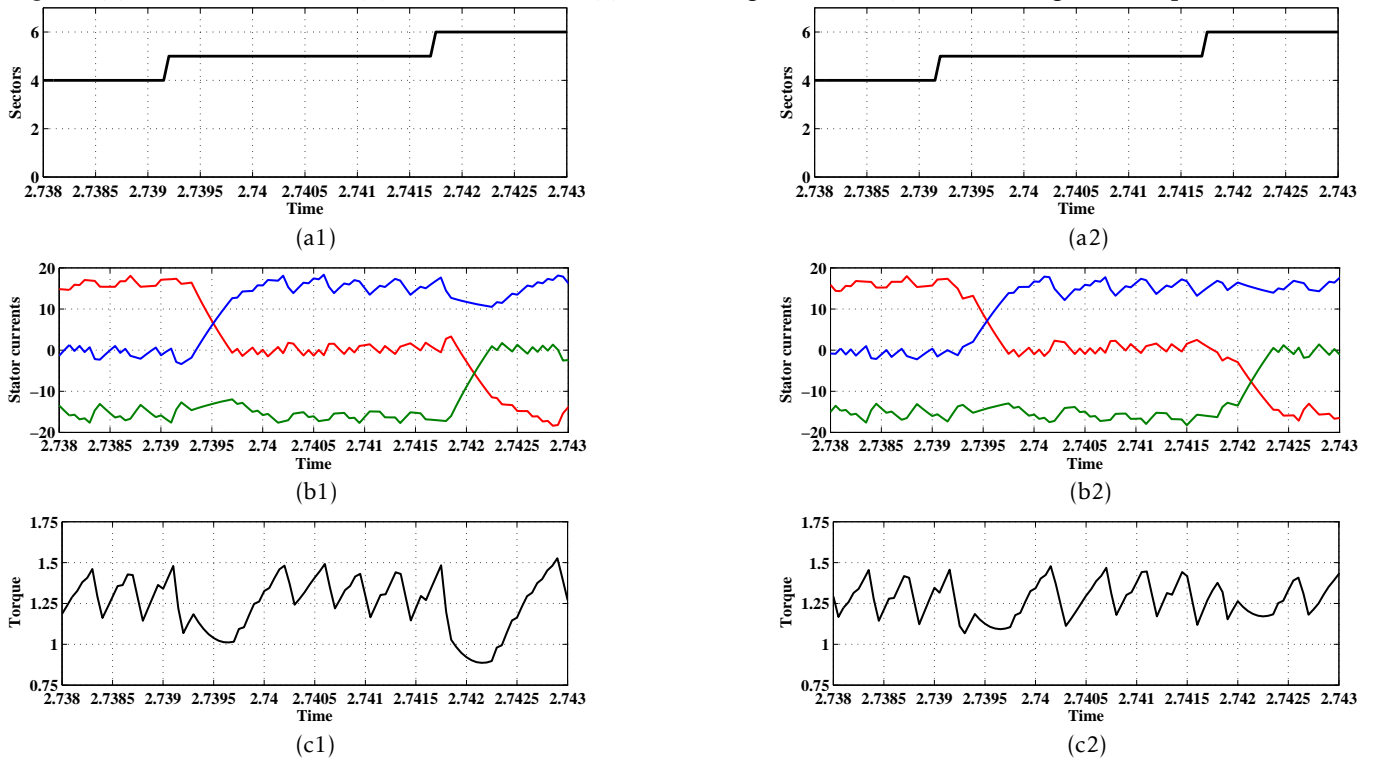
Figure 15: Simulated transient-state variables offered by the DTC-2 strategy (subscript 1) and the proposed DTC-3 one. **Legend** (a): commutations from sector V to sector VI, (b): the stator currents zoom , and (c): the electromagnetic torque zoom .
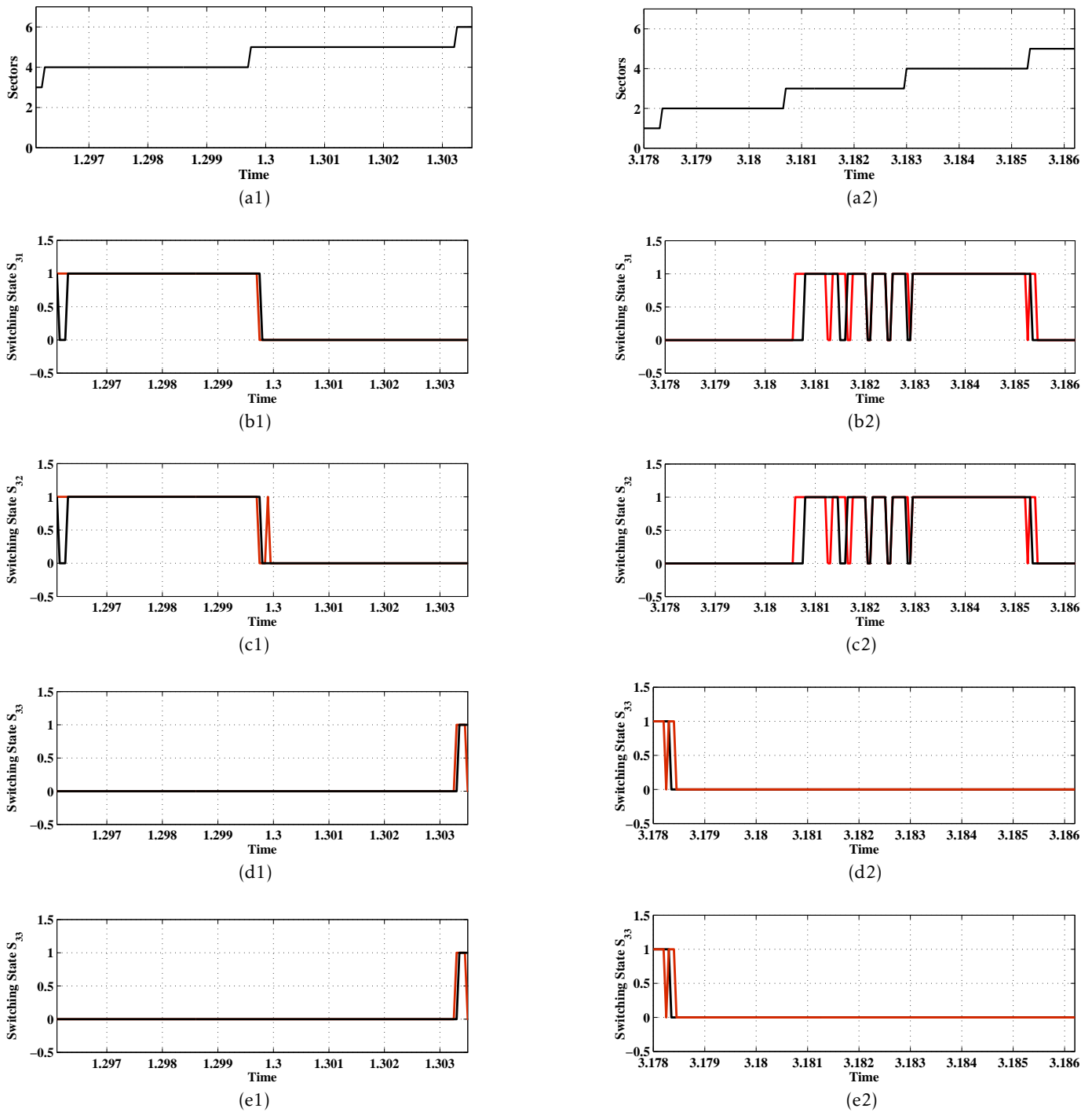
Figure 16: Simulated steady-state variables under the DTC-2 scheme (color black) and the proposed DTC-3 one (color red) for two reference speeds $\Omega =+ 100rad/s$ (subscript 1) and $\Omega =+ 150rad/s$ (subscript 2). **Legend** (a): commutations from sector II to sector IV, (b): S_{31}, (c): S_{32}, (d): S_{33}, and (e): S_{34} .

Moreover, simulation conditions characterizing the simulation study are listed below:

- a sampling period $T_s = 55\mu$s,

- a proportional gain $K_p = 1$,

- an integral gain $K_i = 0.5$,

- a load torque = 1Nm,

- a limit speed $\Omega_{limit} = 110rad/s$

- the hysteresis bounds are: $\varepsilon_{\tau 1} = \pm 0.0012$Nm and $\varepsilon_{\tau 2} = \pm 0.0036$.

To analyze the effectiveness of the proposed DTC-2 and DTC-3 strategies, this section deals with a simulation-based comparison between these strategies. By using MATLAB/SIMULINK, the proposed strategies are implemented. The obtained simulations results are illustrated in Figures. 10, 11, 12, 13, 14, 15.

4.1 Performance of the proposed strategies during sequence operation:

Figures.9, 10 and 12 show the waveforms of the current, the torque controller output c_τ and the torque waveforms at steady-state operation.

By comparing Figures.9(d), 10(d) and 12(d), one clearly notice that the DTC-1 scheme leads to an important torque dips thanks to the application of the Intermediate active voltage vectors. Both DTC-3 and DTC-2 strategies lead to a low torque dips during sequence operation thanks to the application of the Half active voltage vectors.

During sequence operation, both DTC-2 and DTC-3 schemes present almost the same torque waveforms. This is expected thanks to the application of the the same vectors. The difference concerns the torque decrease during commutations from sector to sector .

4.2 Performance of the proposed strategies during commutations from sector-to-sector :

To evaluate the proposed DTC schemes effectiveness during commutations from sector-to-sector , Figures. 11 and 13 give a zoom of the electromagnetic torque and the stator currents at steady-state operation. Figure.16 illustrates the steady-state waveforms of the switching states S_{31}, S_{32}, S_{33} and S_{34} with the sector succession.

As shown in Figures. 11(b1) and 13(b1), during the commutation from sector IV to sector V, the unbalancing between the current rising time and the current falling time affected the electromagnetic torque waveform. As a result of this, the DTC-2 strategy is penalized by a high torque dips. The comparison of Figures.10(c1) and 12(c1) confirms that commutation torque dips is more serious and more strong at high speed.

Compared to DTC-2, this commutation phenomenon is resolved for the DTC-3 strategy. Therefore, during the commutation from sector IV to sector V and in the case of the reference speed $\Omega^* = 150rad/s$ ($\Omega^* > \Omega_{limit}$), the current i_c is forced to flow through the IGBTS (S_{31}, S_{32}) and the supply source V_{dc} instead of the freewheeling diodes (D_{33} and D_{34}) as shown in Figures.16(b1), (b2), (b3) and (b4). However, from the analysis of Figures.16(a1), (a2), (a3) and (a4) in the case of a reference speed $\Omega^* = 100rad/s$ ($\Omega^* \leq \Omega_{limit}$), it can be clearly noticed that the current i_c is forced to flow through the IGBT S_{32}, the clamped diode and the supply source $\frac{V_{dc}}{2}$ instead of the freewheeling diodes (D_{33} and D_{34}). Consequently, $|di_a/dt|$ and $|di_c/dt|$ are equal for both reference speeds which results a high dynamic of the torque without commutation ripple. These benefits are confirmed in Figures. 11(a2), 11(a3), 13(b2) and 13(b3).

The performance of DTC strategies under a transient-state is presented in Figures.14 and 15. Accordingly, compared to DTC-2 and DTC-1 schemes, one can evidently affirm that the developed DTC-3 scheme leads to the lower commutation torque dips for both transient and steady-state operations.

5 Conclusion

BLDC motor is integrated in various applications due to its different advantages. However, the BLDC motor drive has the drawback of high torque ripple which affects the performance of this system. Thus, torque ripple minimization is a crucial necessity. This paper has been focused on the development of two DTC schemes dedicated to BLDC motor drives. Simulation work has been accomplished to highlight the developed schemes effectiveness and performances. It has been found that the developed DTC-3 scheme has guaranteed the less distortion in the line currents and less commutation torque dips. This developed strategy has been based on the compensation of the difference between the slopes of the incoming and outgoing phases currents.

References

[1] J. Rodriguez, J.-S. Lai, and F. Z. ”Peng, Multilevel inverters: A survey of topologies controls and applications”, *IEEE Trans. Ind. Electron.*, vol. 49, no. 4, pp. 724-738, Aug. 2002.

[2] Y. Liu, A. Q. Huang, W. Song, S. Bhattacharya, and G. Tan, ”Smallsignal model-based control strategy for balancing individual dc capacitor voltages in cascade multilevel inverter-based statcom, Industrial Electronics”,*IEEE Transactions on*, vol. 56, no. 6, pp. 22592269, 2009

[3] H. Vahedi,K. Al-Haddad, P.-A. Labbe, and S. Rahmani, ”Cascaded multilevel inverter with multicarrier pwm technique and voltage balancing feature”, *in Industrial Electronics (ISIE)*, 2014 IEEE 23rd International Symposium on. IEEE, 2014, pp. 21552160

[4] J. P. Lyons, V. Vlatkovic, P. M. Espelange, A. A. M. Esser, and F. F. Want, ”Five level high power motor drive converter and control system”, *U.S. Patent 6 058 031*, May 2, 2000.

[5] S. Bernet, T. Bruckner, and P. Stiemer, Three-point converter and method for its operation, U.S. Patent 6 219 265, Apr. 17, 2001.

[6] M. A. Doss, E. Premkumar, G. R. Kumar, and J. Hussain, "Harmonics and torque ripple reduction of Brushless DC Motor (BLDCM) using cascaded H-bridge multilevel inverter" , *in Proc. IEEE ICPEC*, Feb. 2013, pp. 296299.

[7] Chan, C. C., Jiang, J. Z., Xia, W., Chan, K. T. "Novel wide range speed control of permanent magnet brushless motor drives". *IEEE transactions on power electronics*, 10(5), 539-546,1995

[8] T. Shi, Y. Guo, P. Song, and C. Xia, "A new approach of minimizing commutation torque ripple for brushless DC motor based on DC-DC converter", *IEEE Trans. Ind. Electron.*, vol. 57, no. 10, pp. 3483-3490, Oct. 2010.

[9] K. Ohishi, M. Nakao, K. Ohnishi, and K. Miyachi, "Microprocessor-controlled DC motor for load-insensitive position servo system", *IEEE Trans. Ind. Electron.*, vol. 34, no. 1, pp. 44-49, 1987.

[10] X. Huang, A. Goodman, C. Gerada, Y. Fang, and Q. Lu, "A single sided matrix converter drive for a brushless dc motor in aerospace applications", *IEEE Trans. Ind. Electron.*, vol. 59, no. 9, pp. 3542-3552, Sep. 2012.

[11] T. M. Jahns and W. I. Soong, "Pulsating torque minimization techniques for permanent magnet AC motor drives-a review", *IEEE Trans. Ind. Electron.*, vol. 43, no. 2, pp. 321-330, Apr. 1996.

[12] X. Li, C. Xia, Y. Cao, W. Chen, and T. Shi, "Commutation torque ripple reduction strategy of Z-source inverter fed brushless DC motor", *IEEE Trans. Power Electron.*, vol. 31, no. 11, pp. 7677-7690, Nov. 2016.

[13] C. Xia, Y. Xiao, W. Chen, and T. Shi, "Torque ripple reduction in brushless DC drives based on reference current optimization using integral variable structure control", *IEEE Trans. Ind. Electron.*, vol. 61, no. 2, pp. 738-752, Feb. 2014.

[14] G. Buja, M. Bertoluzzo, and R. K. Keshri, "Torque ripple-free operation of PM BLDC drives with petal-wave current supply", *IEEE Trans. Ind. Electron.*, vol. 62, no. 7, pp. 4034-4043, Jul. 2015.

[15] J. Fang, H. Li, and B. Han. "Torque ripple reduction in BLDC torque motor with non-ideal back EMF", *IEEE Trans. Power Electron.*, vol. 27, no. 11, pp. 4630-4637, Nov. 2012.

[16] W. Jiang, Y. Liao, J. Wang, P. Wang, and Y. Xie, "Improved control of BLDCM considering commutation torque ripple and commutation time in full speed range", *IEEE Trans. Power Electron.*, vol. 33, no. 5,pp. 4249-4260, June 2018.

[17] I. Takahashi and T. Noguchi, "A new quick-response and high efficiency control strategy of an induction motor", *IEEE Trans. Ind. Appl.*, vol. 22, no. 5, pp. 820-827, 1986.

ASTES

Using Nullors to Modify Linear Model Parameters of Transistors in an Analog Circuit

Reza Hashemian[*]

Electrical Engineering, Northern Illinois University, DeKalb, IL 60115, USA

A B S T R A C T

Usually, after properly biasing an analog circuit we need to linearize its transistors in order to linearize the entire circuit for performance analysis and design. This linearization is done through modeling of devices, which is becoming more and more complicated as the electronic technology advances. This is not of course a major issue for circuit simulators that use the latest device models and fast computers to deal with them. However, it makes it difficult when manual computation and understanding of the circuit behavior is in question. This is particularly important in teaching and training electrical engineering students.
What the proposed method offers is to simplify the case in two steps. First, adopt a most recent model with nominal parameter values for the transistors and run the circuit. Second, compare the circuit responses with those obtained through simulators with the most advanced device models. Next, with the help of one or more nullors try to modify the manually selected model parameters so that the two responses become close enough together. Several examples demonstrate the way the technique works and how close we can get to the simulated responses.

1. Introduction

Linearizing analog circuits is a major step in circuit analysis and simulation, preparing for performance analysis and design. However, it is also important to understand the circuit behavior for design purposes. This is particularly essential in teaching students/designers in the field of electrical engineering. With so fast changing electronic technology and the transistors' modeling techniques, this linearization process becomes very cumbersome when it comes to hand calculations and manual circuit verifications and evaluations [1-3].

Several strategies can be proposed. A simple and effective one is to apply external probing such as experimenting with the transistors in different circuits and study their behavior, and then start building a working model. Alternatively, we can manually linearize the transistors to our best knowledge, and with an available and appropriate model that we think is working. Then we can compare responses from this circuit with those we get from an advances circuit simulator. This definitely will guide us to modify our model for a better accuracy. Here, in this study we

are trying to focus on this proposed methodology. It is interesting to note that the proposed technique takes a full advantage of the computational power and the precision expected from a modern circuit simulator, in order to direct the manually modeling procedure to achieve its goals.

Here is how the method works. Let us first assume a circuit that has only one transistor. We choose a linear small signal equivalent circuit model for the transistor in the circuit that best represents the device modeling. This apparently linearizes the entire circuit and prepares it for the AC analysis. Our next step is to simulate the circuit. However, to be certain about the direction we take for the modification of the model elements, this simulation is done in combination with the circuit with the original transistor. The combined circuits here are connected in parallel-parallel format. Other formats are also possible, such as series-series, parallel-series, and series-parallel, depending on the types of the input and output ports. In this connection both circuits get the same input signal and will produce identical outputs. Naturally, considering circuit laws, this identical outputs cannot happen unless we allow an extra degree of freedom in the linearized circuit. To fulfill this task one component in the linearized model is allowed to freely change its variables (v and i) values, which basically becomes a

[*]Reza Hashemian, Electrical Engineering, Northern Illinois University, DeKalb, IL 60115, USA. Tel: 815-753-9930, rhashemian@niu.edu

norator. The final step is to find a closest real component (or any two-terminal circuit) that best suits to replace the norator.

In short, there are two main steps involved here. In the first step the analog circuit is manually linearized after the biasing is completed and the regions of the operations for the transistors are specified. This will allow us to roughly assign linearized model parameters to the devices and hence make the entire circuit linear. In the second step this linear circuit is linked to the actual nonlinear circuit, in its original form, and the combined circuit is simulated with the constraint that both circuits produce the same responses, without interfering each other's operations. This process can of course be repeated for as many times as needed, depending on the precision required. In each process a different model component is selected for change, or alternatively, new components can be added to the model until we are satisfied.

Overall, using this procedure gets us closer and closer to the ideal response for the application we have. In exchange the price we need to pay is to modify one or more device model components in a manner in which the output response will improve to the ideal case even when we disconnect the two circuits from each other. This means, the linear circuit responds favorably after being separated from the original circuit. It must also be noted that this similarity in responses must only be valid for a certain region of operation, and for a desired bandwidth. That is, depending on the selections of the model components, the situation only needs to work for the application in hand and the frequency bandwidth described. Outside of this region of operation or bandwidth the response may or may not match with that of the original circuit.

The crucial components that are used for such operations are nullors, which are special types of fixator-norator pairs (FNP) [4-10]. Nullors are among pathological elements that although not ideally available in physical form but have very unique properties particularly for circuit designs [8, 9]. Briefly, a nullor consists of a pair of elements, a nullator and a norator. A nullator element passes zero current and has zero voltage across, whereas, a norator element has both its current and voltage unspecified. The difference between a nullor and an FNP is in the nullator element. In case the nullator in a nullor allows to have any arbitrary (but fixed) current or voltage then the nullor becomes an FNP. When used in pairs the nullator keeps both component variables (v and i) fixed, whereas both variables of the pairing norator are specified by the circuit. It is this unique property that makes a nullor well suited for design purposes. And here we exactly use the pair to adjust and modify the transistor model parameters in order to make the linearized circuit to perform as specified.

Finally, it must be pointed out that there is no claim this methodology can replace a formal and accurate transistor modeling, which is typically used by circuit simulators. The main purpose of using this technique is to educate electrical engineering students to manually experiment with device modeling for different applications. In addition, the approach simplifies and makes it more understandable to study circuit behavior such as finding poles and zeros and frequency behavior of the circuit. Although not still to its perfection, the method seems to be new and practical. As far as this author is aware of, there is no similar method reported in the literature to compare.

2. Modification of Model Parameters in a Linearized Circuit

The objective in this presentation is to linearize an analog circuit with transistors after being biased for a certain application. This is needed to perform further analysis such as AC analysis, transient analysis, and also for performance design [11]. A typical performance design procedure for amplifiers and analog filters is to design for frequency profile and bandwidth. To do this we need to first linearize the devices in the circuit and come up with a linear circuit. The next step is to perform AC analysis on the circuit using a circuit simulator. The procedure is done by first replacing the transistors in the circuit with their small signal linear models that are constructed from the device characterization and parameters extractions. Depending on several factors, such as the fabrication technology used, device sizes, and the transistors' modeling levels (in Spice) being adopted, these model parameters can become very complicated for manual considerations. They may even change whether we use them for analog design or digital. In general getting quite accurate model parameters mainly for short channel MOS transistors, for example, is not a trivial task. Figures 1 and 2 are examples of simpler linearized models for longer channel lengths MOS transistors.
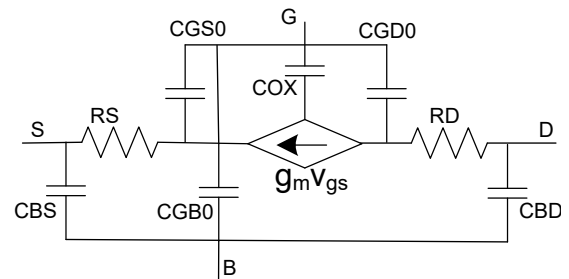
Figure 1. SPICE Level 1 Model for long channel MOS transistors.
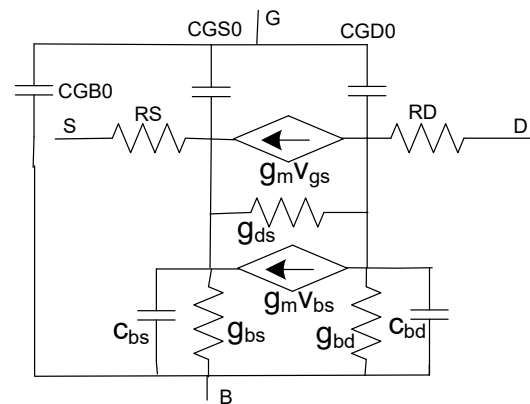
Figure2. SPICE Level 3 Model for long channel MOS transistors.

Figures 1 is the Level 1 model for long channel MOS transistors adopted by SPICE simulation program, and Fig. 2 is the schematic of a more advanced, Level 3, model with the code given below [12].

*Model parameters for long channel nMOS Level=3 models, supply VDD=5V
.MODEL N_1u NMOS LEVEL = 3

```
+ TOX   = 200E-10     NSUB  = 1E17       GAMMA = 0.5
+ PHI  = 0.7         VTO   = 0.8         DELTA = 3.0
+ UO   = 650         ETA   = 3.0E-6      THETA = 0.1
+ KP   = 120E-6      VMAX  = 1E5         KAPPA = 0.3
+ RSH  = 0           NFS   = 1E12        TPG   = 1
+ XJ   = 500E-9      LD    = 100E-9
+ CGDO = 200E-12     CGSO  = 200E-12     CGBO  = 1E-10
+ CJ   = 400E-6      PB    = 1           MJ    = 0.5
+ CJSW = 300E-12     MJSW  = 0.5
```

For short channel devices BSIM models are well advanced that are developed by Berkeley. In comparison with the lower level models, BSIM models are covering sub-micron and way down into nanotechnology cases. An example of 50nm BSIM4 models with V_{DD}=1V is a page long, and it is partially given below [12].

```
* Model parameters for 50nm nMOS BSIM4, supply VDD=1V.
.model N_50n nmos level = 54
+binunit = 1        paramchk= 1        mobmod = 0
+capmod = 2         igcmod = 1         igbmod = 1        geomod = 0
+diomod = 1         rdsmod = 0         rbodymod= 1        rgatemod= 1
+permod = 1         acnqsmod= 0        trnqsmod= 0
+tnom  = 27         toxe  = 1.4e-009   toxp  = 7e-010    toxm  = 1.4e-009
. . . .
. . . .
. . . .
+dmcg  = 0e-006     dmci  = 0e-006  dmdg = 0e-006  dmcgt = 0e-007
+dwj  = 0e-008      xgw   = 0e-007  xgl  = 0e-008
+rshg = 0.4         gbmin = 1e-010  rbpb = 5        rbpd = 15
+rbps = 15          rbdb = 15       rbsb = 15       ngcon = 1
```

As we can see, putting a model BSIM4 into schematic for manual calculation/evaluation is not practical and almost impossible. Although a circuit simulator, such as SPICE, does actually replace MOS transistors with such models for AC analysis, but doing so manually and for educational and training purposes is not simple, and even some of the model parameters cannot be translated into physical entities. So, the model must be simplified. One possibility is to do it for each case separately, and based on the circuit/device behavior. It may take a bit longer but it can be worth understanding the device behavior in that particular region of operation and for that application.
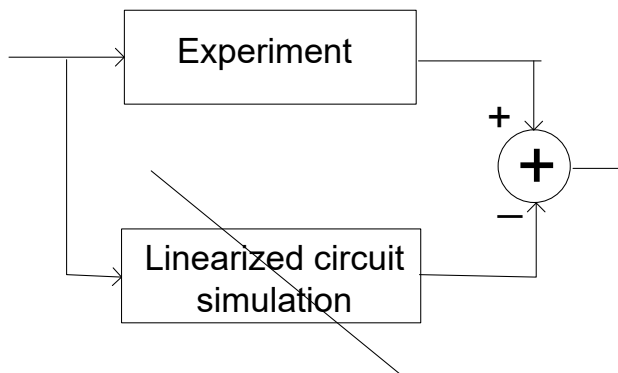
Figure3. Modification of the linearized transistor model parameters is directed by experiments or in comparison with the simulation of the original circuit.

In our proposed methodology the process is done in two steps. First, the circuit is manually linearized by replacing the transistors with their linear model parameters that are best available for the region of operation and the bandwidth specified. We next simulate the linearized circuit as well as the original analog circuit and then

compare the two results. If the responses happened to be close enough then we have hit the target, and the model parameters have been rightly selected. Otherwise, we go through the second step, in which we select one of the model elements in the circuit that seems most effective in changing the output response. Then we will proceed into a procedure that tries to modify the model element aiming at a response that mimics the response from the original circuit.

All we are trying to do in this methodology is to focus on this second step of the process, and find the right model component for the transistor. In case one selected model element does not finish the job to our satisfaction, we can still follow the same procedure with another model component. This is well illustrated in Fig. 3. As shown, the experimental results getting from a spectrum analyzers or a digital oscilloscope are compared with the results coming from simulating the linearized circuit. Alternatively, both the original circuit and the linearized one can be simulated in combination. Now, if the two responses are still too far apart the difference is fed back into the system to further modify the selected model component in the linear circuit. This is a sort of an ad hoc adaptation procedure. The question is, what is the purpose of manually linearizing an analog circuit when we don't have the exact model parameters for its transistors? Or what are the advantages of linearizing a circuit when we do not have access to the accurate model parameters? The answer to these questions mainly rely on the simplicity and the analytical power that exits with linear circuits. The experiments on nonlinear circuits are limited to the conditions and regions that the circuit operates in. After the biasing of the transistors in a nonlinear setting are completed we need to do performance (AC) design. We need to linearize the devices using model parameters that are valid for the selected regions of operations and the bandwidth we need.

So, circuit linearization shifts our circuit analysis and design from nonlinearity into linear domain, with all its properties and advantages. Even if a set of manually selected linearized model parameters are properly working for a certain operating condition then we can claim that we have reached to a regional situation, and in case the operating conditions of the transistors shift to a new region we can still try to modify those parameters again to reach to a solution for the new case. After all, we don't forget that this is just a simulation and the purpose is to explain the devices behavior in certain regions of operations, and in an educational setting.

Another point worth mentioning here is that, depending on the operational environment, it is most preferred to compare the results of a manually linearized circuit with experiments that are done with the real circuit. This is of course more easily done in case of a laboratory environment, but for a class or study room setting where only simulation facilities are available the case is different. Hence, for the latter case, it is more practical to have both original circuit, as the reference circuit, and the linearized circuit combined in one circuit and then simulate.

2.1. Linear model parameter identification

The type of arrangement just described is illustrated in Fig. 4 for a single model parameter change. Figure 4(a) is a symbolic representation of the linearized circuit along with a selected model element, denoted by P, to be modified. In Fig. 4 (b) the actual

circuit, denoted by *M*, is connected in parallel with the linearized circuit *N* ready for simulation. The responses of the circuits, i.e., the linearized *N* and the original circuit *M,* are then continuously compared. To keep the two responses identical, a nullator is used in the output port. Notice that the nullator, in addition of keeping the two output responses identical, it also provides independence for each circuit to operate without being interfered by the other. In exchange, it is the pairing norator that is responsible to provide the condition needed for such situation.
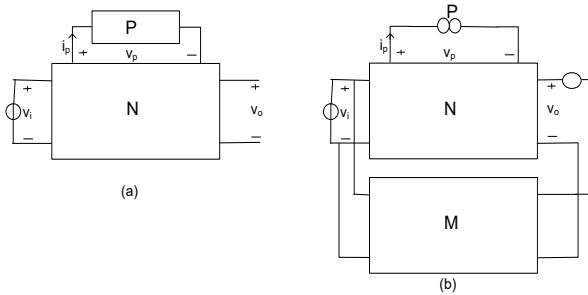
Figure4. (a) Manually linearized circuit *N* with a variable element *P* to be added; (b) nullor is added to be able to realize the element *P*.

Now, there are two cases that might happen. Either the v – i characteristic of the norator closely matches with that of the selected model component or they are way apart. For the former case the experiment is over and the model component remains unchanged for the linearized circuit.
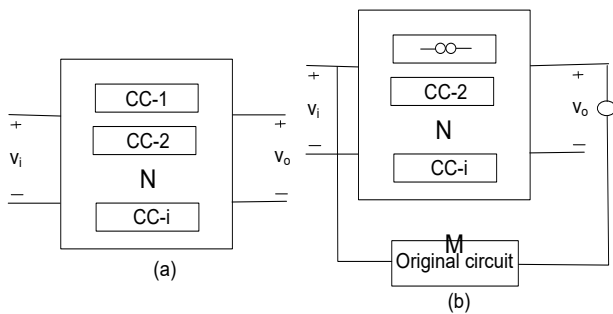
Figure 5. (a) Manually linearized circuit *N* with one or more possible sub-circuits *CC*; (b) nullor is added to keep the interconnection between the linear circuit and the original circuit ordered.

In the second case, we find that the model component and the norator do not match, which means we need to modify the model component so that its v – i characteristic is getting close enough to that of the norator. However, it is always possible that a single model parameter (component) may not completely solve the problem. In this situation we need to select another model component and continue with the same process, and so on. This situation is displayed in Fig. 5 with multiple Circuit Components (CCs) to be considered. Again, Fig. 5(a) shows the linearized circuit with *i* number of model elements, denoted by CC, that are candidates for changes. In Fig. 5 (b), on the other hand, the actual circuit, denoted by M, is added and connected in parallel with the linearized circuit. The rest of the testing procedure is the same as was explained previously. The difference, however, is that here we go through simulations multiple number of times and in each

time one of the CC components is replaced with a norator. This certainly continues improving the performance of the linearized circuit in reference to the actual circuit, in multiple steps until we get satisfied with the results.

Now we reach to a point that we can put the entire process into an algorithm for implementation.

Algorithm 1 – Consider a circuit *M* with transistors. Let *M* be manually linearized and call the resulted circuit *N,* as described earlier. We can then go through a stepwise procedure that modifies and corrects one or more model elements of the transistors in order to make the response of the linearized circuit to closely follow the response given by the original (nonlinear) circuit. The rest of the procedure is as follows:

1. Search for one or more transistor small signal model elements in *N* and call them CC_j, for all j, as shown in Figs. 4(a) and 5(a). Connect *N* and *M* circuits together through a nullator. For now make this connection parallel-parallel (for other possibilities see Footnote 1), as shown in Figs. 4(b) and 5(b).
2. Chose one of the model components *CCs*, say CC_j. This is a candidate to be modified for the design. Replace CC_j with the paring norator.
3. Perform a simulation on the combined circuit, *M* and *N*, for AC analysis and for the bandwidth that is specified.
4. Assume $v_j(s)$ and $i_j(s)$ to be the voltage across and the current through the norator CC_j, respectively. Then find the equivalent pseudo-impedance $z_j(s) = v_j(s)/i_j(s)$ of the norator, and plot its Bode plot p_j for the bandwidth given.
5. Next, try to replace the norator CC_j with an actual component, or a two-terminal circuit, that ideally has the same impedance $z_j(s) = v_j(s)/i_j(s)$. In case such an ideal case is not reached, find the closest component (or a two-terminal circuit) CC_j so that its Bode plot response p'_j is close enough to p_j, for the given bandwidth.
6. Replace the norator (CC_j) with the component (or the two-terminal circuit) CC_j found. Next, disconnect *M* from *N* circuit, and then simulate *N* alone for its response.
7. If more modifications are still needed, go to step 2 and select a different model component, say CC_k in *N* and follow the rest of the procedure. Continue with the procedure, as many times as needed, until an acceptable response is obtained.

3. Examples

Several examples are worked out here that illustrate the way the proposed method works. Example 1 assumes a circuit with a single nMOS transistor. Instead of using the exact model parameters, given in the data sheets, we simply use level 3 Spice model for long channel devices with values most appropriate. Next, we try to modify the parameter values until we get close enough to the desired response. To do this we use the FNP technique, described before, to adjust the model parameters within the bandwidth specified. Example 2 is given for a two stage nMOS amplifier with multiple transistors, where the type and size of the transistors are the same as those in Example 1. Therefore, we should be able to utilize the same transistor model parameters developed for Example 1. Example 3 is yet another nMOS transistor amplifier that utilizes FNP technique to modify the transistor model parameters. Here also we keep our previous level 3 transistor model used in Examples 1 and 2 unchanged. The point

of repeating the same model for all three cases is to see how expanding the circuits with more transistors but with the same transistor linear modeling affects the accuracy and the validity of the proposed method.

Example 1: – Take a single stage MOS amplifier, as shown in Fig. 6(a). Next, proceed to linearize the amplifier by using level 3 model parameters for its transistor. The assumption is that we

(a)

(d)
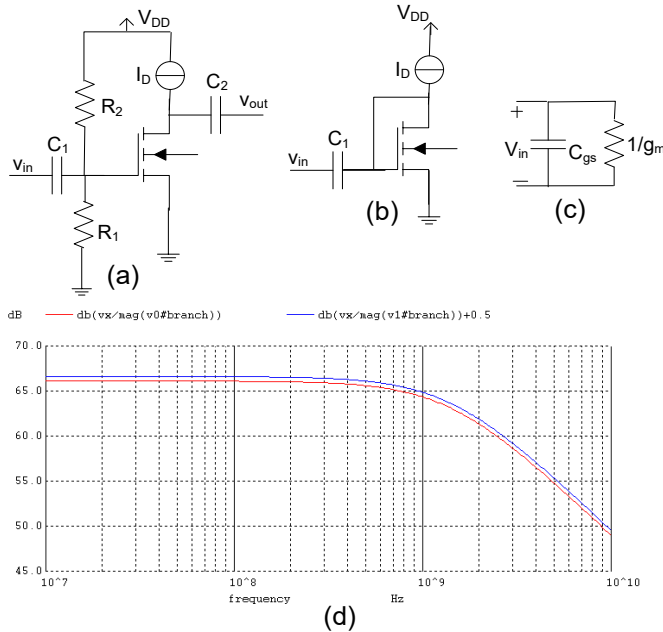
Figure 6. (a) An nMOS transistor amplifier; (b) diode connection; (c) equivalent diode connection in the linearized amplifier; (d) comparing plots for the input impedances in (b) and (c).

have no access to the Spice parameter values at this point, although the simulator is assumed to use the latest device model. To get the pseudo-parameters for our linearized model we go into two operational steps. First, we connect the transistor in a diode format (in the original circuit as well as in the linearized circuit) and then simulate the two circuits combined. Figures 6(b) and (c) are the two circuit schematics, and the Bode plots resulted from the two impedances, looking through the input ports, are demonstrated in Fig. 6(d). Note that for readability purposes one plot is artificially raised by 0.5 dB, otherwise the two are exactly the same.

Now, in the next step we use an FNP (similar to nullor) procedure to find the remaining parts of the model parameters for the transistor. To do that, we can simply use the actual amplifier (Fig. 6(a)) as the "model" circuit to help to adjust the model parameter in the linearized circuit, as given in Fig. 7(a). The Spice circuit code for the combined circuit is also given below, where a voltage controlled voltage source (e1) is used to represent an FNP.

```
.control
destroy all
set units = degrees
ac dec 1000 1.0e6 1.0e11
plot db(v(3)) db(v(6))+1
.endc
******** Combined circuits *******
Vin     2       0       DC      0       AC      1
x11     2       3       non-Lin
x21     6       2       0       Lin-Mos
```

```
*************** Circit 2 *********
.subckt Lin-Mos    7        6        0
gm1     7       0       6       0       0.495m
cgs1    6       0       55.8f
*VCVS e1 6      7       3       7       1.0e06
rgd1    6       7       2.2856MEG
cgd1    6       7       9.7f
.ends
**************** Ciruit 1 ******
.subckt non-Lin    3        7
VDD 2   0       DC      5
id1     2       5       20u
M11     5       4       0       0       N_1u L=1u W=50u
r11     4       0       198k
r21     2       4       1MEG
c11     3       4       10u
c21     5       7       10u
.ends
.include cmosedu_models.txt
.end
```


(a)   (b)

(c)

(d)
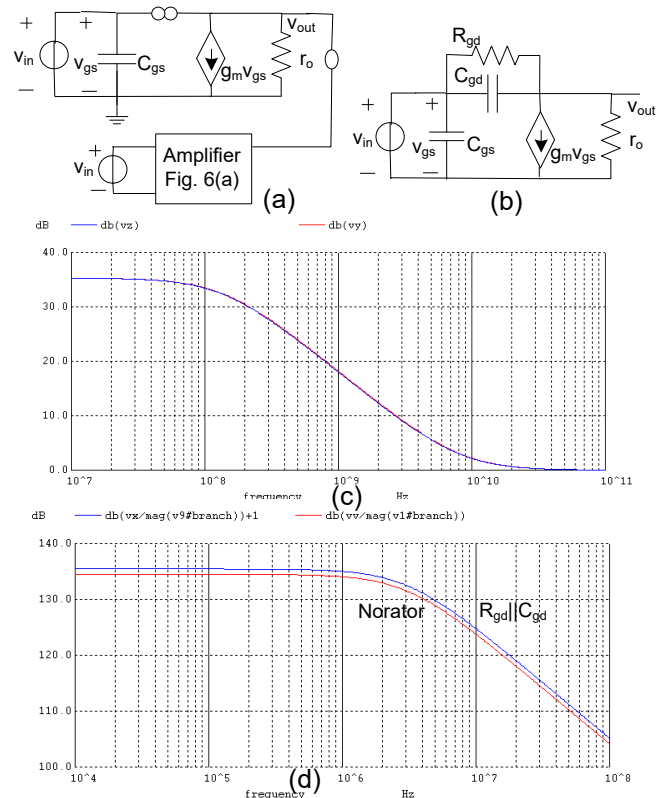
Fig.7. Getting the model parameters for the nMOS; (a) assumsed linearized circuit is parallel with the amplifier circuit; (b) nMOS linear model circuit; (c) gain responses for the actual amplifier and its linear equivalent circuit; (d) plots for the impedance functions, the norator and the replacement with the parallel R_{gd} and C_{gd} circuit.

In Fig. 7(c) we notice one gain plot from the actual amplifier and one from the linearized circuit. Obviously the two must be identical because a nullator enforce them to. Next, we need to replace the norator with one or more real component. To find this component we first run the circuit with the norator and make a plot of the impedance function of the norator. This plot is given in Fig. 7(d). The ideal solution is possible if we could find one or more combination of real components that represent the same impedance plot. Here we notice the norator plot is in fact very close to the impedance plot of a parallel RC circuit. This RC

circuit is represented by R_{gd} and C_{gd} in the linearized circuit given in Fig. 7(b), and the impedance function of the $R_{gd} \parallel C_{gd}$ is also plotted in Fig. 7(d), which is almost identical to the norator impedance (except for the 1dB clearance). This concludes our Example 1.

Example 2: – Consider a two stage nMOS shunt - shunt feedback amplifier, as shown in Fig. 8 [12]. There are four identical nMOS transistors used in this amplifier. In order to minimize our modeling efforts, we are going to use the same type and size transistor we used in Example 1. With this choice, to linearize the amplifier circuit all we need to do is to replace each transistor with its linear model already found in Example 1 with a bit of adjustment. This adjustment is done on the resistance r_{gd}, which is in parallel with c_{gd} in the transistor model. The value of r_{gd} in Example 1 was found to be close to 2.3 MEG ohms, whereas in Example 2 it was increased to 40 MEG ohms, where both resistances are quite large anyway. The list of the Spice program for this example is given below.
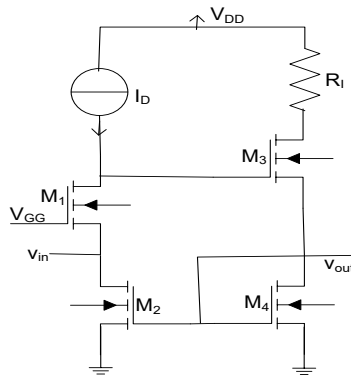
Figure 8. A two stage nMOS current amplifier.

```
.control
destroy all
set units = degrees
ac dec 1000 1.0e5 1.0e10
plot db(I(v21)) db(I(v31))
plot ph(I(v21)) ph(I(v31))
.endc
****** The combined circuit ******
ii1    0    2    DC    0    AC    1
x11    2    3    Mos-2
x21    2    4    M-Linear
v21    3    0    DC    0
v31    4    0    DC    0
****** Linear circuit ********
.subckt        M-Linear  8    9
x11    6    0    3    Lin-Mos
x21    3    4    0    Lin-Mos
x31    7    6    4    Lin-Mos
x41    4    4    0    Lin-Mos
r11    0    6    160k
r21    0    7    100k
c11    3    8    10u
c21    4    9    10u
.ends
************** Linear-Modeling ********
.subckt        Lin-Mos   7    6    3
gm1    7    3    6    3    0.52m
cgs1   6    3    55.8f
rgd1   6    7    40MEG
cgd1   6    7    9.7f
.ends
*************** The original Amplifier ********
```

```
.subckt        Mos-2    8    9
VDD1  2    0    DC    5
VG1   5    0    DC    1.770833
M11   6    5    3    0    N_1u L=1u W=50u
M21   3    4    0    0    N_1u L=1u W=50u
M31   7    6    4    0    N_1u L=1u W=50u
M41   4    4    0    0    N_1u L=1u W=50u
r11   2    6    160k
r21   2    7    100k
c11   3    8    10u
c21   4    9    10u
.ends
.include cmosedu_models.txt
.end
```
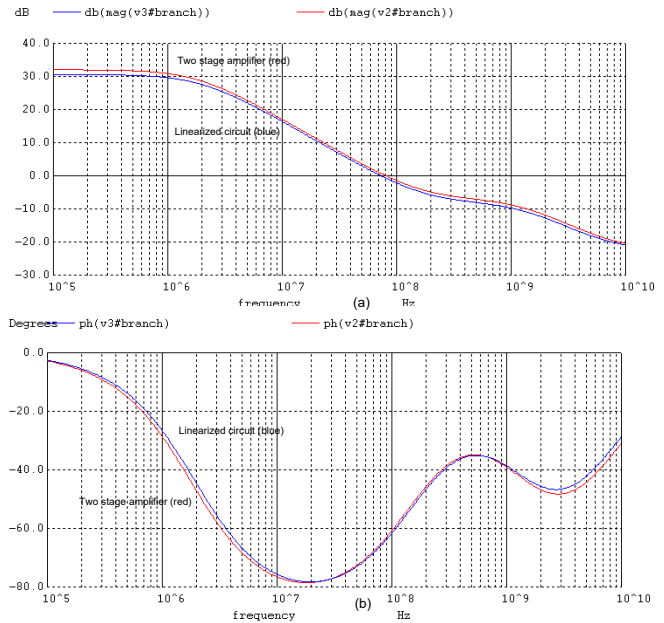

Figure 9. Bode plots given for the gains for the two stage nMOS amplifier and its linearized circuit; (a) magnitude plots; (b) phase plots.

Now it is time to simulate the amplifier and its linearized equivalent circuit in combination. The simulation generates the output transfer functions plots given in Figs. 9(a) and (b), where the plots in (a) represent the magnitude and in (b) the phase angle plots. Notice that the plots are almost identical for the original amplifier and its linearized circuit. Further experiments show that with the body effect removed in this example the two responses become even closer and tighter together.
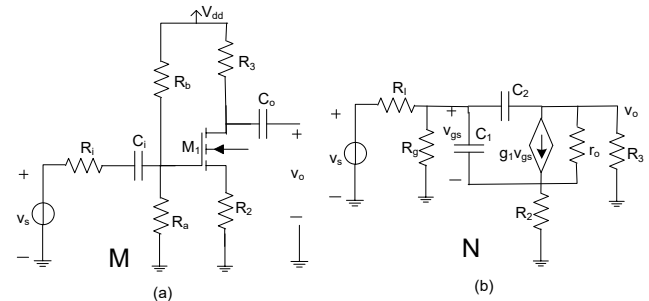
Figure 10. nMOS amplifier circuit and its linearized equivalence. (a) The original nMOS amplifier, and (b) its AC small signal linearized equivalent circuit.

Example 3: – Consider a single stage nMOS amplifier *M* with

feedback, as shown in Fig. 10(a). As usual, we first replace the transistor with its small signal model suitable for this case. This linearizes the amplifier and ready to compare its transfer function with the original amplifier. The model parameters used in this example are assumed to be the Spice Level 3 parameters provided for long channel nMOS transistors. Figure 10(b) shown the amplifier circuit after being linearized.

Our next move it to simulate both circuit, N and M, and plot the transfer functions. Figures 11 represents the plots for (a) the magnitude, and (b) the phase. In comparing the results from the two circuits we notice the differences, particularly in the magnitude plots; although the plots show the same patterns of variations vs the frequency and also in term of poles and zeros locations.

Now the question is what has gone wrong here, and how can this be corrected? In our efforts to find out about this discrepancy we realize that the ratio between the values of the two transistor model capacitors, i. e., $C_{gs} = 2.05\ pF$ and $C_{gd} = 13.83\ fF$, is very high and close to 150. So, C_{gs} seems to be too high and must be modified. We pick up C_{gs} and replace it with a norator, and go through Algorithm 1 for the adjustments.
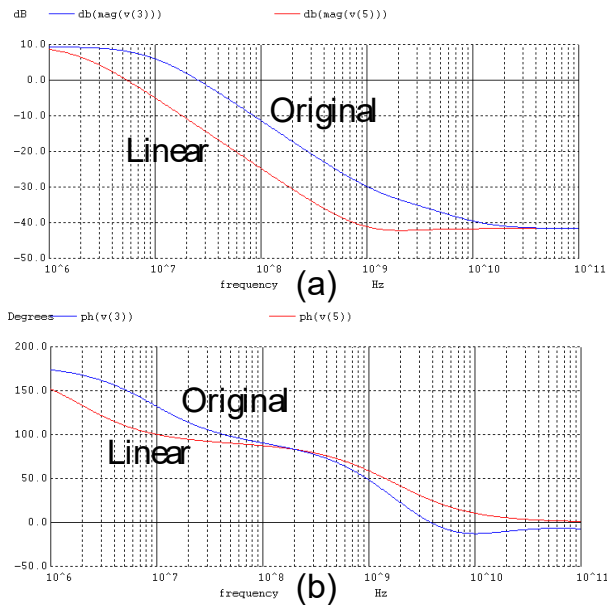
Figure 11. Transfer function responses given for the original amplifier circuit and it linearized equivalent circuit. The differences shown in the responses are mainly the results of mismatch in model parameters. (a) The magnitude, and (b) the phase.

As stated in Algorithm 1, step 2, we are going to make a parallel-parallel connection between the original circuit, *M,* and the linearized one, *N,* through a nullor (this connection is the same as that in Fig. 7(a)) and is not shown here. For the norator we choose a current controlled voltage source (CCVS). Next, we use Spice simulation tool to simulate the combined circuit. The partial Spice programming code for this example is given as.

```
.control
destroy all
let vm1 = v(5)-v(7)
plot db(v(6)) db(v(4))
plot db((vm1)/I(v11)) db((v(8))/I(v31))
```

```
.endc
.option scale = 1u
.ac dec 1000 1Meg 100G
VDD1   2        0            DC        5
Vs1    3        0            DC        0         AC        1
*************** The actual amplifier ****************
.subckt         model-1    2         3         6
M11    8        5            7         0         NMOS L=2 W=100
ra1    5        0            200k
rb1    2        5            330k
ri1    5        3            100k
r31    2        8            5.1k
r21    7        0            1k
ci1    5        4            10u
co1    8        6            10u
.ends
************** The linearized circuit ***************
ri1    3        5            100k
r21    7        0            1k
rg1    5        0            82k
r31    8        0            5.1k
ro1    8        7            180k
c11    5        7            2.05p
c21    5        8            13.83f
g11    8        7            5         7         1.648m
*********** Circuits combined ***************
x11    2        3            4         model-1
v01    7        4            DC        0
v11    5        b            DC        0
h11    b        7            v0        1.0e8
* *********** Testing Bench **************
v31    3        9            DC        0
rx1    9        a            190
c31    a        0            300f
.include cmosedu_models.txt
.end
```

We notice from the simulation results that in order for the linearized circuit to respond exactly similar to that of the original amplifier is the followings. We must replace the norator with one or more components that in combination represent an impedance function with the characteristic plot given in Fig. 12(a), plot *CC*. Now, with a close look at this plot we come to a conclusion that this plot cannot be just from a capacitor C_{gs} alone. There must be a series combination of a capacitor and a resistor. This is because we see at least one zero on the real axis in the LHP. This *RC* circuit is shown in Fig. 12 (b), where we find $C = 300\ fF$ and $R = 190\ \Omega$. Next, when we plot the impedance function of the RC circuit we come up with the plot *RC* given in Fig 12(a). As noticed, the two plots, RC and CC, are now so close together, which are almost undistinguishable from each other. Our next move is to replace the norator in N with the RC circuit just found.
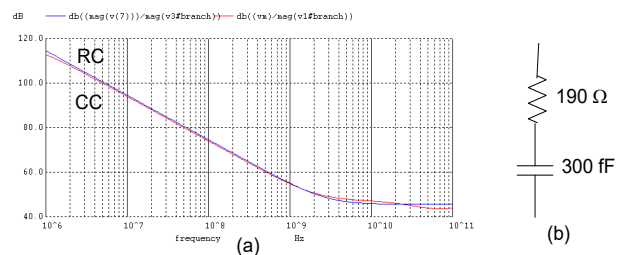
Figure 12. (a) Comparing Bode plots associated with two impedances; one representing the norator (CC), and the other one the RC circuit. (b) The series RC circuit.

After we have made the substitution (i. e., the original model element $C_{gs} = 2.05\ pF$, is replaced by the components $C = 300\ fF$

and $R = 190\ \Omega$ in series) we simulate the circuits, this time disconnected from each other, except for the source signal. The frequency responses of both circuits are now plotted in Fig. 13, (a) for magnitude, and (b) for the phase. Comparing the plots in Figs. 13 (a) and (b) with those in Figs. 11 (a) and (b) we see drastic changes. There are much more improvements, and this is particularly evident from the magnitude plots given in Fig. 13(a).

It is of course possible to go one step further and pick up another component in the transistor model for change. The process will be the same as explained before, i. e., replacing the component with a norator first and then follow Algorithm 1 for further processing. The result is expected to get the linearized circuit response closer to that of the original amplifier. However, we think the error is negligible at this point and there is no need to go further.
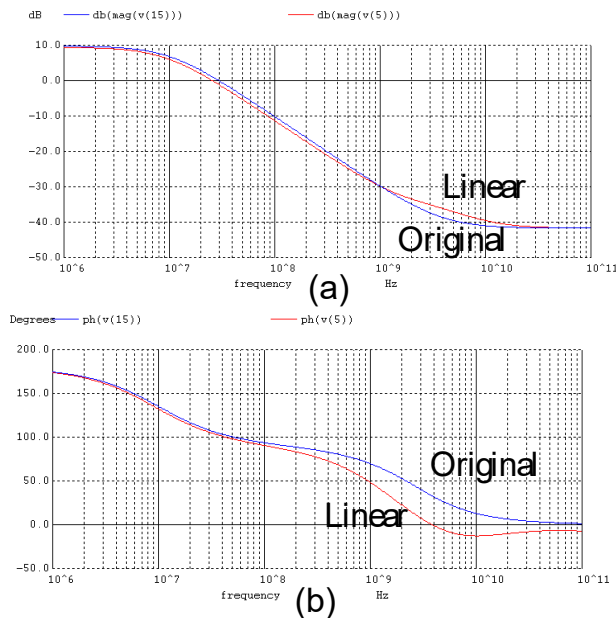
Figure 13. Comparing two transfer functions plots, one associated to the original nonlinear amplifier, and the other one to the linearized equivalent circuit; (a) magnitude, and (b) phase responses.

4. Conclusion

A simple and practical techniques is presented to manually linearize an analog circuit. The method starts with linearizing the transistors based of simple and available models but it gradually modifies one or more model components until an acceptable result is reached.

The method simplifies the case in two steps. First, adopt a most recent model with nominal parameter values for the transistors and simulate the circuit. Second, compare the circuit responses with those getting from the original circuits when simulated or in experiments. Then with the help of one or more nullors modify the manually selected model parameters so that the two responses become close enough for the application. Several examples demonstrate the way the technique works.

The method is especially useful for teaching electronic circuits to students and trainees, where understanding of how linearized modeling of devices works is of prime interest. It is shown that this modeling procedure is completely external to the device and works based on the circuit application and for an assigned

bandwidth.

In this procedure the circuit is first manually linearized, which means replacing the transistor with a linear model that is available to the designer. The next step is to combine the linearized circuit with the original circuit in such a way that they behave independently without any interfering each other, but at the same time their responses remains identical. From circuit theory stand point, this is only possible if we allow a component (norator) in the linearized circuit to have varying (v and i) variables. The next step in the process is to replace the norator with a real component or sub-circuit in the transistor model. Evidently, this may not always work quite accurately due to the realizability of the component. In case the result obtained after this substitution is still unacceptable the process can go on including other components of the linearize model. That is, the process of modifying and correcting the manual modeling of the device can continue until the final response we receive is working and acceptable.

References

[1] T.L. Pillage, R.A. Rohrer, C. Visweswariah, Electronic Circuit & System Simulation Methods, McGraw-Hill 1995.

[2] A. S. Sedra, and K. C. Smith, Microelectronic Circuits, Seventh Edition, Oxford University Press, 2015.

[3] E. Tlelo-Cuautle, Advances in Analog Circuits, InTech, 2011.

[4] C. Sánchez-López, A. Ruiz_Pastor, R. Ochoa-Montiel, and M. A. Carrasco-Aguilar, "Symbolic Nodal Analysis of Analog Circuits with Modern Multiport Functional Blocks," Radio Eng., vol. 22, no. 2, pp.518-525, June 2013.

[5] E. Tlelo-Cuautle, C. Sa´nchez-Lo´pez, E. Martı´nez-Romero, Sheldon X.-D. Tan, "Symbolic analysis of analog circuits containing voltage mirrors and current mirrors," Analog Integr Circ Sig Process 65:89–95, DOI 10.1007/s10470-010-9455-y, 2010.

[6] M. Fakhfakh, E. Tlelo-Cuautle, F. V. Fernández, "Design of Analog Circuits Through Symbolic Analysis," *https://books.google.com/books?isbn =1608050955, Computers 2012.*

[7] C. Sánchez-López *et al*, "Pathological Element-Based Active Device Models and Their Application to Symbolic Analysis," IEEE Trans. Circuits Syst. I: Regular Papers, vol. 58, no. 6, pp 1382 – 1395, June 2011.

[8] R. Hashemian, "Application of Fixator-Norator Pairs in Designing Active Loads and Current Mirrors in Analog Integrated Circuits," IEEE Trans. Very Large Scale Int. (VLSI) Syst., vol. 20, no. 12, pp 2220 – 2231, Dec. 2012.

[9] R. Hashemian, " Fixator-Norator Pairs vs Direct Analytical Tools in Performing Analog Circuit Designs," IEEE Trans. Circuits Syst. II, Exp. Briefs, vol. 61, no. 8, pp. 569 - 573, August 2014.

[10] R. Hashemian, "Application of Nullors in Designing Analog Circuits for Bandwidth", IEEE Inter. Conf. on Electro/Information Tech., EIT2016 conference proceedings, Grand Forks, ND, May 19 - 21, 2016.

[11] R. Hashemian, "Nullors in Finding/Modifying Transistor Parameters, Blindly", IEEE International Conference on Information and Computer Technologies (ICICT), DeKalb, Illinois, March 23-25, 2018.

[12] R. J. Baker, CMOS, Circuit Design, Layout, and Simulation, John Wiley &Sons, New Jersey, 2008.

Economic and Social Sustainability for Iraqi Middle Provinces

Maysoon Abdullah Mansor*

College of Engineering, Dept. of Civil Engineering, University of Tikrit, Iraq

A R T I C L E I N F O	A B S T R A C T
	This research is part of a study concerned with building an integrated framework for development of sustainable urban planning in Iraqi provinces. Seven Iraqi middle provinces are covered in this study. The aim of the study is to determine, measure and evaluate economic and social sustainability in the Iraqi middle provinces. *The stages to achieve this research were based on establishing a two levels hierarchy framework (a level for categories and a level for indicators). In this study, the Economic dimension comprised of 6 categories and 18 indicators and the Social dimension comprised of 8 categories and 41 indicators. Through using the Multi Criteria Analysis method and the Analytic Hierarchy Process, the Sustainability Index has been determined and used to measure and evaluate the economic and social sustainability in the Iraqi middle provinces*

1. Introduction

More than 50% of the worldwide population live in urban or urbanizing areas. In just five years, we will add another people, all will need jobs, food, security and opportunity. Sustainable development have become imperative to meet their own needs. [1], [2]

Generally , sustainable development concept comprises of three aspects: economic, environmental and social. There are other sources that show other aspects: sustainability, aaccording to LA21, has four dimensions, namely: economic, environmental, social and governance or institutional aspects [3], shows that the three sustainability main dimensions, all located within the governance dimension that manages and controls these dimensions.

1.1. The economic dimension

It is considered as a base for development and any disruption of it or to the draining of its resources eventually will lead to the weakening of its future development opportunities, and then a long-term economic perspective must be taken to solve the problems in order to provide effort, money and resources [4].

1.2. The social dimension:

It is the right of a normal person in a clean and sound environment in which he can practice all activities while ensuring his right in a fair share of natural resources and environmental and social services, and investing it in order to serve his basic needs (shelter, food, clothing, air ....) as well as the complementary needs to raise his standard of living (business, entertainment, fuel ....) and without reducing the chances of future generations[4].

1.3. Indexes

An index is a synthesis of indicators or composite indicator. Formalization of an indicator that necessitates the aggregation of several data or variables results in. In the field of sustainable development, the use of indexes facilitates the interpretation and understanding of given phenomenon indicators, particularly for the public. Indexing offers an instrument for benchmarking performance and provides information to enable decision-making and directing development [5]. According to its specific goals, Indicators should tell us in what fields the city is doing better than in others, and a single sustainability index should tell us whether the city is becoming more sustainable or not. Based on an aggregation method, the index is then calculated [6].

Composite indicators development is considered to be a unique approach for sustainable development evaluation. Calculating aggregated values is one of the common methods for establishing indices. Depending on its application, composites indices can be established without or with weights [7].

2. Methods Used in the Developing Sustainability Index

For developing mathematical sustainability index, the Methods used are:

Maysoon Abdullah Mansor, Email: maysoonm810@gmail.com dr.maysoonabdullah@tu.edu.iq.

2.1   The Analytic Hierarchy Process (AHP)

This method is most widely used Multi Criteria Decision Making method, as it presented a deriving priorities and structuring the decision problem logically and representatively. The use of pair-wise comparisons, in AHP method is, to compare the alternatives regarding to the various criteria and to assess weights of criteria" [8].

The comparisons and priority calculation of AHP pairwise are executed to get the relative weights for objectives, sub objectives, and measures on every level of criteria in the hierarchy [9]. Figure 1 describes the principal steps for achieve the AHP. Table (1) shows a preference scale using for comparison. Ref. [10] suggested using the following consistency index (CI) as in (1):

$$CI = \frac{\lambda_{max} - n}{n - 1} \tag{1}$$

Where
CI: is the consistency index
λ_{max}: is the largest Eigen value of matrix and
n: is the number of elements within a branch being compared. This consistency index can also be expressed as a consistency ratio as in (2):

$$CR = \frac{CI}{RI} \tag{2}$$

where:
CR: is the consistency ratio
RI: is a known random consistency index obtained from a large number of simulation runs and varies depending upon the order of matrix.

Table 2 shows the random consistency index values (RI) for matrices of order 1 to 10 obtained by approximating random indices using a sample size of 500[11].
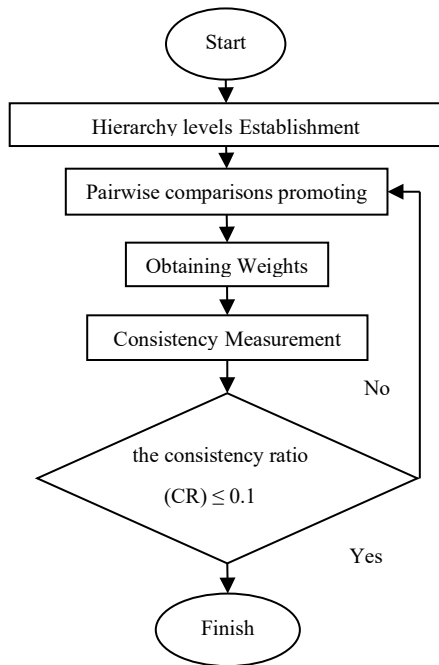
Figure 1 The principal steps for achieve the AHP [Researcher depending on [9]]

2.2   The Geometric Mean [12]

Equation (3) or (4) is used to find geometric mean:

$$G = \sum f_i \sqrt{\left(x_1^{f_1}\right)\left(x_2^{f_2}\right)\left(x_3^{f_3}\right)...\left(x_n^{f_n}\right)} \tag{3}$$

or

$$\log(G) = \frac{\sum f_i \log(x_i)}{\sum f_i} \tag{4}$$

Where, G: Geometric Mean
x_n: weights of each respondents' number.
f_i: respondents number for each weights.
n: $\sum f_i$: the all respondent sum.

Table 1 Saaty Rating scale

Importance	Definition	Explanation
1	Equal importance of both elements	Two elements contribute equally
3	Moderate importance of one element over another	Experience and judgment favour one element over another
5	Strong importance of one element over another	An element is strongly favoured
7	Very strong importance of one element over another	An element is very strongly dominant
9	Extreme importance of one element over another	An element is favoured by at least an order of magnitude
2,4,6,8	Intermediate values	Used to compromise between two judgments

Table 2 Average random index according to matrix size [11]

Matrix Size(n)	1	2	3	4	5	6	7	8	9	10
Random Index	0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49

2.3   Analysis Grey Relational (AGR) [13]

(GRA) is a measurement method for discussing the consistency of discrete uncertain sequence and its target. It initiates by the original sequence factor creation, the measurement space factor using grey-relational generation, that   is classified into three types:

- Smaller-the-better grey-relational generation: the minimum of the sequence factors is the ideal factor.
- Larger-the-better grey-relational generation: the maximum of the sequence factors is the ideal factor is.
- Nominal-the-better grey-relational generation: the one with the sequence factors target value in line is The ideal factor.

In grey relational generation, (5) used for normalized data corresponding to criterion, when Lower-the-Better (LB):

$$X_{i(k)} = \frac{\max y_{i\ (k)} - y_{i(k)}}{\max y_{i(k)} - \min y_{i(k)}} \tag{5}$$

Equation (6) used for the normalized data when Higher-the-Better (HB) criterion:

$$X_{i(k)} = \frac{y_{i(k)} - \min y_i}{\max y_{i(k)} - \min y_{i(k)}} \tag{6}$$

Where:

$X_{i(k)}$: is the value of $y_{i(k)}$ after the grey relational generation,

min $y_{i(k)}$: is the smallest value of $y_{i(k)}$, and max $y_{i(k)}$: is the largest value of $y_{i(k)}$.

The purpose of Grey relational grade is to appear the relation degrees between the sequences say, $[X_{0\ (k)}$ and $X_{i\ (k)}$, $i = 1, 2, 3…, n]$.

The purpose of using GRA, in this research, is to normalize the values of realistic indicators by using (6) as the research intends maximum sustainability (higher is better).

3. Developing Economic and Social Sustainability Index

Reference [6] suggested steps for building the framework of composite indicators (index) as described in Figure 2. the steps for developing economic and social sustainability index in current study, are the following:
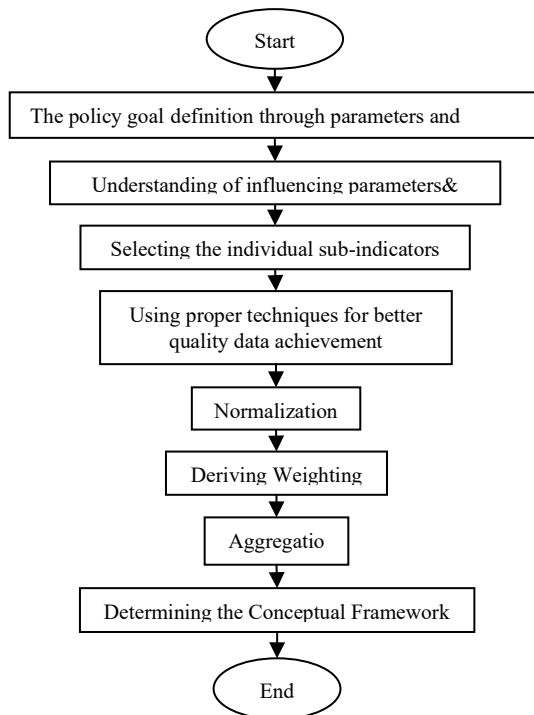
Figure 2: The Steps for establishing the framework of composite indicators [Researcher depending on [6]]

3.1. Developing Appropriate Sustainable Urban Framework

Reference [14] illustrated Developing three levels hierarchy Sustainable Urban framework for Iraqi cities (Dimensions,
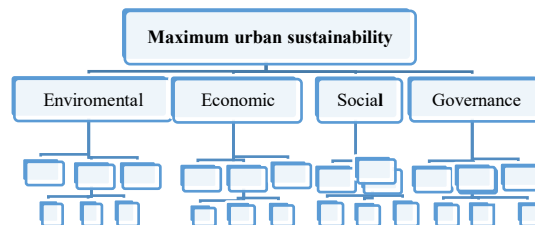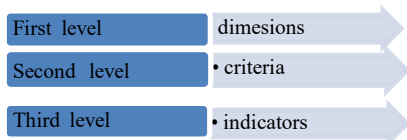
categories& indicators), as in Figure 3. This study deals with the economic and social dimension. Each dimension consists of two levels (categories level and indicators level). Six categories and eighteen indicators in The Economic dimension, and eight categories and forty-one indicators in The social dimension. the components for each dimension are shown in Table 3.

3.2. Determining Weightings Using AHP Method

Figure 4 illustrate the steps followed in the current research for determining framework components weighting by AHP method (Analysis Hierarchy Process).

The questioners used for finding the weight of categories, and indicators for economic and social dimension. The process of preparing questionnaire depends on research hierarchy framework and pair wise comparison. The questionnaire contains multiple issues and the answer accuracy needs a logical respondent with the following characteristics:

- Knowing about sustainability or urban sustainability.
- Involved in and active with projects and development of the governorate.

The researcher suggested to distribute at least (9) questionnaires for each governorate as in [15]. A list of the experts' respondents and their other information is summarized in Table 4. The number of received questionnaire respondents varies from one province to another, and the highest number of questionnaire received is (9) as in governorate of Baghdad and Salah Al-Deen and the lowest number is (4) as in the province of Wasit.

After accomplishing categories and indicators pairwise comparisons by the specialists' responses about the questionnaire, the aggregation for the individual judgments using the geometric mean as Saaty suggested in [16].

3.3. Determining Indicators Realistic Values

The researcher spent more than four months for collecting indicators realistic values for the study provinces. additional four months were needed for determining indicators values because of the difficulties to get fit value of indicator. The process of information gathering mainly depended on the Ministry of Planning / Central Statistical Organization & Information Technology (COSIT) in all its divisions, otherwise the values of indicators were obtained by other relevant ministries, authorities and departments, if information was unavailable. the process of determining values of indicators is as follows:

Figure 3: The hierarchy three levels Framework[Researcher]

Table 3: Economic and social sustainable framework component

Code	Details
2.	The economic dimension
2.1	The use of resources and energy
2.1.1	Electricity used per capita
2.1.2	Gasoline consumed per capita
2.1.3	Energy use per capita per year
2.2	Efficiency of economic development
2.2.1	Contribution of tourism to GDP
2.2.2	Investment share of GDP
2.2.3	Number of industrial projects
2.2.4	The share of government investment
2.3	Local government finance
2.3.1	Capital spending as a percent- age of total expenditures
2.3.2	Spending on research and development as a percentage of
2.3.3	Local government revenue
2.4	Services
2.4.1	Cost recovery
2.4.2	Loss in revenue as a result of power outages
2.5	Disasters
2.5.1	Percent of population living in areas at risk
2.5.2	Economic loss resulting from human and natural disasters
2.5.3	Safe design of the building of the topographic and geological
2.8	Economy population
2.8.1	Employment
2.8.2	Average family income
2.8.3	Per capita GDP
3.	The social dimension:
3.2	Health
3.2.1	The proportion of the number of hospitals and health
3.2.2	The proportion of the provision of potable water in urban areas, according to national standards
3.2.3	Hospital beds for every ten thousand people
3.2.4	rate of Mortality and birth
3.2.5	Number of doctors per capita
3.2.6	Access to sanitation facilities
3.2.7	Nutritional status of children
3.2.8	Immunization of children against disease infections
3.2.9	Prevalence of tobacco use
3.3	Education
3.3.1	The number of schools to population ratio
3.3.2	The number of researchers
3.3.3	Illiteracy rate for adults
3.3.4	Enrollment rates
3.3.5	The number of higher education graduates
3.3.6	The percentage of students who reach the fifth grade of primary education
3.3.7	The proportion of the number of students to teachers
3.3.8	The number of people who have graduated from junior
3.4	Housing
3.4.1	Housing shortage
3.4.2	The dimensions of the deficit housing and shortage ;neighborhoods, slums
3.4.3	Rent-to-income ratio
3.4.4	Optional decent housing or authorized
3.4.5	Type of tenure
3.4.6	Random housing
3.4.7	Abandoned housing
3.5	Social development and poverty eradication
3.5.1	percent of population living below the poverty line
3.5.2	The proportion of poor women , children with special needs who do not have access to community facilities and services
3.5.3	Child labor
3.5.4	Death under the age of five suffer from malnutrition
3.5.5	The proportion low-birth weight infants
3.5.6	The incidence of serious diseases
3.6	Violence and Crime
3.6.1	Violence in urban areas
3.6.2	Number of recorded crimes per 100,000 population
3.7	Urbanization and urban housing
3.7.1	The population growth rate
3.7.2	The number of families in one housing unit
3.7.3	Total net migration
3.7.4	The average household size for each family
3.7.5	The proportion of population living in urban areas
3.8	Culture of the community in sustainability
3.8.1	Public awareness of the population to hazardous
3.8.2	Percentage of the population who have gardens
3.9	Household Spending
3.9.1	Household spending
3.9.2	Average savings per household

- Some of the values were found directly.

- The researcher calculated some of these values with many equations and methods.

- Some indicators were calculated from some surveys that had a relationship to the indicators.

- Some indicators were new or that the relevant ministry responsible for providing the indicators values refused to provide the researcher with, so the researcher used an open interview questionnaire with Likert scale (i.e.,1: very low, 2: low, 3: medium ,4: high ,5: very high) to find its values.

- For the same reasons above, some of the economic indicators were found without monetary units.

- Some indicators included urban and rural regions together and some included urban only. In the first case, the indicator to be found was only for the urban through multiplication by the percentage of relevant urban in the governorate.

Start
↓
Establishment of a structural hierarchy
↓
Comparative pair wise for framework components
↓
Preparing questionnaire
↓
Determining nature & size sample respondents
↓
Questionnaire distribution
↓
Received questionnaire responses
↓
Finding Geometric mean for the response received
↓
Deriving framework components weighting
↓
End
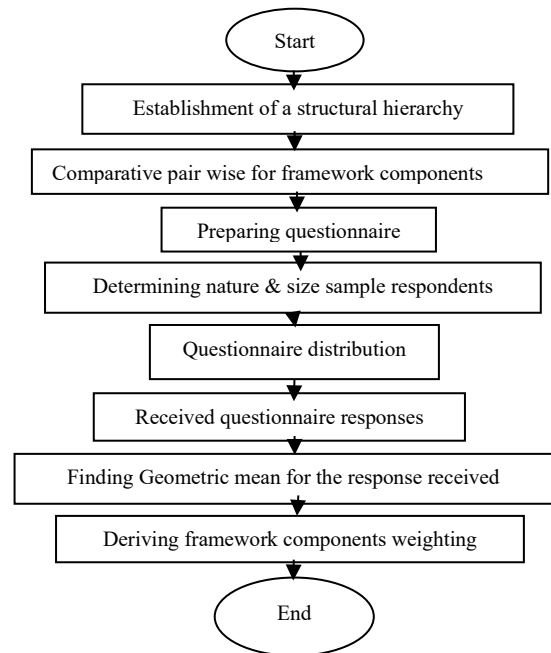
Figure 4: Flowchart for weighting framework components by AHP method [Researcher]

Table 4: Summary the respondents' information

provinces		Diala	Al-Anbar	Baghdad	Babylon	Kerbela	Wasit	Salah ALDeen
No. of respondents		8	5	9	7	5	4	9
Certificate	B.A.	2	0	0	0	2	2	3
	M.A.	2	0	4	3	2	0	1
	Ph.D.	4	5	5	4	1	2	5
profession	Planner	1	1	1	0	1	1	3
	Engineer	4	1	1	1	2	2	1
	Lecturer	4	3	3	6	2	1	5
Years of practice	10 to20	3	2	2	4	3	3	1
	more than 20	5	3	3	3	2	1	8

3.4. Indicators Influence identifying on Sustainability

Table 5 illustrated the direction of indicators against sustainability.

Table 5: Signs of indicators against sustainability [Researcher]

Signs of indicators	Means of the signs
-	If the indictor value increased sustainability decreased.
+	If the indictor value increased sustainability increased .

3.5. Using Analysis Grey Relational for Normalization indicators values

Because of the indicators values were not the same each other units, the normalization was performed to obtain one unit for all values of indicators, additionally to obtain sequential in the indicators values.

Equation (6) was used for normalization indicators values, the limits of indicators values normalization are between (0-1).

3.6. Determining Dimension Sustainability Index (DSI)by Aggregation

The base for determining the economic and social sustainability Indices is the aggregation as in (7).

$$DSI=\Sigma[W_c\ (\Sigma W_i X_i A_i)] \qquad (7)$$

Where:
DSI: Dimension Sustainable Index
W_c: Weight of Category c; W_i: Weight of Indicator i;
X_i: Normalized obtained of indicator value i in the study area;
A_i: 1 or -1, that depends on indicator i influence on sustainable; i: indicator Number for the category c;
c: categories Number.

4. Results

4.1. Categories and Indicators weight for Economic and Social Dimensions

Tables 6 and 7 show the weights and ranks of economic dimension categories.

It is clear from the table that the category(C2.1) ranking in two provinces is the second and fourth ranking and the highest weight in the Babylon province. the category(C2.2) ranking in two provinces is the first ranking and the highest weight in the Wasit province. The category(C2.3) ranking in three provinces is the first ranking and the highest weight in the Salah ALDeen province and sixth rank in two other provinces. The category(C2.4) ranking in three provinces is the first ranking and the highest weight in the Diala province .and second rank in two other provinces. The category(C2.5) ranking in three provinces is the sixth ranking and the highest weight in the Salah ALDeen province .and second rank in two other provinces. The category(C2.8) ranking in three provinces is the fifth ranking and the highest weight in the Diala province, third and fourth rank in

two other provinces. Tables 8,9 and10 show the weights and ranks of social dimension categories.

It is clear from the table that the category(C3.2) ranking in three provinces is the first and the highest weight in the Kerbela province. the category(C3.3) ranking in two provinces is the first ranking and the highest weight in the Wasit province and second rank in four other provinces. the category(C3.4) ranking in two provinces is the fourth ranking and the highest weight in the Al-Anbar province and fifth rank in two other provinces.

the category(C3.5) ranking in two provinces is the third ranking and the highest weight in the Diala province and fifth rank in three other provinces. the category(C3.6) ranking in two provinces is the sixth ranking and the highest weight in the Salah ALDeen province and seventh rank in three other provinces. the category(C3.7) ranking in three provinces is the eighth ranking and the highest weight in the Al-Anbar province. the category(C3.8) ranking in two provinces is the third ranking and the highest weight in the Diala province and fifth rank in two other provinces. the category(C3.9) ranking in two provinces is the seventh ranking and the highest weight in the Baghdad province and eighth rank in two other provinces. Table 11shows the priority summary for economic and social dimension categories.

Table 6: Weights and ranks of economic dimension categories

Category	C2.1		C2.2		C2.3	
Provinces	Rank	Value	Rank	Value	Rank	Value
Diala	2	0.157	5	0.115	4	0.117
Al-Anbar	1	0.072	2	0.078	1	0.283
Baghdad	4	0.109	6	0.055	1	0.179
Babylon	2	0.177	1	0.199	6	0.030
Kerbela	4	0.148	3	0.159	6	0.074
Wasit	3	0.153	1	0.241	2	0.213
Salah ALDeen	6	0.072	5	0.075	1	0.238

Table 7: Weights and ranks of economic dimension categories

Category	C2.4		C2.5		C2.8	
Provinces	Rank	Value	Rank	Value	Rank	Value
Diala	1	0.181	6	0.080	3	0.132
Al-Anbar	2	0.273	6	0.065	3	0.086
Baghdad	2	0.178	3	0.122	5	0.093
Babylon	5	0.075	3	0.144	4	0.142
Kerbela	1	0.163	2	0.160	5	0.111
Wasit	1	0.163	6	0.037	5	0.053
Salah Aldeen	3	0.174	2	0.200	4	0.098

Table 8: Weights and ranks of social dimension categories

Category	3.2		3.3		C3.4	
Provinces	Rank	Value	Rank	Value	Rank	Value
Diala	4	0.115	2	0.148	6	0.107
Al-Anbar	8	0.068	4	0.111	2	0.169
Baghdad	1	0.160	2	0.149	5	0.135
Babylon	1	0.261	2	0.176	5	0.089
Kerbela	1	0.206	2	0.133	3	0.121
Wasit	3	0.132	1	0.220	4	0.128
Salah Aldeen	7	0.093	1	0.163	4	0.121

4.2. Economic and Social Sustainability Index

In the current research, many mathematical equations, Microsoft Excel program was used. Table 12 shows a sample of sustainability index calculations for the economic dimension of Diyala province in excel sheet: After making calculations, Table13 illustrate the values of each sustainability dimension index in the current research. Figures 5and6 illustrate economic and social index for each province in study area. In the economic dimension, Baghdad got the first ranking followed by Wasit, and Salah ALDeen got the last ranking followed by Babylon. As for the social dimension, Baghdad got the first rank followed by Babylon and Wasit got the final ranking followed by Diyala.

Table 9: Weights and ranks of social dimension categories

Category	C3.5		C3.6		C3.7	
Provinces	Rank	Value	Rank	Value	Rank	Value
Diala	1	0.153	5	0.109	8	0.077
Al-Anbar	5	0.094	6	0.083	1	0.200
Baghdad	3	0.143	7	0.066	8	0.037
Babylon	5	0.089	7	0.068	3	0.167
Kerbela	5	0.109	6	0.102	4	0.111
Wasit	2	0.139	7	0.060	6	0.074
Salah Aldeen	3	0.128	2	0.151	8	0.062

Table 10: Weights and ranks of social dimension categories

Category	C3.8		C3.9	
Provinces	Rank	Value	Rank	Value
Diala	3.000	0.131	7	0.092
Al-Anbar	3.000	0.118	7	0.075
Baghdad	6.000	0.081	4	0.137
Babylon	4.000	0.105	8	0.036
Kerbela	7.000	0.097	8	0.076
Wasit	5.000	0.095	8	0.051
Salah ALDeen	5.000	0.113	6	0.113

Table 11: The priority summary for economic and social dimension categories.

Economic Category priority		Social Category priority	
Category	Rank	Category	Rank
C2.4	1	C3.2	1
C2.3	2	C3.3	2
C2.2	3	C3.5	3
C2.1	4	C3.8	4
C2.8	5	C3.4	5
C2.5	6	C3.6	6
		C3.7	7
		C3.9	8

2. Conclusions

In general, Degree of sustainability dimensions of Iraqi provinces, was very low: and Baghdad province got first in the economic and social index maybe it's because it is the capital. Salah ALDeen province got the last position in the economic index. Wasit province got the last position in the social index. The negative sign of the dimension index value means the sustainability was declined in that dimension. The economic

index consisted of only one province with negative signal (Kerbela).

Table 12: A sample of economic sustainability index calculations of Diala

Code	Wc	Ci	Wi	Xi	Ai	WiXiAi	CI
2.1	0.157						
		2.1.1	0.416	0.026	-1	-0.011	
		2.1.2	0.200	0.001	-1	0.000	
		2.1.3	0.384	0.016	-1	-0.006	
					∑	-0.017	-0.003
2.2	0.115						
		2.2.1	0.228	0.000	1	0.000	
		2.2.2	0.211	0.049	1	0.010	
		2.2.3	0.377	0.028	1	0.011	
2.3	0.117					0.021	0.002
		2.2.4	0.185	0.023	1	0.004	
		2.3.1	0.295	0.051	1	0.015	
		2.3.2	0.431	0.013	1	0.006	
		2.3.3	0.274	0.037	1	0.010	
2.4	0.181					0.035	0.004
		2.4.1	0.439	0.016	1	0.007	
		2.4.2	0.561	0.006	-1	-0.003	
2.5	0.080					0.003	0.001
		2.5.1	0.209	1.000	-1	-0.209	
		2.5.2	0.284	0.600	-1	-0.171	
		2.5.3	0.506	0.506	1	0.256	
2.8	0.132					-0.124	-0.010
		2.8.1	0.184	0.084	1	0.015	
		2.8.2	0.494	0.282	1	0.139	
		2.8.3	0.323	0.065	1	0.021	
						0.176	0.023
						DI	**0.018**

Table 13: The values of economic and social index

Provinces	EconomicIndexDI2	Social IndexDI3
Diala	0.018	-0.116
Al-Anbar	0.02	-0.062
Baghdad	0.202	0.092
Babylon	0.003	0.084
Kerbela	-0.008	0.005
Wasit	0.046	-0.128
Salah ALDeen	-0.014	-0.044

The social index consisted of four provinces with negative values (Wasit, Diyala. Al-Anbar and Salah ALDeen). The dimension sustainability indexes (DSI) were varied from one province to another because of the index value depends on the categories and indicators weights and the indicators realistic value where the weights are based on the nature of region and the opinions of experts while the indicators realistic value depend on several things, as the following:

- The difference in topography and geography of the region, which contributes to different capacities.

- The difference in population density in each province, which affects the distribution of economic projects and development plans.

- The direction of the longitudinal transport is determined along the Tigris and Euphrates rivers, neglecting the horizontal transport, which affects the distribution of economic projects and development plans in the study provinces in addition to the other aspects (environmental and social).

- The priorities of the governorate determined by the central and local governments in the province.
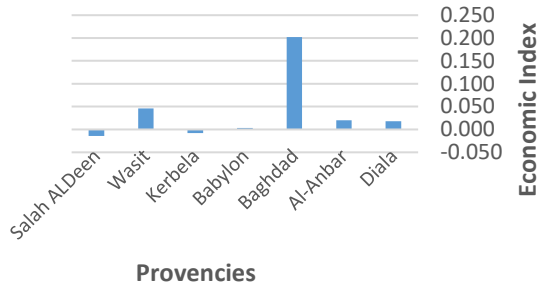
Figure 5: Economic indexes for each province in study area
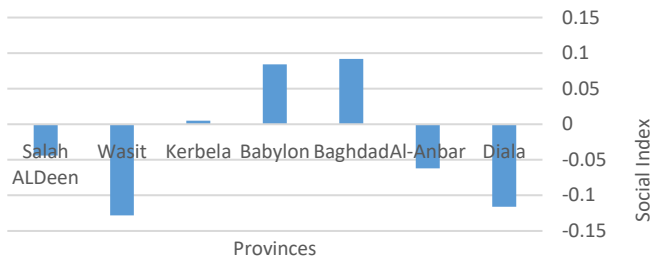
Figure 6: Social indexes for each province in study area

Conflict of Interest

There is many conflict such as respondents sample characteristics as mentioned in Section (3.2) are rare and if we find them, they will be very busy, so the researcher did the following steps for questionnaire distribution:

- Traveling for each governorate,
- Selecting most of the respondents,
- Distributing at least (9) questionnaires for each governorate, and
- Attending an open interview with respondents to ensure the understanding of questionnaire and to insuring the consistency of answers

Acknowledgment

References

[1] Portland State University, Twelfth International Conference on Environmental Cultural, Economic & Social Sustainability "Urban Sustainability – Inspiration and Solution", Portland State University | Portland, USA, 2016.

[2] Environment Management Group Wang, A Framework For Advancing Environmental And Social Sustainability In The United Nations System, United Nations, Geneva, 2012.

[3] Spangenberg, J.H). "The changing contribution of unpaid work to the total standard of living in sustainable development scenarios". Int. J. Sustainable Development, 5(4);461–475,2002.

[4] سليمان مهنا, ريدة ديب ,التخطيط من اجل الاستدامة ,مجلة جامعة دمشق للعلوم الهندسية ,المجلد 25 العدد 1, 2009

[5] Hsiang-Hsuan Chih, Yu-En Lin & Pin-Huang Chou, "Disposition Effect, Escalation of Commitment and Mutual Fund Performance".Management Review methodology,28 (4); 103-107, 2009.

[6] Tanguay, G. Rajaonson, U., Lefebvre, J.F., Lanoie, P.," Measuring the Sustainability of Cities: A Survey-Based Analysis of the Use of Local Indicators". Ecological Indicators;10: 407-418,2009. Available on www.cirano.qc.ca /pdf/publication/2009s-02.pdf .

[7] Singh, R.K., Murty, H.R., Gupta, S.K., Dikshit, A.K., "An overview of sustainability assessment methodologies". Ecological Indicators , 15; 281–299,2012.DOI of original article:10.1016/j.ecolind.2008.05.011. E-mail address: r.singh@pe- international.com (R.K. Singh). (Meadows, 1998). 1470-160X/\$ – see front matter © 2011 Elsevier Ltd. All rights reserved. doi: 10.1016/j.ecolind.2011.01.007.

[8] Loken, E. . Use of multi-criteria decision analysis methods for energy planning problems. Renewable and Sustainable Energy Reviews, 11(7): 1584-1595,2007.

[9] Kendrick, J. D. &Saaty, D., "Use Analytical Hierarchy Process For Project Selection". ASQ Six Sigma Forum Magazine, 6(4); 22-29 ,2007.

[10] Saaty, T.L.. A scaling method for priorities in a hierarchichal structure. Journal of Mathematical Psychology, 15, 234–281,1977.

[11] Saaty, T.L. . Fundamentals of the Analytic Hierarchy Process, RWS Publications, Pittsburgh, PA,2000.

[12] محمد أماني موسى, "التحليل الإحصائي للبيانات". مركز تطوير الدراسات العليا والبحوث، كلية الهندسة، جامعة القاهرة، الطبعة الأولى 2007. Available in www.capscu.com.

[13] Lin, S.T., Horng, S.J., Lee, B.H., Fan, P., Pan,Y., Lai, J.L, Chen, R.J and Khan,M.K.."Application of Grey-relational Analysis to Find the Most Suitable Watermarking Scheme", International Journal of Innovative Computing, Information and Control,7(9): 5389-5401,2011. ICIC International 2011 ISSN 1349-4198

[14] Mansor, M. A.& Mohammed ,S. R. "Developing Sustainable Urban Planning Indicators Framework for Iraqi Cities by Using Content Analysis and Delphi Method" International Journal of Modern Sciences and Engineering Technology (IJMSET)ISSN 2349-3755; Volume 2, Issue 3, pp.46-56, 2015.

[15] Akadiri,O.P., "Development of a Multi-criteria Approach for the Selection of Sustainable Materials for Building Projects". Ph.D. thesis, School of Engineering and the Built Environment, University of Wolverhampton, UK ,2011.

[16] Saaty, T.L., "Decision Making for Leaders: The Analytic Hierarchy Process for Decisions in a Complex World". RWS Publications, Pittsburgh, 2001.

Minimally Invasive, Thermal Energy Based, Cost-Efficient Method to Measure Fluid Flows in Compact Systems

Saurin Patel*, Rick Walker

SMC Ltd., Product Design and Development, 44087, USA

A R T I C L E   I N F O

A B S T R A C T

Flow measurements are important for numerous industries as it directly relates to proper equipment functioning and product quality. This work demonstrates a novel thermal energy based compact flow measuring sensor. The sensor injects thermal energy into the core which is then extracted by the moving flow through the center of the core. Temperature changes produced by thermal energy exchange is evaluated over time and then used to calibrate and measure flows. The sensor can detect extremely low energy changes therefore measure relatively low flow rates as well. Being minimally intrusive, it does not disturb the flow eliminating any additional resistance commonly introduced by intrusive type sensors. It is made from inexpensive components and is easy to assemble therefore making it affordable. The simple geometry allows for it to be easily incorporated into compact systems and can be used in miniature devices to measure low flows. We declare this minimally intrusive only due to breaking the flow path with the sensor element otherwise it is non-intrusive.

1. Introduction

Fluid flow measurement is an essential as well as critical step in numerous processes and finds applications in many fields such as medical diagnostic devices, pharmaceutical industry, food industry, automotive industry, chemical industry, compressed air systems, and others (1). Precise fluid management is essential to regulate basic functionalities of a system. In certain applications accurate flow measurement becomes the key player by directly influencing the quality of the product (2). Rapid research and developments in microfluidics also dictates a need to develop ways to accurately measure and analyze extremely low flow rates (3). Flow measuring can be classified broadly into two categories: intrusive and non or minimally intrusive. In an intrusive method, the flow encounters a constriction or other mechanical interference. Intrusive methods typically depend on the forces produced by the flow which are then correlated to flow rate values. Mechanical flow meters, pressure dependent meters, and vortex based flow meters are examples of intrusive type flow meters. These meter types are relatively inexpensive and are available in variety of shapes and sizes. However, they are limited by their repeatability, accuracy and operating ranges. Being intrusive they reduce the flow rate within the system due to added resistance above the inherent system flow resistance. At low flow rates the kinetic energy associated is also low and inadequate to be

converted into measureable mechanical movement (4). Fluids temperature changes affect the internal structure of these types of meters rendering this type of devices not suitable for applications where fluid temperatures changes are frequent and occupy a wide range (5).

Non-intrusive methods include optical, sonar, thermal, electromagnetic, Coriolis, and laser Doppler based flow meters. These types of flow meters do not interfere with the fluid motion or path and can measure most liquids. However electromagnetic flow meters can only measure the flow rate of a fluid conducting electricity and are expensive compared to their intrusive counterparts (6). Other flow meters based on optical, sonar, and Coriolis methods are accurate but very expensive to implement and have limited capabilities at low flow rate (6). Traditional thermal sensors need complex control system involving pulse width modulation (PWM) to constantly provide power to the heated element and maintain its temperature (6). The energy loss from the heated element due to fluid flow is correlated to the flow rate for measurements and requires extensive data acquisition. We herein present a novel way to measure flow rate and fluid velocity in a minimally intrusive, easy to assemble and cost-effective manner by simple cooling rate measurement obtained from temperature readings. The sensor only needs to be powered once for a reading, measures the energy loss from the system and establishes correlation with the flow rate. Moreover, the sensor was very easy to calibrate, capable of detecting low flows and has excellent

*Saurin H. Patel, Twinsburg Ohio, +1 (330) 342 7800, saurin.patel@smcltd.com

linearity and response time. In addition to the advantages mentioned above, the sensor is calibrated at a defined inlet temperature and therefore accounts for fluid property variations resulting in accurate readings.

2. Experimental Method

2.1. Material Selection and Sensor Design

A cylindrical aluminum rod with dimensions as shown in Figure 1a and 1b was used as the central core for the sensor. The density of aluminum being low compared to other conductive metals is capable of producing rapid temperature changes with small amount of stored thermal energy loss thus enabling low flow rates to produce measurable temperature differences. A circular opening equivalent to the inner diameter of the polymer tube was machined first at the center of the block for the fluid pathway. Counter boring operation was performed on the two flat outer surfaces for a short length to create an opening equal to the outer diameter of the polymer tubing. The tubing was then inserted from both sides and is represented in Figure 1a. The resulting geometry ensured uniform opening at the center of the aluminum core for the fluid motion, with minimal flow disturbances and alteration to system flow rate. Figure 1b shows the side view of the aluminum core. An opening with its center at the midway of inner diameter and outer diameter of the block was machined to accommodate the thermal sensor. Before inserting the sensor, adequate thermal grease was introduced into the opening to maintain thermal contact between the aluminum core and the sensor. The *Biot number* associated with the selected block is small (<0.1) which signifies analysis based on lumped heat capacitance method for the current system is valid with an error of less than 5%. Therefore, it is safe to assume that the temperature measured at any location on the aluminum block is at least 95% accurate and closely resembles the temperature of the entire block.
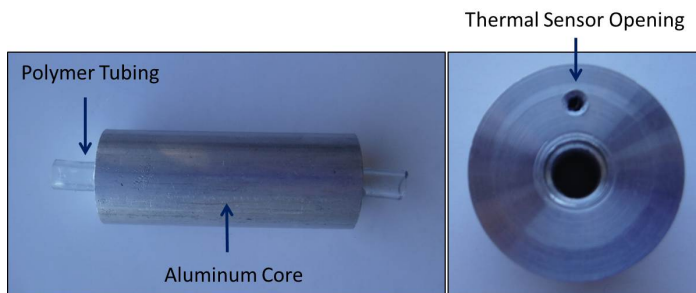
Figure 1: (a) Aluminum core with connected polymer tubes and (b) side view with counter bore and thermal sensor opening

2.2. Data Acquisition

Figure 2a represents the resistive heating element utilized to supply thermal energy to the aluminum core. A 12V DC power supply (YH-302D Yihua, China) connected to the heating element was controlled by Arduino UNO that used the temperature readings obtained from the thermal sensor (10K thermistor, see Figure 2b). The thermal sensor was calibrated using the Steinhart-Hart equation which is capable of producing accurate results between the ranges of 0 - 50° Celsius (Error -0.005 – 0.002° C). Code was written and embedded into to the Arduino UNO to monitor the temperature readings continuously and perform logical and arithmetic operations as needed. Initially, fluid was cooled to a desired temperature (14°C and 15°C in the present testing) using a novel system which also contained a pump to produce necessary

volumetric flow rate. The system continuously monitored fluid temperature just before it entered into the sensors' aluminum core. Once fluid reached desired stable temperature a constant flow was allowed to pass through the core and the flow rate was measured using a calibrated flow meter (intrusive). The embedded code then turned the power supply on when the temperature of the aluminum core dropped to 16° C and turned the power supply off when the temperature reached 21° C. The constant flow of cold fluid dropped the temperature of the core. Time required by the fluid moving at a particular flow rate (recorded from the intrusive flow meter) to drop the temperature of the aluminum core by 2° C from 20 to 18° C was recorded. Cooling rate of the flow was calculated using (8). The flow rate to the aluminum sensor core was then randomly altered by pinching the polymer tube and all the previously mentioned steps were repeated for four different flow rates and two different inlet flow temperatures. The cooling rates obtained from the code and volumetric flow rates from the flow sensor were tabulated and are presented in the results and discussion section. The correlation obtained was then utilized to calibrate the minimally intrusive sensor.
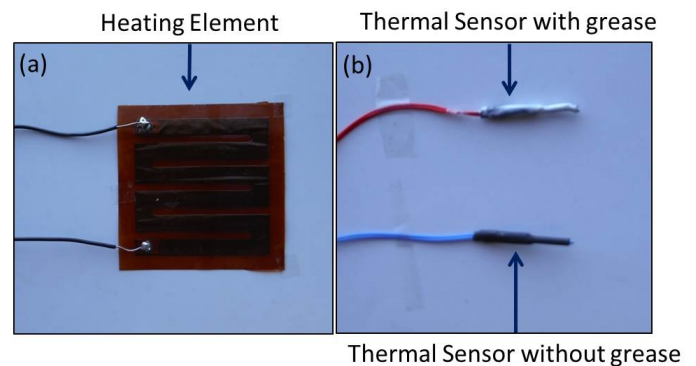
Figure 2: (a) Heating element with cables that are to be connected to a DC power supply and (b) thermal sensor used to monitor temperature readings (with and without thermal grease)

3. Theory

3.1. Thermal Analysis

The heater element attached to the peripheral surface of the aluminum core elevates the temperature of the entire metal by converting electrical energy into thermal energy. Thermal energy received by the aluminum core from the outer surface is conducted inwards towards the center. The energy reaching the center is carried away by the fluid making contact with the aluminum core through the mode of convective heat transfer as represented in Figure 3.

$$\dot{E}_{st} = \dot{E}_{in} - \dot{E}_{out} \qquad (1)$$

$$\dot{E}_{st} = \dot{E}_{out} \qquad (2)$$

The amount of energy stored is equal to vector summation of energy entering and exiting the system in the absence of internal energy generation. The energy balance of flow rate measuring system is represented by (1). When the temperature of the aluminum core as measured by the thermal sensor reaches 21° Celsius, the power supply to the heater is terminated. As a result the energy entering the system is effectively zero from that moment onwards. Eliminating the energy entering term (1) can be condensed into (2) which suggest that the energy stored in the aluminum core is gradually conveyed away by the energy exiting

the flow rate measuring system. In the present system, energy can exit through three different medium namely (a) convective heat transfer – water and (b) convective heat transfer – air and (c) radiative heat loss. The insulating layer of Poron© covering the entire outer aluminum core surface minimizes the convective heat
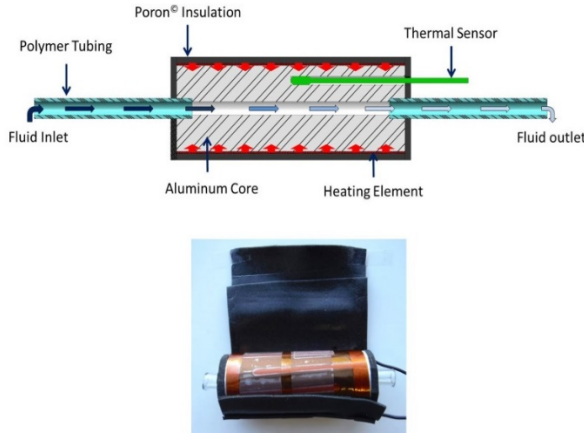
Figure 3: Schematic representation of the sensor including the heater element, aluminum core, thermal sensor, fluid tubing, and insulation. The sensor used for experimentation is also shown in the image.

transfer produced by air flow and therefore can be neglected. The radiative heat transfer losses for the temperatures at which the current system operates is also minimal and can be neglected as well. Eliminating other modes of heat transfer it is safe to conclude that the energy stored in the aluminum rod after the power supply to the heater is terminated is conveyed out by the fluid (14°C and 15°C) flowing through the center of the core. The energy balance between stored energy in the aluminum core and fluid motion is represented in (3). The amount of energy stored in the fixed mass aluminum core can be quantified as shown in (3) where m_{al}, C_P, ΔT_{al}, h, A, ΔT_f, and Δt are mass, heat capacity, temperature difference in the aluminum core, convective heat transfer coefficient, area fluid makes contact with the aluminum core, temperature difference between fluid entering and exiting the center and time interval respectively (7).

$$m_{al}C_p\left[\frac{\Delta T_{al}}{\Delta t}\right] = hA[\Delta T_f] \qquad (3)$$

Mass, heat capacity, area of contact, and the temperature difference of the aluminum core (2° Celsius) in the current test system can be termed constant. Moreover, the system is designed to produce minimum variation in the fluid temperature exiting the core by utilizing a relatively small length of the core and thus temperature difference of fluid can be assumed constant for most practical purposes. Arranging all the constants on one side and variables on the other (3) can be represented in a different form as shown in (4). The modified equation dictates that the product of convective heat transfer coefficient and time interval is practically a constant. In other words, the fluid condition for which the convective heat transfer coefficient is large, requires a smaller time interval to extract stored energy compared to a fluid condition with smaller convective heat transfer coefficient for the above mentioned flow rate measuring system. The correlation suggests, with the increase in convective heat transfer coefficient of the fluid the time needed to cool the aluminum core by 2° Celsius will decrease.

$$\frac{m_{al}C_p\Delta T_{al}}{A\Delta T_f} = h\Delta t \qquad (4)$$

The convective heat transfer coefficient based of Nusselt number evaluated using Gnielinski correlation that accounts for friction factor (surface roughness) f, Reynolds number Re_D, Prantl number Pr, thermal conductivity of fluid k, and hydraulic diameter D is given by (5). The correlation in (5) can be used with the hydraulic diameter, independent of the surface thermal condition (7). As seen from (5), the convective heat transfer coefficient is directly proportional to the Reynolds number of the fluid which in turn is proportional to the flow rate of the fluid (6). Q, ϑ and A_C are volumetric flow rate, kinematic viscosity, and cross-sectional area respectively in (6). Based on the above analysis we can deduce that with the increase in flow rate the convective heat transfer coefficient associated with the fluid also increases. The proposed minimally invasive flow rate measuring sensor exploits the above mentioned correlation between the fluid flow rate and the convective heat transfer coefficient.

$$h = \frac{\left(\frac{f}{8}\right)(Re_D - 1000)Pr}{1 + 12.7\left(\frac{f}{8}\right)^{0.5}(Pr^{0.67} - 1)}\frac{k}{D} \qquad (5)$$

$$Re_D = \frac{QD}{\vartheta A_C} \qquad (6)$$

$$h \propto Re_D \propto Q \qquad (7)$$

3.2. Working Principle - Sensor

In order to quantify the energy exiting the aluminum core at different volumetric flow rate, a cooling rate term $\Delta T_{al}/\Delta t$ was defined that is a direct function of convective heat transfer coefficient h. The cooling rate as shown in (8) can be defined as the decrease in temperature of the aluminum core per unit time. From the above analysis we predict that with the increase in flow rate the convective heat transfer coefficient increases and therefore the cooling rate.

$$\left[\frac{\Delta T_{al}}{\Delta t}\right] = h\left[\frac{A\Delta T_f}{m_{al}C_p}\right] \qquad (8)$$

Theoretically from (8) we can say that the cooling rate or energy loss from the aluminum core will increase with an increase in the convective heat transfer coefficient and the other terms on the right side of the equation are practically constant therefore acting as a multiplication factor only. Cooling rate at several different flow rates and couple different inlet fluid temperatures (14°C and 15°C) were experimentally obtained and tabulated. The results are discussed in the next section. The experimentally obtained data was then utilized as calibration factor to evaluate flow rate in a minimally invasive manner.

4. Results and Discussion

After Figure 4 represents the relationship between volumetric flow rates through the sensor core and the associated cooling rates obtained from the code for inlet fluid temperature of 14° Celsius. A multiplication factor was used to represent the cooling rates with a better resolution. Time measurements for the flow to reduce the aluminum core temperature by 2° Celsius were extracted from the code and manual calculations were performed to cross check the cooling rates obtained directly from the code. The cooling rates obtained using both the methods were found to be identical. As can be seen from the experimentally obtained plot, the cooling rate increases with the increase in volumetric flow rate. The

experimental results are consistent with the theoretical evaluation performed previously (7 & 8) which predicted that the cooling rate is proportional to the flow convective heat transfer coefficient h, which in turn is proportional to the volumetric flow rate Q. Additionally, it is important to note that the relationship between the cooling rate and volumetric flow rate was linear for the tested flow rates. We do acknowledge minor error in flow rate measurements originating from the accuracy of the intrusive flow meter. The error in the experimental measurements is accounted by the error bars as represented in the plots. Utilizing a Flow meter with greater accuracy can be used to obtain more accurate results for the calibration. Nevertheless, the result predicts a linear relation between the entities of concern in the given test performed.
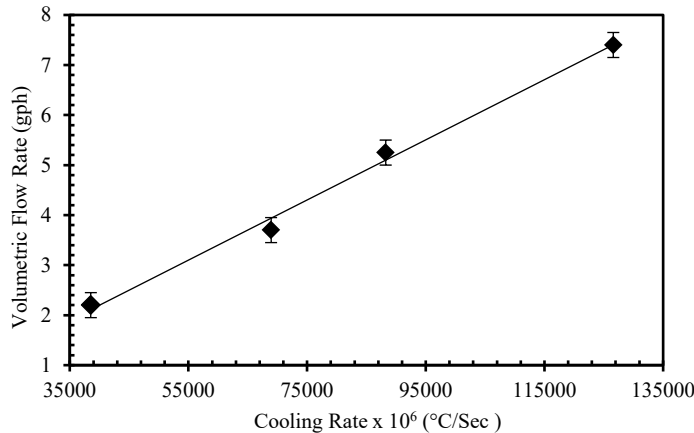
Figure 4: Cooling rates obtained from the code at various volumetric flow rates for inlet fluid temperature of 14° C

Figure 5 represents the relationship between volumetric flow rates through the sensor core and the associated cooling rates obtained from the code for inlet fluid temperature of 15° Celsius. Consistent with the earlier study for inlet temperature of 14° Celsius the cooling rate increases with the increase in volumetric flow rate through the aluminum core. Moreover, the coorelation between the entities for 15° Celsius was also linear. Suggesting cooling rate as defined earlier can be used as a good calibration factor to measure flow rates and velocity (from fluid properties and sensor geometry) in compact systems using a minimally intrusive method.
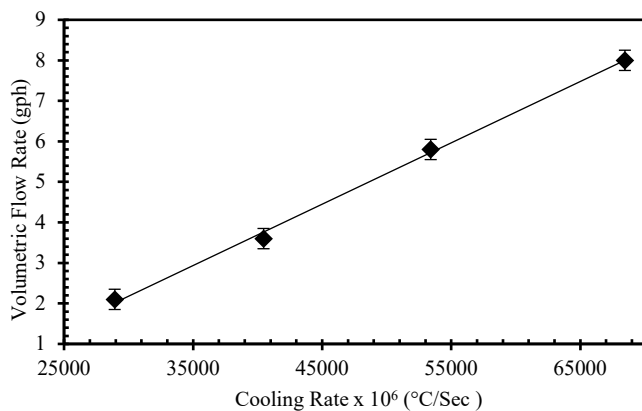
Figure 5: Cooling rates obtained from the code at various volumetric flow rates for inlet fluid temperature of 15° C

On comparison of the two plots presented, the cooling rates for the 14° Celsius inlet fluid temperature are higher than the 15° Celsius for a given volumetric flow rate. This can be expected as a colder fluid can extract more thermal energy from the aluminum core compared to a relatively warm fluid in a defined time interval. Therefore, the time required for a colder fluid to reduce the core temperature by 2° Celsius was smaller compared to a warmer fluid which results in a higher cooling rate (8). The correlation obtained for the two different fluid inlet temperatures were programmed back into the code and volumetric flow rates were measured using them. The relation produced accurate readings in both the cases.

The slope values calculated from the plots above for the 14° Celsius and 15° Celsius inlet fluid temperatures are 200 °C/sec/gph and 60°C/sec/gph respectively. If we compare the slopes of the two plots, it can be seen than the colder fluid (14° Celsius) has a smaller slope compared to the warmer fluid (15° Celsius) reconfirming the cooling rate was higher for a colder fluid compared with a warmer fluid. Analysis the slope valves, we can also observe that the slope values decrease with the decrease in the inlet fluid temperature. Knowing that the relation between the cooling rate and flow rate is linear for both the inlet fluid temperature and realizing that the cooling rate values at no flow condition will be same (i.e. small cooling rate at no flow condition defining the origin for the curves) for any inlet fluid temperature, we can define a single calibration factor or correlation by collecting adequate experimental data. The correlation can then be used to develop a single sensor that can identify volumetric flow rates for a broad range of inlet fluid temperature conditions.

5. Conclusion

We have developed a prototype sensor to measure volumetric flow rate and fluid velocity in compact flow systems using a minimally intrusive method for a defined inlet fluid temperature. The sensor was developed using an aluminum core that is heated and is allowed to be cooled by convective heat transfer produced by the fluid motion. The time associated with the aluminum core cooling was used to define the calibration factor and eventually flow rate. The linear relationship between the cooling rate and volumetric flow rate can be effectively exploited to develop a universal sensor that can measure flow rates for several different inlet fluid temperatures. The absence of any mechanical components or flow restrictors in the sensor's measurement path prevents any volumetric flow rate modifications within the system. No flow restriction in the system from the sensor means the fluid driving mechanism (e.g. pumps & motors) can be efficiently used as it does not have to overcome the additional losses created by the presence of sensor restriction. Moreover, the sensor uses commonly available items making it extremely economical to fabricate and easy to assemble providing a significant advantage over expensive ultra-sonic, electromagnetic, Coriolis and other flow meters. However, we appreciate the fact that the response time of the system can be slow for extremely low flow rates and the sensors need to be calibrated for the operating temperature.

Conflict of Interest

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] D. N. Ku, "Blood flow in arteries" Annual Review in Fluid Mechanics. 21, 399–434, 1997. https://doi.org/10.1146/annurev.fluid.29.1.399

[2] L. X. Yu, "Pharmaceutical Qulaity by Design: Product and Process Development, Understanding, and Control" Pharma Res., 25, 2463-2463, No. 10, 2008, https://doi.org/10.1007/s11095-007-9511-1

[3] S. Patel, D. Showers, P. Vedantam, T-R. Tzeng, S. Qian, X. Xuan "Microfluidic separation of live and dead yeast cells using reservoir-based dielectrophoresis" Biomicrofluidics, 6(3), 034102, 2012, https://doi.org/10.1063/1.4732800

[4] J. G. Santiago, S. T. Wereley, C. D. Meinhart, D. J. Beebe, R. J. Adrian, "A particle image velocimetry system for microfluidics" Experiments in Fluids, 25, 316-319., 1998, https://doi.org/10.1007/s003480050235

[5] H.A. Stone, A.D. Stroock, A. Ajdari, Engineering flows in small devices: "Microfluidics towards a Lob-on-a-Chip" Annual Review in Fluid Mechanics. 36:1, 381-411, 2004, https://doi.org/10.1146/annurev.fluid.36.050802.122124

[6] R. C. Baker, Flow measurement handbook, Cambridge University Press, 2016

[7] T. L. Bergman, A. S. Lavine, F P. Incropera, D. P. DeWitt, Fundamentals of Heat and Mass Transfer, 2011.

Framework for a Chemical Substance Reporting System

Sukhraj Singh Takhar[1,*], Kapila Liyanage [2]

[1]*Materials Management and Chemical Reporting, Assent Compliance Inc. & PhD Research Student, College of Engineering and Technology, University of Derby, Derby, DE22 3AW, United Kingdom.*

[2]*College of Engineering and Technology, University of Derby, Derby, DE22 3AW United Kingdom.*

A R T I C L E I N F O

A B S T R A C T

In this paper a chemical substance reporting system is presented to enable industry to assess the impacts of increasing chemical regulations. Chemical regulations impose the need to monitor, control and restrict the use of hazardous substances. As chemical substances become more regulated, industry requires mechanisms to identify potential business continuity risks, posed by increased regulation. A chemical reporting system is one such mechanism that identifies chemical substances used on their (i) own, (ii) in mixtures, (iii) in materials, (iv) in internally defined articles (products) and (v) from articles (products) procured from the supply chain. The chemical reporting system will contrast product related information against chemical regulation substance lists, identifying the applicable reporting obligations and potential supply chain risks.

1. Introduction

Regulatory control is a method in which governments can impose a set of measures upon society to act in a certain manner. The development of new chemical substances gathered pace following the end of the second world war, where the vast amount of new chemicals at the time were being introduced to market at a rapid rate, with much reduced testing cycles. Following the Thalidomide and Asbestos scandals, chemical regulations first began to appear during the 1960's [1].

The core aims of chemical regulations are to limit the use of hazardous substances, by applying appropriate monitoring, control and restriction measures [2].

Chemical regulations within Europe have evolved from regulators reacting to chemical substance data, when hazards become known, as in the case of the Thalidomide and Asbestos scandals, to requiring industry to provide data on regulated chemical substances within products in order to gain access to the market, this coined the term '*no data no market*'.

Chemical regulations such as the European Union (EU) Restriction of Hazardous Substances (RoHS) [3, 4]. EU Registration Evaluation Authorisation and restriction of CHemicals (REACH) [5] and other international regulations based on both EU RoHS and EU REACH have resulted in industry

having to generate vast amounts of information in relation to chemical substances used in articles (products) [6, 7, 8].

A feature of traditional engineering organizations [9] is to compartmentalize data within specific functional areas, which results in multiple sub-systems storing data. This makes the process of identifying and collating chemical substance related information difficult.

Chemical regulations will aim to regulate more and more hazardous substances over time, this results in industry having to develop systems which record the use of chemical substances. Not understanding internal and supply chain usage of hazardous substances has the potential to create business continuity risks, where the hazardous chemicals substances become further regulated.

This paper contributes to chemical substance reporting literature by identifying a research gap of multiple organisational data silos, which currently store the required data elements. The purpose of this paper is to derive a set of data tables which could then be applied to existing systems or create a new system (cloud based, data warehouse, access database, etc) to perform chemical substance reporting activities.

2. Methodology

This study follows a three-step approach to conduct the literature review. The research consisted of (1) initial literature

* Sukhraj Singh Takhar, Email: Raj.Takhar@assentcompliance.com
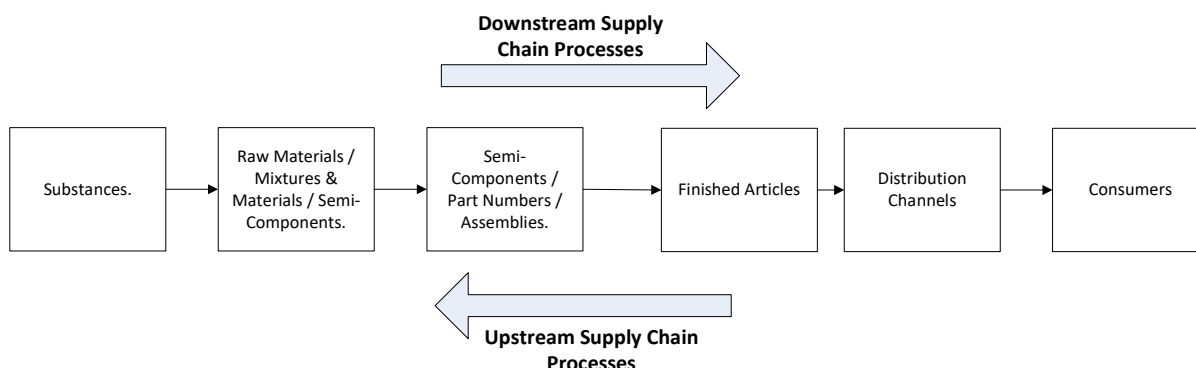
Figure 1. Article transformation cycle (simple supply chain)

search against chemical regulations, industry examples of chemical reporting systems appearing within the title of an article. selecting the most relevant articles; (2) cross-referencing supply chain management articles; (3) application of previous work experience in development and implementation of chemical reporting systems within the Aerospace and Defence sector, this includes expert interviews with users of previously implemented systems.

3. Findings

3.1. Supply Chain Evolution

The term 'Article Transformation Cycle' is the process of utilising raw materials, processing substances and mixtures to produce finished articles. Figure 1 presents the article transformation cycle.

Supply chains are a collection of organizations / elements, selling / flowing articles and services. The traditional supply chain focused on: (1) maintaining deliverables; (2) reducing costs; (3) increasing customer satisfaction levels [10, 11] Supply chains evolved from simple, labour intensive articles through to highly complex modern articles [12].

Understanding the article transformation cycle is important to see where chemical substances, mixtures and materials are being applied to articles.

3.2. Impacts of Chemical Regulations

Each chemical regulation presents industry with potentially new impact(s) as shown in Figure 2. Chemical regulations define lists of substances against which specific actions / obligations are to be undertaken: (1) regulatory notifications; (2) supply chain declarations; (3) customer declarations; (4) safe use guidance for consumers and article users; (5) continued authorized use of a substance request; (6) through to prohibited use of a substance, this information requires substance identification at the article level [13, 14].

As the number of chemical regulations increased, so too has the number of reportable substance lists [15, 16, 17]. This has resulted in specific industry sectors having to develop harmonized substance lists for their supply chains to report against, these have appeared in the Automotive [18], Electronics [19, 20] and AD sectors [21].

3.3. Traditional Organizational Information Flows

To develop a chemical reporting system, existing data silos and flow of information need to be examined. Figure 3 presents a systems diagram, based on expert interviews:

- Materials function: generate the specifications (material / process based) with details of applicable chemical substance(s); mixture(s) or material(s).
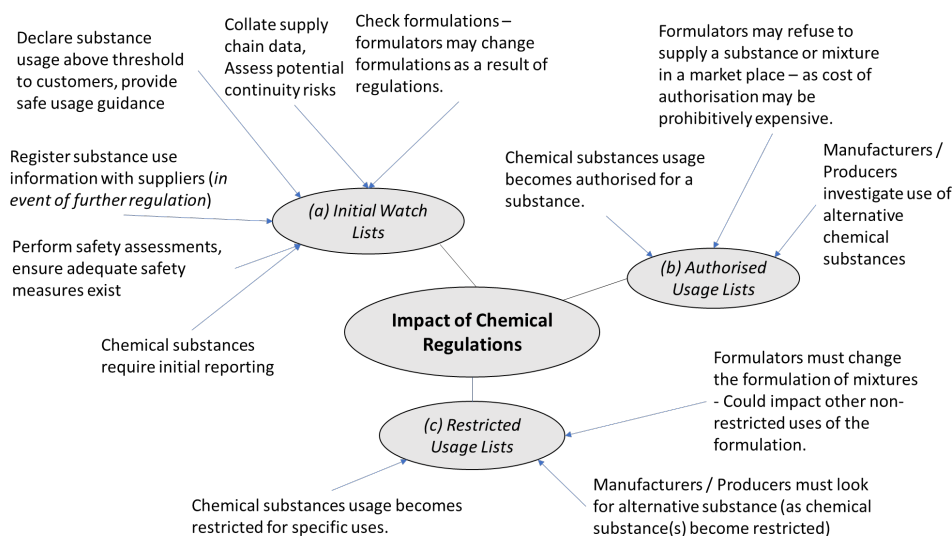
Figure 2. Impacts of chemical regulations

- Design function: generate individual engineering drawings for article(s) referencing specifications.

- Purchasing function: responsible for purchasing articles from supply chain.

- Engineering function: responsible for article transformation cycle for internal produced articles.

- Stores function: receipt, dispatch articles.

- Compliance function: (1) highlight potential business continuity risks of articles which contain regulated substances. (2) perform regulatory reporting and obligation (notifications / authorisation) tasks.

- Health and Safety function: ensure safety of employees, users of articles and the environment.

- Sales function: communicate with the customer and provide any regulatory reporting as required.

- Quality function: maintain consistent quality procedures and practices across the organization.

- Chemical regulators: define chemical substance lists and applicable obligations.

- Customers: consumer(s) of articles.

Issues arise where variability in data consistency and quality causes incorrect analysis taking place.

3.4. Aerospace and Defence (AD) Sector Organizational Information Flows

Based on expert interviews with companies from the AD sector, the following common practices were identified:

- Specifications: (1) configuration management appears to be limited; (2) standard practices focus on updating specification data with additional substance and mixture information as opposed to creating new specification numbers, this was primarily to avoid the cost updating geometry drawings with new specification numbers.

- Drawings: (1) drawings were not altered unless a change occurred to an article fit, form or function; (2) with specifications detailing the material and process specification data.

- Contracts: (1) supplier agreements; (2) audit data (3) payment terms and conditions; (4) required document(s): (5) statements of conformity; (6) supplier chemical substance declarations; (7) conflict mineral reporting; (8) safe use information; (9) industry reporting standards, etc.

- Data ambiguity: poor quality control from (1) requirements capture; (2) configuration management of specification data cause issues, as shown in Figure 4.

- Data flows: a typical finished AD article may consist of multiple sub-articles, all of which need to be collated and rolled up, with the chemical substance reporting system.

Evolving supply chains: AD supply chains are not simple, as a result of a number of vertical and horizontal integrations, it is possible at one level for customers and competitors to be suppliers at another level, for example Bombardier are both a component supplier and a customer for Rolls-Royce aero engines, if as a supplier it does not supply adequate substance information, this will make it hard for them to request full substance information as
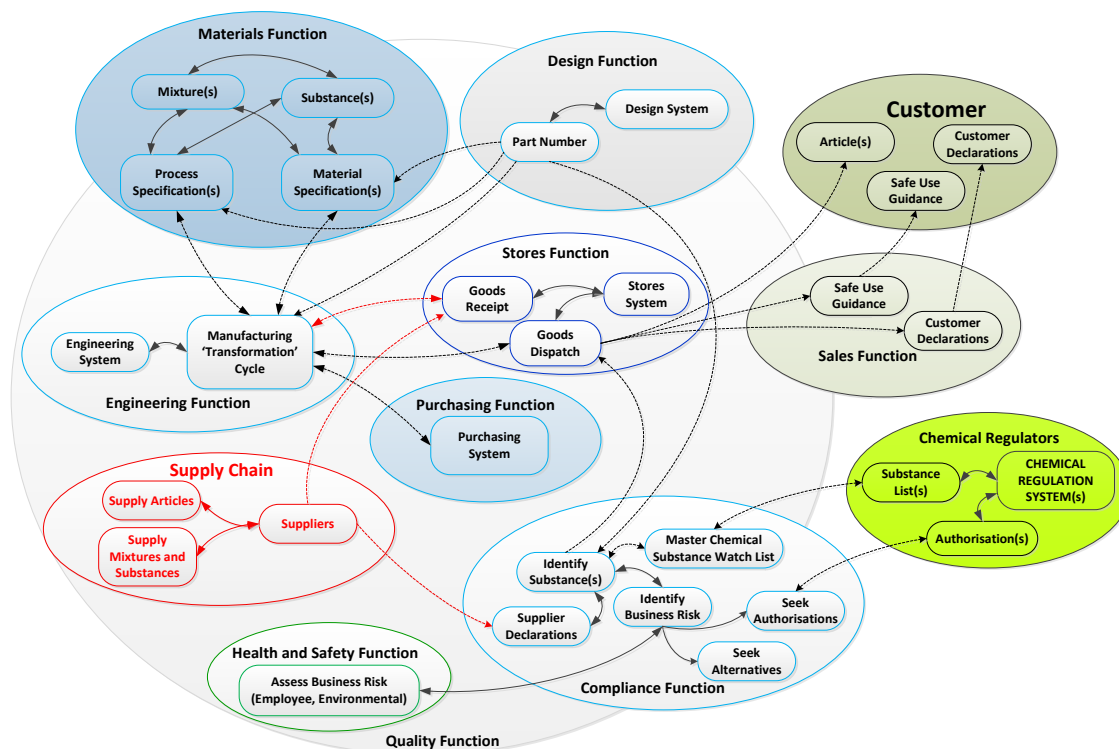
Figure 3. The chemical substance reporting system context model

a customer. A number of high-level acquisitions by Airbus, Boeing, GE Aerospace and United Technologies Aerospace Systems has resulted in this type of scenario. Figure 5 depicts a supply chain for a typical complex product, showing the flow of products which will require data relating to chemical substances, mixtures and materials to flow along the supply chain.

- Multiple actors: AD articles can be considered as complex articles which can consist of multiple sub-articles joined together to complete the finished article, the article transformation cycle may consist of multiple actors performing various roles.

- Maintaining data ambiguity by using many chemical substances to one specification (many to one), incurs cost inefficiencies of having to (1) checking all data from substance(s) to specification; (2) verifying actual substances used to enable determination of declarable and non-declarable chemical substance(s).

- Adopting a one chemical substance or mixture for one specification approach is the recommended way forward, however it does require AD companies to understand that it will result in improved configuration management enabling greater substance traceability. The negative aspect will be in the more frequent generation of new engineering drawings.

Category	Sub-Category	Examples of Information	Data Ambiguity Caused By
Contractual Documents	Supplier Agreements	Specific terms – Article supply (batch sizes) / Definition of obligations	(1) Supplier agreements signed for several years in advance; (2) Generic contract clauses for reporting cause confusion; (3) Not defining precise reporting standards; (4) Not stating any reporting requirements.
	Payment Terms	Based on article acceptance	
	Required Documents	Statements of conformity	(1) Data may be spread across multiple sub-systems; (2) Tracing substance(s) across transformation cycle; (3) Understanding declarable data and safe use guidance data to be provided
		Material declarations	
		Conflict mineral reporting	
		Safe use guidance	
		Industry reporting standards	
	Audit Data	Auditing supplier processes and checking required documentation is in place.	(1) Lack of supplier awareness; (2) Poor quality control; (3) Data may be spread across multiple sub-systems.
Design Documents	Engineering Design Documents	Dimensional data.	(1) Identifying substances in articles; (2) Poor document version control management.
	Materials Information	Material specifications	(1) Identifying substances in processes; (2) Poor document version control management.
		Process specifications	

Figure 4. Chemical substance reporting examples of data ambiguity
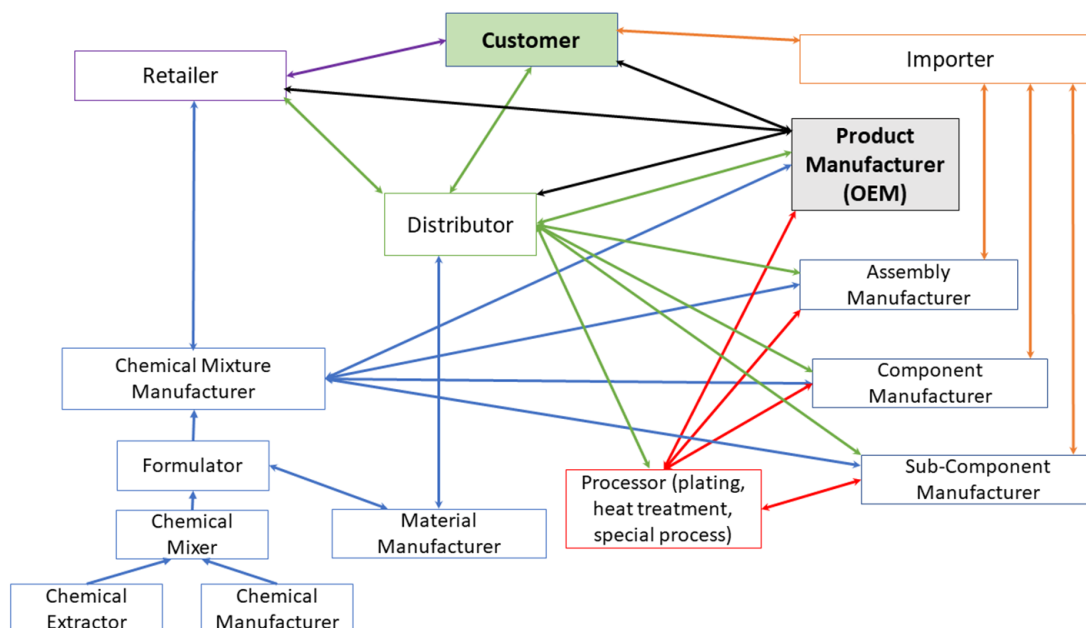
Figure 5. Complex supply chains

4. The Chemical Substance Reporting Framework Model

4.1. Define Data Elements and Formats

Examine internal data flows as shown in Figure 3, understand the common attributes which may be passed between sub-systems. List these attributes down as they become the central data elements for the chemical reporting system.

Data formats define the manner in which data is collated and transmitted, from simple spreadsheets, through to CSV, TXT and XML files to parse data between systems in a consistent manner. XML is deemed to be the most flexible.

4.2. Understand Your Internal Data Silos

The following sub-sections examine internal functional areas in terms of progressing from a functional design, purchasing and handling raw materials, manufacture and distribution of articles to consumers.

Design Function Data Elements:

The design function is observed as being the starting point for the end to end analysis. Figure 6 presents the product definition based on substances, mixtures and materials being assigned to specifications, which would be attached to the article (product) geometry drawings.

1. Chemical substances, mixtures / formulations (2 or more chemical substances); materials (generally defined as a chemicals and mixtures that constitute a physical thing) are identified via specifications.

2. Specifications can be (1) material specifications which identify substances; mixtures / formulations and materials which are present within the final manufactured product; (2) process specifications which identify substances; mixtures / formulations and materials which are used in the process of manufacture (sometimes referred to as process chemicals).

3. Engineering will create applicable machining programs, material handling information, assembly instructions, inspection data; test data to manufacture and release products.

4. Design drawings define product geometry data, which is used by engineering to produce products, identifying applicable material and process specifications.

5. Products are then distributed to consumers, where they may be used on their own, or rolled up into higher level assemblies.

Figure 6. Basic Product Definition

Note = **Bold Text** denotes specific data requirements of the Design Function
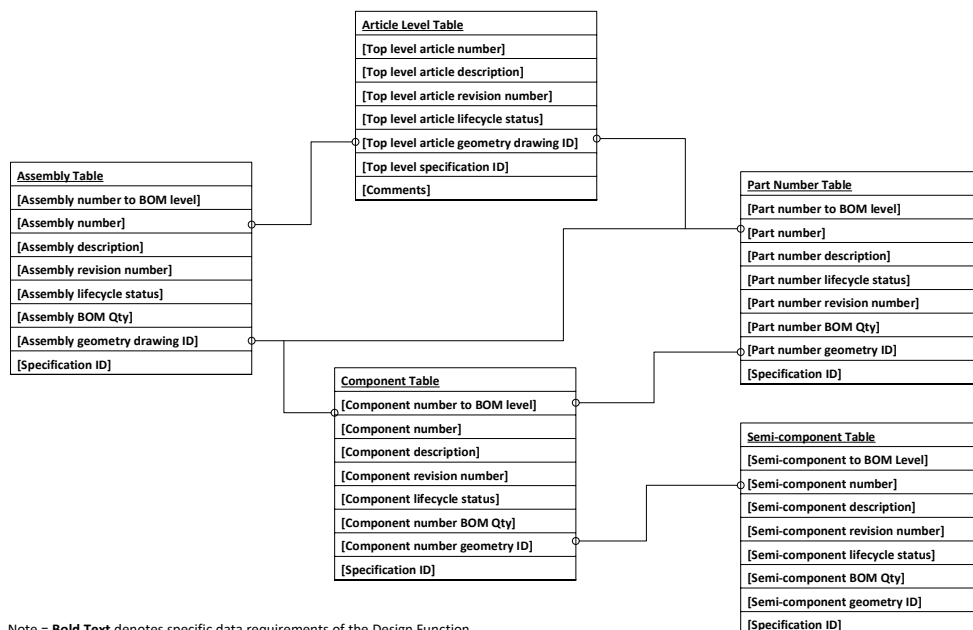
Figure 7. Design Function Data Elements
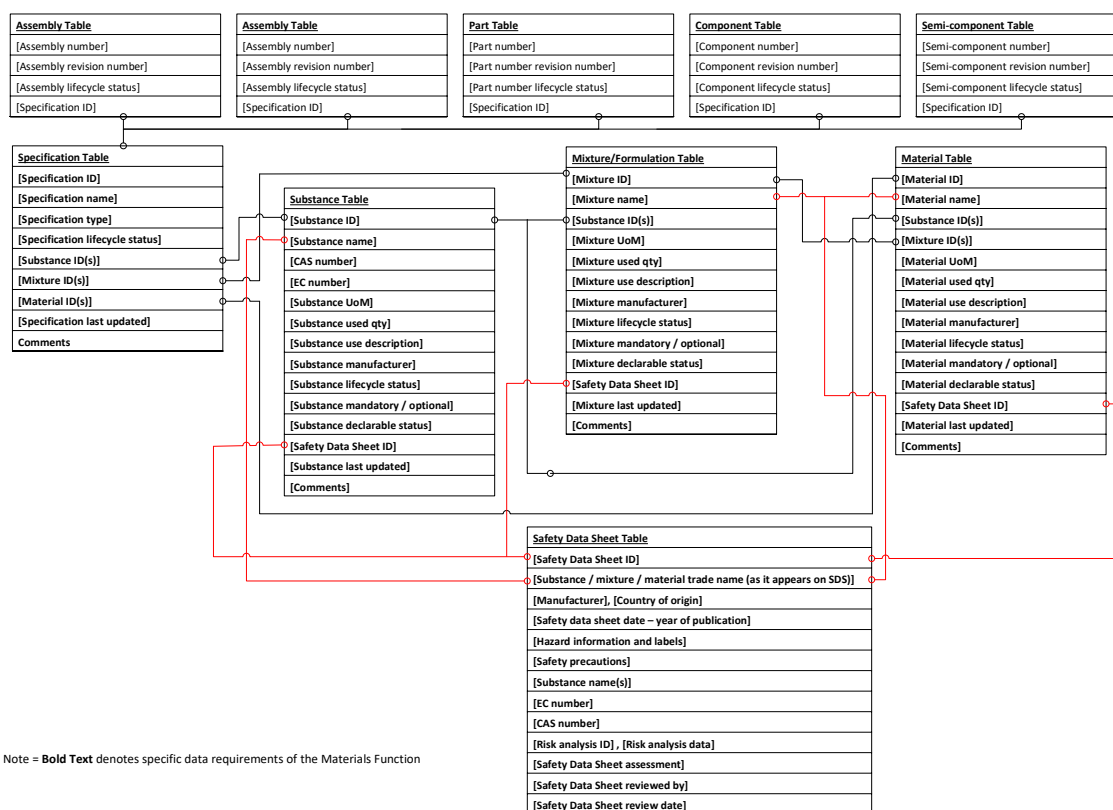
Figure 8. Materials Function Data Elements

The role of the design function is to define the product designs, Figure 7 presents the assumed data requirements / data sources which would occur from the design cycle. The basic premise is that products (top level article numbers) may consist of (1) assemblies (collection of parts and/or components), or; (2) individual part numbers, which may consist of separate components or; (3) components consisting of lower level semi-components

Materials Function Data Elements

The role of the materials function is to define chemical substances, mixtures and materials to a given type of specification. The specification types can be either: (1) material specifications – the chemical substance, mixture or material appears on the finished product, or; (2) process specifications – chemical substance, mixtures and materials are used as part of the manufacturing cycle.

Figure 8 depicts data utilised by the materials function, the standard substance, mixture and materials are extended using IDs as opposed to enable more granular reporting of a specific usage of a substance, mixture or material, in relation to specification which may be referenced in a geometry drawing created by the design function.

Master Chemical Inventory List

A master chemical inventory list identifies chemical substances and mixtures physically held in stock. The logic would be to (1) locate all physical chemical substances, mixtures and materials; (2) note the trade name, manufacturer, year of manufacture and country of origin data; (3) input the data into the

applicable chemical substance, mixture or material tables; (4) additionally add the SDS/MSDS data to the applicable SDS table; (5) adjust any stock level data within the stores table.

Engineering Function Data Elements

The role of the engineering is to produce products. To transform substances, mixtures and materials into finished products, the data from the design and material functions is extended as shown in Figure 9 to include machining, testing, and assembly instructions.

A specific make / buy flag would be utilized in conjunction with the purchasing function to determine if an article, assembly, part, component, semi-component is internally produced or externally sourced from the supply chain.

Purchasing Function Data Elements

The purchasing function is responsible for the procurement of substances, mixtures and materials, from suppliers. The make / buy flag determined by the engineering function determines if a product is manufactured internally or procured from the supply chain. Figure 10 presents the basic data elements required by the chemical reporting system, detailed purchasing related information is expected to maintain within a purchasing system. Sales Function Data Elements

The sales function is responsible for the sale of products (articles) to customers. Figure 11 presents the basic data elements required by the chemical reporting system, detailed sales related information is expected to maintain within a sales system.

Article Level Table

[Top level article number]
[Top level article description]
[Top level article revision number]
[Top level article lifecycle status]
[Top level article geometry drawing IDs]
[Specification ID]
[Comments]
[Top level article machine program]
[Top level article inspection data]
[Top level article assembly instructions]
[Top level article test data]
[Make / buy flag]

Assembly Table

[Assembly number]
[Assembly description]
[Assembly revision number]
[Assembly lifecycle status]
[Assembly geometry drawing IDs]
[Specification IDs]
[Comments]
[Assembly machine program]
[Assembly inspection data]
[Assembly assembly instructions]
[Assembly test data]
[Make / buy flag]

Part Table

[Part number]
[Part number description]
[Part number revision number]
[Part number lifecycle status]
[Part number geometry drawing IDs]
[Specification IDs]
[Comments]
[Part number machine program]
[Part number inspection data]
[Part number assembly instructions]
[Part number test data]
[Make / buy flag]

Component Table

[Component number]
[Component description]
[Component revision number]
[Component lifecycle status]
[Component geometry drawing IDs]
[Specification IDs]
[Comments]
[Component machine program]
[Component inspection data]
[Component assembly instructions]
[Component test data]
[Make / buy flag]

Semi-component Table

[Semi-component number]
[Semi-component description]
[Semi-component revision number]
[Semi-component lifecycle status]
[Semi-component geometry drawing IDs]
[Specification IDs]
[Comments]
[Semi-component machine program]
[Semi-component inspection data]
[Semi-component assembly instructions]
[Semi-component test data]
[Make / buy flag]

Note = **Bold Text** denotes specific data requirements of the Engineering Function

Figure 9. Engineering Function Data Elements

Article Level Table

[Top level article number]
[Top level article description]
[Top level article revision number]
[Top level article lifecycle status]
[Top level article geometry drawing IDs]
[Specification ID]
[Make / buy flag]

Assembly Table

[Assembly number to BOM level]
[Assembly to BOM Qty]
[Assembly number]
[Assembly description]
[Assembly revision number]
[Assembly lifecycle status]
[Assembly geometry drawing ID]
[Specification ID]
[Make / buy flag]

Part Number Table

[Part number to BOM level]
[Part number to BOM Qty]
[Part number]
[Part description]
[Part number lifecycle status]
[Part number revision number]
[Part number geometry ID]
[Specification ID]
[Make / buy flag]

Component Table

[Component number to BOM level]
[Component number BOM Qty]
[Component number]
[Component description]
[Component revision number]
[Component lifecycle status]
[Component number geometry ID]
[Specification ID]
[Make / buy flag]

Semi-component Table

[Semi-component number to BOM level]
[Semi-component number BOM Qty]
[Semi-component number]
[Semi-component description]
[Semi-component revision number]
[Semi-component lifecycle status]
[Semi-component number geometry ID]
[Specification ID]
[Make / buy flag]

Purchasing Function Table

[Internal number (Top level / assembly / part / component / semi-component)]
[Make / buy flag]
[Supplier product name]
[Purchased product description]
[Supplier ID]
[Supplier Name]
[Supplier contact ID]
[Supplier phone (public)]
[Qty ordered last 6 months]
[Last order date]

Purchasing Function Personal Data Table [GDPR]

[Supplier ID]
[Supplier contact ID]
[Supplier contact name]
[Supplier contact phone number]
[Supplier contact email address]
[Supplier address (street)]
[Supplier address (city)]
[Supplier address (state)]
[Supplier address (zip code)]

Note = **Bold Text** denotes specific data requirements of the Purchasing Function

Figure 10. Purchasing Function Data Elements

Article Level Table

[Top level article number]
[Top level article description]
[Top level article revision number]
[Top level article lifecycle status]
[Top level article geometry drawing IDs]
[Specification ID]

Sales Function Table

[Internal number (Top level)]
[Customer product number]
[Product description]
[Customer ID]
[Customer contact ID]
[Qty sold last 6 months]
[Last order date]

Sales Function Personal Data Table

[GDPR]
[Customer ID]
[Customer contact ID]
[Customer contact name]
[Customer contact phone number]
[Customer contact email address]
[Customer address (street)]
[Customer address (city)]
[Customer address (state)]
[Customer address (zip code)]

Note = **Bold Text** denotes specific data requirements of the Sales Function

Figure 11. Sales Function Data Elements

Stores Function Data Elements

The stores function is responsible for storage of chemicals, mixtures and materials, releasing them into the engineering function as required. Figure 12 presents the basic data elements required by the chemical reporting system.

Transportation Function Data Elements

The transportation function covers packaging, storing and distribution of the end article to consumers. Figure 13 presents the basic data elements required by the chemical reporting system.

Figure 12. Stores Function Data Elements

Figure 13. Transportation Function Data Elements

Figure 14. Health and Safety Function Data Elements

Health, Safety and Environment (HSE) Function Data Elements

The assumption is that the HSE function will review data in conjunction with the Materials Function inputting into the risk analysis and risk actions of the Safety Data Sheet table, as shown in Figure 14:

Quality Function Data Elements

The role of the quality function is to define procedures for all other functional areas of the business to adhere to, as shown in Figure 15:

Quality Function Table
[Quality procedure ID]
[Applicable function name]
[Quality procedure name]
[Quality procedure details]
[Quality procedure status]
[Cross reference standard]
[Quality procedure last review date]

Note = **Bold Text** denotes specific data requirements of the Quality Function

Figure 15. Quality Function Data Elements

Note = **Bold Text** denotes specific data requirements of the Compliance Function

Figure 16. Compliance Function Data Elements for Internal and Procured Products

Figure 17. Compliance Function Data Elements for Customer Declarations

Compliance Function Data Elements

The compliance function plays two pivotal roles: (a) internally manufactured and procured products from the supply chain as shown in Figure 16; (b) generation of appropriate declarations to customers, as shown in Figure 17:

Technical Publications Function Data Elements

The technical publications function produces user guides, service manuals which are used by end user, as shown in Figure 18:

4.3. Understand Your External Data Silos

The following sub-sections examine external data silos which require data to be stored within a chemical substance reporting system, in order to highlight areas of potential business continuity risk:

Develop a Master Chemical Substance List

Examine all the regions where articles are produced, distributed and sold to consumers. This will inform you of the applicable chemical regulations which may affect your organisation.

Develop the concept of a 'Master Chemical Substance List', which details the chemical substances which are at subject to chemical regulations as shown in Figure 19, also see Appendix 1, Declarable Substance List Table.

Understand Data Flows Across Your Supply Chain

Figure 7 presented the view that articles are defined by drawings which detail the specifications, which in turn relate to substances, mixtures and materials. Figure 5 depicted a supply chain for a typical complex product. Figure 20 presents an example of how the data may be collated and rolled, as part of the article transformation process across a supply chain:

Note = **Bold Text** denotes specific data requirements of the Technical Publications function

Figure 18. Technical Publications Function Data Elements
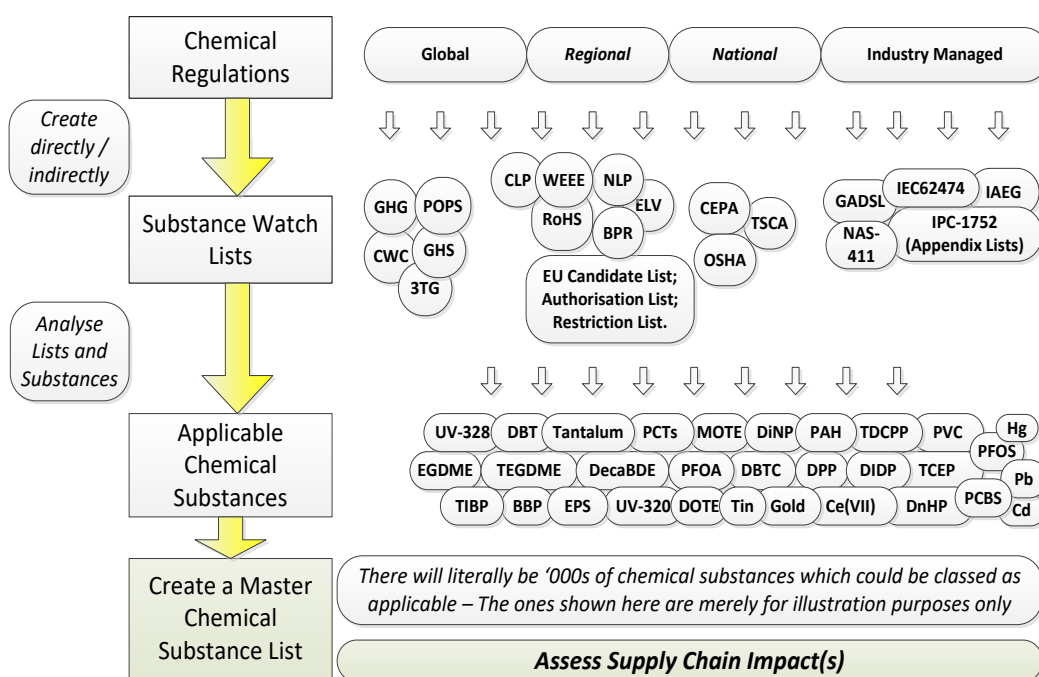
Figure 19. Create a master chemical substance list
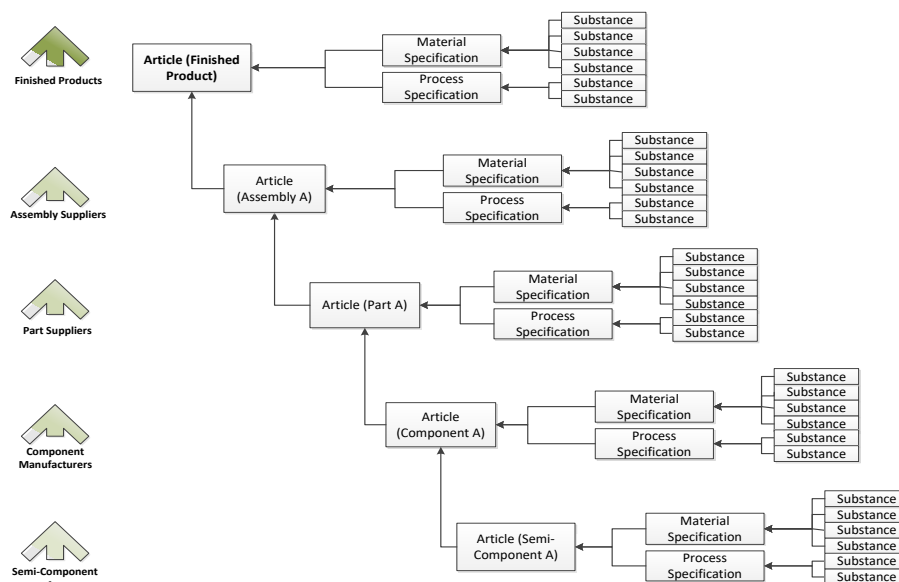
Figure 20. Substance usage tracking within the article transformation process

Develop a Supply Chain Reporting Questionnaire

A supply chain questionnaire is used to capture data from the supply chain for procured articles. The data elements to be used come from the supplier declaration table elements defined in Figure 17 and also shown in Supplier Declaration Table, Declaration Data Table. The data elements allow for existing data exchange standard templates to be used [19, 20].

Define Supply Chain Reporting Tasks

- Is any supply chain reporting currently being undertaken, if so, how?
- Which functional area will bear responsibility for collecting data from the supply chain? [*purchasing / compliance / materials / HSE*].

Supply Chain Engagement Tasks

- Establish early supply chain engagement.

- Ensure suppliers contracts state the need for chemical substance reporting as a standard requirement. Agree this requirement with your suppliers.

- Engage with suppliers on any reporting templates to be used, test the use of these templates and make any necessary adjustments.

Supply Chain Training

- What are the anticipated training needs? – engage with the supply chain to derive requirements.

- Training should inform the supply chain in relation to legal obligations to report chemical substances above any applicable threshold level.

- How is the training expected to be communicated? [*on-line training manuals / printed training matter / webinars / face-to-face meetings*]

- Define applicable processes for handling supply chain queries.
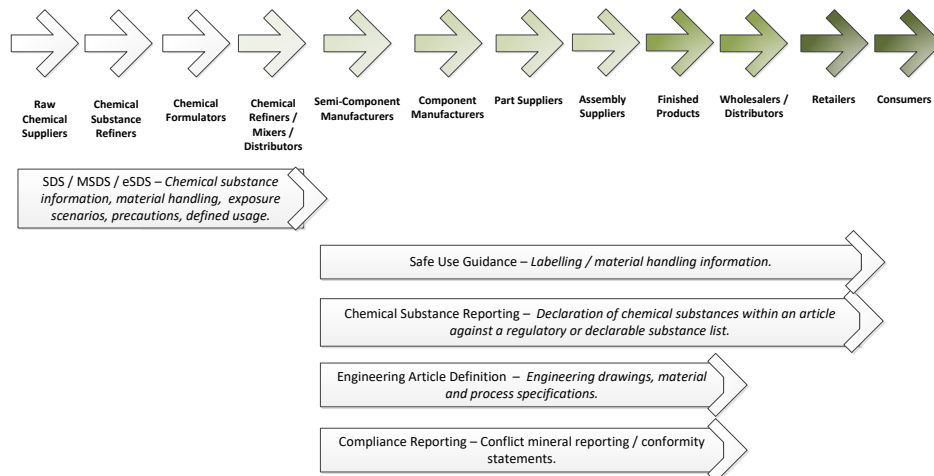
Figure 21. Article transformation cycle: push chemical substance reporting system

Execute Supply Chain Requests – Push or Pull?

- Push type systems exist where the lowest supply chain tier flows data from the bottom upward as shown in Figure 21. Push type systems work best in industries which are highly regulated and have well defined articles.

- Pull type systems are used within industries that contain complex articles, for example AD industries. Pull type systems involve the highest tiers in the supply chain initiating the chemical substance reporting requests [28].

Receiving Supply Chain Data Back

- Supply chain reporting questionnaires transmitted to the supply chain and start to flow back into the organisation.

- A location for storing the supply chain reporting questionnaires should be defined.

- From the storage location, the supply chain reporting questionnaires, should then be processed (using a manual or automated process) and then imported into a target system (where the supplier declaration tables reside).

- The target system may then process the data, rejecting incomplete reporting questionnaires and performing additional processing tasks on successfully imported supply chain reporting questionnaires.

- Back-up copies of processed and rejected files should be stored onto applicable folders, which may then be used to recover records in the event of a system failure.

5. Discussion

5.1. Where Could the Chemical Substance Reporting Data be Stored?

The intent of the chemical reporting data model is to define the required data elements as shown in Table 1, needed to enable accurate chemical substance risk analysis against chemical regulation substance risks. The data model is intended to be implemented on a wide range of IT systems including existing engineering design (PDM, PLM), supply chain management (ERP), or onto other platforms (spreadsheets, databases, custom applications, data warehouses or cloud-based systems).

5.2. Where Could the Chemical Substance Reporting Data be Stored?

- Develop metrics to: (1) highlight supplier response rates to the supply chain reporting questionnaires; (2) flag data quality issues back to suppliers.

- Using data from (1) substances and mixtures on hand (master chemical inventory list); (2) internally defined products (engineering data), and; (3) supply chain reporting information for procured products; (4) run reporting against all applicable database tables to highlight areas of potential business continuity risk, identifying substances which appear on different regulatory lists and perform the appropriate obligations and actions.

- Develop action plans to act on the business continuity risks identified.

5.3. Evaluate the Chemical Reporting System

- Maintain the data within the chemical inventory lists, to ensure an accurate representation of physical substance, mixtures, materials on hand.

- Maintain the master chemical substance list, as chemical regulations are updated on a regular basis.

- Use feedback comments from your supply chain to main any necessary adjustments to the supply chain questionnaire. Adjust as required, and ask the suppliers to identify any areas which impact response rates.

Table 1. Functions, Systems and Related IT System Types

Functional area	Safety Data Sheet System	Material Specification System	Compliance Reporting (*REACH / RoHS /CMR / CLP / BPR / ELV / Supply Chain Engagement*)	CAD Drawing System	PDM / PLM Systems	ERP System	Manufacturing Execution System (*Machining / Inspection*)	Technical Publications
Chemical Substance Reporting Questionnaire			X					
Compliance Function			X		X			
Master Chemical Substance list	X	X	X					
Health and Safety Function	X		X				X	
Quality Function	X	X	X		X	X	X	X
Materials Function	X	X	X					
Design Function			X	X	X			
Engineering Function			X	X	X		X	X
Purchasing Function			X		X	X		
Transportation Function			X			X		
Sales Function						X		X

- In terms of chemical substance reporting connect with stakeholders in your business and observe feedback comments relating to system reporting, adjusting as required.

6. Conclusions

As technological advancements are being made, more and more substances will continue to be developed and released onto the marketplace. Industry will continually strive to develop new product offerings which utilise new chemical substances, mixtures and materials.

As awareness of the risk of specific chemical becomes known, chemical regulators will take the necessary steps to monitor, control and restrict the use of substances which are deemed to be hazardous substances. The net result will over time, be a continually growing list of substances which industry will need to adhere to.

Industry needs to develop chemical reporting systems which observe (1) chemical substances, mixtures and materials held in hand; (2) chemical substances, mixtures and materials internally defined for articles manufactured internally or outsourced to the supply chain; (3) chemical substances, mixtures and materials which are contained with articles procured from the supply chain.

By not implementing a chemical reporting system, industry may not be able to identify the risks of: (1) not clearly understanding details any hazardous substances being used and the application of appropriate safety measures; (2) loss of market access, if no substance or safety information is provided; (3) financial penalties for not identifying regulated chemical substance(s) in a timely manner; (4) supply chain disruption if raw material(s) cannot be sourced as they contain restricted substance(s); (5) potential product failures if a supplier has changed a material / formulation, which affects the performance of a product, which has not been recorded correctly within existing systems; (6) not realising the benefits of sustainability [12], green chemistry [34], recycling and reusing materials as prescribed by the circular economy.

The current framework has examined the need to identify and report on substances within an article, against the chemical regulation substance lists.

The framework model presented in this research paper should allow reporting against a wide range of chemical regulations: EU REACH [5]; EU RoHS [3, 4]; EU Waste Framework Directive [29]; EU Classification; Labelling and Packaging (CLP) [30]; EU Biocidal Properties Reporting (BPR) [31, 32] and EU Conflict Mineral Reporting (CMR) [33] reporting.

The data elements presented in the figures and Table 2 outline the required fields which enable accurate collation of chemical substance data at the material level (appearing on the finished article) and process level (required in order to produce the finished article).

The expectation is that organisations will be able to utilize the findings of this research paper and apply the information to develop appropriate internal chemical reporting systems.

Future extensions to this framework will encompass (i) developing the model with a real-world Delphi study; (ii) developing flow analysis [35] and decision-based modelling [36].

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

References

[1] European Commission, Dangerous Substances Directive 67/548/EEC, 1967, [Online]. Available from: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:31967L0548.

[2] Regulation of chemicals wiki, last modified 3 July 2018, [Online]. Available from: https://en.wikipedia.org/wiki/Regulation_of_chemicals.

[3] European Commission, EU RoHS Directive 2002/95/EC, 2002, [Online]. Available from: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32002L0095.

[4] European Commission, EU RoHS recast Directive 2011/65/EU, 2011 [Online]. Available from: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32011L0065.

[5] European Commission, EC REACH Regulation EC 1907/2006, 2006 [Online]. Available from: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32006R1907.

[6] H. Selin, "Global governance and regional centers: multilevel management of hazardous chemicals and wastes" Procedia - Social and Behavioral Sciences, 14, 40-43. 2011,40-43, doi:10.1016/j.sbspro.2011.03.018.

[7] L. Molander, C. Rudén. "Narrow-and-sharp or broad-and-blunt – Regulations of hazardous chemicals in consumer products in the European Union" Regulatory Toxicology and Pharmacology, 62(3), 523-531, 2012, doi:10.1016/j.yrtph.2011.11.003.

[8] P. Sivaprakash, L.M. Karthikeyan, S. Joseph "A Study on Handling of Hazardous Chemicals in Engineering Industries" APCBEE Procedia, 9, 187-191, 2014, doi:10.1016/j.apcbee.2014.01.033.

[9] W. Skinner, Manufacturing in the corporate strategy, Wiley, 1978.

[10] L.S. Dias, M.G Ierapetritou, "From process control to supply chain management: An overview of integrated decision making strategies" Computers and Chemical Engineering, 106, 826-835, 2017, doi:10.1016/j.compchemeng.2017.02.006.

[11] M.E Porter, Competitive strategy: techniques for analyzing industries and competitors, The Free Press, 1980.

[12] A. Woinaroschy, "A paradigm-based evolution of chemical engineering" Chinese Journal of Chemical Engineering, 24(5), 553-557, 2016, doi:10.1016/j.cjche.2016.01.019.

[13] M.F. Ashby, Materials and the Environment Eco-Informed Material Choice, Butterworth-Heinemann imprint of Elsevier Inc, 2009.

[14] L. Molander, et al. "Are chemicals in articles an obstacle for reaching environmental goals? — Missing links in EU chemical management" Science of The Total Environment, 435–436, 280-289, 2012, doi:10.1016/j.scitotenv.2012.07.021.

[15] ECHA candidate list, 2018, [Online]. Available from: https://echa.europa.eu/candidate-list-table.

[16] ECHA authorisations list, 2018, [Online]. Available from: https://echa.europa.eu/authorisation-list.

[17] ECHA restricted substances list, 2018, [Online]. Available from: https://echa.europa.eu/substances-restricted-under-reach.

[18] IMDS Global automotive declarable substance list - background information, 2018, [Online]. Available from: https://public.mdsystem.com/en/web/imds-public-pages/gadsl.

[19] IPC-1752, 2018 [Online]. Available from: http://www.ipc.org/ContentPage.aspx?pageid=Materials-Declaration#1752a.

[20] IEC-62474, 2018 [Online]. Available from: http://std.iec.ch/iec62474,.

[21] IAEG AD-DSL, 2018 [Online], Available from: http://www.iaeg.com/workgroups/wg1/addsl/..

[22] K. Botsford, "Preparing for product environmental compliance a method for the madness" Product Compliance Engineering (ISPCE), IEEE Symposium on, 2010, doi: 10.1109/PSES.2010.5637803.

[23] D. Cox, A. Sweatman, "Management of restricted substances in consumer products" IEEE Xplore Conference: Environmentally conscious design and inverse manufacturing, 1999. Proceedings. EcoDesign '99: First International Symposium On, doi: 10.1109/ECODIM.1999.747669.

[24] W.E. Deming, Out of the crisis, 2nd ed, MIT Center for Advanced Engineering Study, Massachusetts, 2000.

[25] M.A. Campion, "Doing competencies well: best practices In competency modeling" Personnel Psychology, 64(1), 225-262, 2011, doi: 10.1111/j.1744-6570.2010.01207.x.

[26] L.C. Müller-Frommeyer. et al, "Introducing competency models as a tool for holistic competency development in learning factories: Challenges, example and future application" Procedia Manufacturing, 9, 307-314. 2017, doi:10.1016/j.promfg.2017.04.015.

[27] P. Hawken, A.B. Lovins, L.H., Lovins, Natural capitalism: The next industrial revolution, Routledge, 2013.

[28] S. Takhar, K. Liyanage, "Top down or Bottom up? - Supply chain engagement for material compliance reporting" Advances in Transdisciplinary Engineering, 6, 77-83, 2017, doi:10.3233/978-1-61499-792-4-77.

[29] EU Waste Framework Directive EU 2018/851, 2018, [Online], Available from: https://eur-lex.europa.eu/eli/dir/2018/851/oj.

[30] EU CLP, 2018 [Online], Available from: https://echa.europa.eu/regulations/clp/understanding-clp.

[31] EU BPR Regulation 528/2012, 2012 [Online], available from: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2012:167:0001:0123:en:PDF.

[32] EU BPR Regulation 1062/2014, 2014 [Online], available from: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014R1062&from=EN

[33] EU Conflict Minerals Regulation, 2017 [Online], available from: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32017R0821.

[34] J.C Charpentier, "What Kind of Modern 'Green' Chemical Engineering is Required for the Design of the 'Factory of Future? "Procedia Engineering, 138, 445-458, 2016, doi:10.1016/j.proeng.2016.02.104.

[35] E. Zschieschang, et al., "Resource Efficiency-oriented Optimization of Material Flow Networks in Chemical Process Engineering" Procedia CIRP, 15, 372-378, 2014, doi:10.1016/j.procir.2014.06.066.

[36] D. Dumas, L.C. Schmidt, P.A. Alexandera, "Predicting creative problem solving in engineering design" Thinking Skills and Creativity, 21,50-66, 2016, pp. 50-66, doi:10.1016/j.tsc.2016.05.002.

Table 2. Framework Tables (Names, Fields, Field Types)

Table	Field Name	Field Type	Mandatory / Optional	Comments
Article level table	[Top level article revision number]	nvarchar	Mandatory	
	[Top level article lifecycle status]	nvarchar	Optional	
	[Top level article geometry drawing ID]	nvarchar	Optional	
	[Top level article specification ID]	nvarchar	Optional	
	Comments	nvarchar	Mandatory	
	[Top level article machine program]	nvarchar	Mandatory	Required by Engineering function [see later]
	[Top level article assembly instructions]	nvarchar	Mandatory	Required by Engineering function [see later]
	[Top level article inspection data]	nvarchar	Mandatory	Required by Engineering function [see later]
	[Top level article test data]	nvarchar	Mandatory	Required by Engineering function [see later]
	[Make / buy flag]	nvarchar	Optional	Required by Engineering function [see later]
Assembly Table	[Assembly number to BOM level]	numeric	Mandatory	
	[Assembly number]	nvarchar	Mandatory	
	[Assembly description]	nvarchar	Mandatory	
	[Assembly revision number]	nvarchar	Mandatory	
	[Assembly lifecycle status]	nvarchar	Mandatory	
	[Assembly BOM qty]	numeric	Mandatory	
	[Assembly geometry drawing ID]	nvarchar	Mandatory	
	[Assembly specification ID]	nvarchar	Mandatory	
	[Comments]	nvarchar	Optional	
	[Assembly machine program]	nvarchar	Optional	Required by Engineering function [see later]
	[Assembly assembly instructions]	nvarchar	Optional	Required by Engineering function [see later]
	[Assembly inspection data]	nvarchar	Optional	Required by Engineering function [see later]
	[Assembly test data]	nvarchar	Optional	Required by Engineering function [see later]
	[Make / buy flag]	nvarchar	Optional	Required by Engineering function [see later]
Part Number Table	[Part number to BOM level]	numeric		
	[Part number]	nvarchar		
	[Part number description]	nvarchar		
	[Part number revision number]	nvarchar		
	[Part number lifecycle status]	nvarchar		
	[Part number BOM qty]	numeric		
	[Part number geometry drawing ID]	nvarchar		
	[Part number specification ID]	nvarchar		
	[Comments]	nvarchar		
	[Part number machine program]		Optional	Required by Engineering function [see later]
	[Part number assembly instructions]	nvarchar	Optional	Required by Engineering function [see later]
	[Part number inspection data]	nvarchar	Optional	Required by Engineering function [see later]
	[Part number test data]	nvarchar	Optional	Required by Engineering function [see later]
	[Make / buy flag]	nvarchar	Optional	Required by Engineering function [see later]
Component Table	[Component to BOM level]	numeric	Optional	
	[Component number]		Optional	
	[Component description]	nvarchar	Optional	
	[Component revision number]	nvarchar	Optional	
	[Component lifecycle status]	nvarchar	Optional	
	[Component BOM qty]	numeric	Optional	

Table	Field Name	Field Type	Mandatory / Optional	Comments
	[Component geometry drawing ID]	nvarchar	Optional	
	[Component specification ID]	nvarchar	Optional	
	[Comments]	nvarchar	Optional	
	[Component machine program]	nvarchar	Optional	Required by Engineering function [see later]
	[Component assembly instructions]	nvarchar	Optional	Required by Engineering function [see later]
	[Component inspection data]	nvarchar	Optional	Required by Engineering function [see later]
	[Component test data]	nvarchar	Optional	Required by Engineering function [see later]
	[Make / buy flag]	nvarchar	Optional	Required by Engineering function [see later]
Semi-Component Table	[Semi-component to BOM level]	numeric	Optional	
	[Semi-component number]	nvarchar	Optional	
	[Semi-component description]	nvarchar	Optional	
	[Semi-component revision number]	nvarchar	Optional	
	[Semi-component lifecycle status]	nvarchar	Optional	
	[Semi-component BOM qty]	numeric	Optional	
	[Semi-component geometry drawing ID]	nvarchar	Optional	
	[Semi-component specification ID]	nvarchar	Optional	
	[Comments]	nvarchar		
	[Semi-component machine program]	nvarchar	Optional	Required by Engineering function [see later]
	[Semi-component assembly instructions]	nvarchar	Optional	Required by Engineering function [see later]
	[Semi-component inspection data]	nvarchar	Optional	Required by Engineering function [see later]
	[Semi-component test data]	nvarchar	Optional	Required by Engineering function [see later]
	[Make / buy flag]	nvarchar	Optional	Required by Engineering function [see later]
Supplier Declaration Table	[Declaration ID]	nvarchar	Mandatory	
	[Supplier ID]	nvarchar	Mandatory	
	[Supplier declaration reporting status]	nvarchar	Mandatory	
	[Requested date]	date	Mandatory	
	[Respond by date]	date	Mandatory	
	[Requestor comments]	nvarchar	Mandatory	
	[Terms and conditions]	nvarchar	Mandatory	
	[Reportable substance list]	nvarchar	Mandatory	
	[Declaration detail ID]	nvarchar	Mandatory	
	[Positive declaration ID]	nvarchar	Optional	
	[Supplier sign off ID]	nvarchar	Mandatory	
Requestor Contact Details (GDPR)	[Requestor company name]	nvarchar	Mandatory	
	[Requestor individual name]	nvarchar	Mandatory	
	[Requestor address (street)]	nvarchar	Mandatory	
	[Requestor address (city)]	nvarchar	Optional	
	[Requestor address (state)]	nvarchar	Optional	
	[Requestor address (zip / postal code)]	nvarchar	Optional	
	[Requestor phone number]	nvarchar	Optional	
	[Requestor email address]	nvarchar	Optional	
	[Requestor form ID]	nvarchar	Mandatory	
Referenced DSL Table	[Declarable Substance List (DSL) name]	nvarchar	Mandatory	
	[DSL version ID]	nvarchar	Optional	
	[DSL version date]	date	Mandatory	
Declarable Substance List Table	[Declarable Substance List (DSL) name]	nvarchar	Mandatory	
	[Unique ID]	nvarchar	Mandatory	
	[Chemical substance name]	nvarchar	Mandatory	
	[CAS number]	nvarchar	Optional	
	[EC number]	nvarchar	Optional	
	[Status]	nvarchar	Mandatory	
	[Date of inclusion]	date	Mandatory	
	[Reason for inclusion]	nvarchar	Mandatory	
	[Internal priority ranking]	nvarchar	Optional	
	[Date of last update]	date	Mandatory	
	[Comments]	nvarchar	Optional	
Declaration Data Table	[Requestor article number]	nvarchar	Mandatory	
	[Requestor article name]	nvarchar	Mandatory	
	[Requestor article version]	nvarchar	Mandatory	
	[Requestor article BOM level]	numeric	Mandatory	
	[Supplier article number]	nvarchar	Mandatory	
	[Supplier article name]	nvarchar	Mandatory	
	[Supplier article version]	nvarchar	Mandatory	
	[Country of manufacture]	nvarchar	Mandatory	
	[Number of instances]	numeric	Optional	
	[Supplier article weight]	numeric	Optional	
	[Supplier article UoM]	nvarchar	Optional	

Table	Field Name	Field Type	Mandatory / Optional	Comments
	[Specification / Standard used]	nvarchar	Optional	
	[Material name]	nvarchar	Optional	
	[Material manufacturer]	nvarchar	Optional	
	[Material weight]	numeric	Optional	
	[Material UoM]	nvarchar	Optional	
	[Material use description]	nvarchar	Optional	
	[Mixture name]	nvarchar	Optional	
	[Mixture manufacturer]	nvarchar	Optional	
	[Mixture UoM]	nvarchar	Optional	
	[Mixture Mass]	numeric	Optional	
	[Mixture – on finished product / just manufacturing process]	nvarchar	Optional	
	[Chemical substance name]	nvarchar	Optional	
	[CAS number]	nvarchar	Optional	
	[EC Number]	nvarchar	Optional	
	[DSL ID No]	nvarchar	Optional	
	[Substance use description]	nvarchar	Optional	
	[Substance mass]	numeric	Optional	
	[Substance UoM]	nvarchar	Optional	
	[Substance – on finished product / just manufacturing process]	nvarchar	Optional	
	[Substance is homogenous?]	nvarchar	Optional	
	[Substance affected by RoHS?]	nvarchar	Optional	
	[RoHS exemption if applicable]	nvarchar	Optional	
	[is an article?]	nvarchar	Mandatory	
Supplier Declaration (Positive) Table	[Positive declaration ID]	nvarchar	Mandatory	
	[Supplier ID]	nvarchar	Mandatory	
	[Requested date]	date	Mandatory	
	[Supplier declaration reporting status]	nvarchar	Mandatory	
	[List of suppliers affected articles]	nvarchar	Mandatory	
	[List of substance(s) of concern on affected articles]	nvarchar	Optional	
	[Sunset date awareness]	nvarchar	Optional	
	[Sunset date of substance(s)]	date	Optional	
	[Supplier regions where substance(s) used]	nvarchar	Optional	
	[Supplier replacement strategy]	nvarchar	Optional	
	[List of exemption substance(s)]	nvarchar	Optional	
	[List of exemptions applied]	nvarchar	Optional	
	[List of exemptions received]	nvarchar	Optional	
	[List of authorisation substance(s)]	nvarchar	Optional	
	[List of authorisations applied]	nvarchar	Optional	
	[List of authorisations received]	nvarchar	Optional	
Customer Declaration Table	[Declaration ID]	nvarchar	Optional	
	[Customer ID]	nvarchar	Optional	
	[Customer contact ID]	nvarchar	Optional	
	[Internal part number(s)]	nvarchar	Optional	
	[Customer part number(s)]	nvarchar	Optional	
	[Substances identified on part number(s)]	nvarchar	Optional	
	[Chemical regulation – applicable substance list(s)]	nvarchar	Optional	
	[Chemical regulation – date of applicable substance list(s)]	date	Optional	
	[Substances identified on part number(s)]	nvarchar	Optional	
	[Customer declaration statement]	nvarchar	Optional	
	[Customer declaration transmitted date]	date	Optional	
Technical Publications Table	[Safe use guidance statement(s)]	nvarchar	Mandatory	
	[Compliance statement(s)]	nvarchar	Mandatory	
	[User guide(s)	nvarchar	Mandatory	
	[Manual name]	nvarchar	Optional	
	[Manual version]	nvarchar	Optional	
	[Publication date]	date	Optional	
Specification Table	[Specification ID]	nvarchar	Mandatory	
	[Specification name]	nvarchar	Mandatory	
	[Specification type]	nvarchar	Mandatory	
	[Specification lifecycle status]	nvarchar	Mandatory	
	[Substance ID]	nvarchar	Mandatory	
	[Mixture ID]	nvarchar	Optional	
	[Material ID]	nvarchar	Optional	
	[Specification last updated]	date	Mandatory	

Table	Field Name	Field Type	Mandatory / Optional	Comments
	[Comments]	nvarchar	Mandatory	
Substance Table	[Substance ID]	nvarchar	Mandatory	
	[Substance name]	nvarchar	Mandatory	
	[CAS number]	nvarchar	Mandatory	
	[EC number]	nvarchar	Optional	
	[Substance UoM]	nvarchar	Mandatory	
	[Substance used qty]	numeric	Mandatory	
	[Substance use description]	nvarchar	Mandatory	
	[Substance manufacturer]	nvarchar	Optional	
	[Substance lifecycle status]	nvarchar	Mandatory	
	[Substance mandatory / option]	nvarchar	Mandatory	Substance in relation to specification.
	[Substance declarable status]	nvarchar	Mandatory	Substance <> threshold level.
	[Safety Data Sheet ID]	nvarchar	Mandatory	
	[Substance last updated]	date	Mandatory	
	[Comments]	nvarchar	Optional	
Mixture / Formulation Table	[Mixture ID]	nvarchar	Mandatory	
	[Mixture name]	nvarchar	Mandatory	
	[Substance ID]	nvarchar	Mandatory	
	[Mixture UoM]	nvarchar	Mandatory	
	[Mixture used qty]	numeric	Mandatory	
	[Mixture use description]	nvarchar	Mandatory	
	[Mixture manufacturer]	nvarchar	Optional	
	[Mixture lifecycle status]	nvarchar		
	[Mixture mandatory / optional]	nvarchar	Mandatory	Mixture / Formulation in relation to specification.
	[Mixture declarable status]	nvarchar	Mandatory	Mixture / Formulation becomes declarable as it contains a reportable substance > threshold
	[Safety data sheet ID]	nvarchar	Mandatory	
	[Mixture last updated]	date	Mandatory	
	[Comments]	nvarchar	Optional	
Material Table	[Material ID]	nvarchar	Mandatory	
	[Material name]	nvarchar	Mandatory	
	[Substance ID]	nvarchar	Mandatory	
	[Mixture ID]	nvarchar	Optional	
	[Material UoM]	nvarchar	Mandatory	
	[Material used qty]	numeric	Mandatory	
	[Material use description]	nvarchar	Mandatory	
	[Material manufacturer]	nvarchar	Optional	
	[Material lifecycle status]	nvarchar	Mandatory	
	[Material mandatory / optional]	nvarchar	Mandatory	
	[Material declarable status]	nvarchar	Mandatory	Material becomes declarable as it contains a reportable substance > threshold
	[Safety data sheet ID]	nvarchar	Mandatory	
	[Material last updated]	date	Mandatory	
	[Comments]	nvarchar	Optional	
Safety Data Sheet Table	[Safety data sheet ID]	nvarchar	Mandatory	
	[Substance / Mixture / Material trade name]	nvarchar	Mandatory	
	[Manufacturer]	nvarchar	Mandatory	
	[Country of origin]	nvarchar	Mandatory	
	[Safety data sheet - Year of publication]	nvarchar	Mandatory	
	[Hazard information and labels]	nvarchar	Optional	
	[Safety precautions]	nvarchar	Optional	
	[Substance name]	nvarchar	Mandatory	
	[CAS number]	nvarchar	Mandatory	
	[EC number]	nvarchar	Optional	
	[Risk analysis ID]	nvarchar	Mandatory	
	[Risk analysis data]	nvarchar	Mandatory	
	[Safety data sheet assessment]	nvarchar	Optional	
	[Safety data sheet reviewed by]	nvarchar	Mandatory	
	[Safety data sheet review date]	date	Mandatory	
	[Comments]	nvarchar	Mandatory	
Purchasing Function Table	[Internal Number]	nvarchar	Mandatory	Top level article number / assembly number / part number / component number / semi-component number. Internal product number.
	[Make / Buy]	nvarchar	Mandatory	Internally manufactured or sourced from supply chain.
	[Supplier product name]	nvarchar	Optional	Supplier part number (SKU)
	[Purchased product description]	nvarchar	Optional	
	[Supplier ID]	nvarchar	Optional	
	[Supplier name]	nvarchar	Optional	

Table	Field Name	Field Type	Mandatory / Optional	Comments
	[Supplier contact ID]	nvarchar	Optional	
	[Supplier phone (public)]	nvarchar	Optional	
	[Qty ordered last 6 months]	nvarchar	Optional	Extracted from procurement system as an indicator of purchased volume of product
	[Last order date]	date	Optional	
Purchasing Function (GDPR) Table Sales Function Table	[Supplier ID]	nvarchar	Mandatory	
	[Supplier contact ID]	nvarchar	Mandatory	
	[Supplier contact name]	nvarchar	Mandatory	
	[Supplier contact phone number]	nvarchar	Optional	
	[Supplier contact email address]	nvarchar	Optional	
	[Supplier address (street)]	nvarchar	Mandatory	
	[Supplier address (city)]	nvarchar	Optional	
	[Supplier address (state)]	nvarchar	Optional	
	[Supplier address (zip / postal code)]	nvarchar	Optional	
	[Internal number]	nvarchar	Mandatory	
	[Customer product number]	nvarchar	Mandatory	
	[Product description]	nvarchar	Mandatory	
	[Customer ID]	nvarchar	Mandatory	
	[Customer contact ID]	nvarchar	Mandatory	
	[Qty sold last 6 months]	numeric	Optional	
	[Last order date]	date	Optional	Date of last sales order
Sales Function (GDPR) Table	[Customer ID]	nvarchar	Mandatory	
	[Customer contact ID]	nvarchar	Mandatory	
	[Customer contact name]	nvarchar	Mandatory	
	[Customer contact phone number]	nvarchar	Optional	
	[Customer contact email address]	nvarchar	Optional	
	[Customer address (street)]	nvarchar	Mandatory	
	[Customer address (city)]	nvarchar	Optional	
	[Customer address (state)]	nvarchar	Optional	
	[Customer address (zip / postal code)]	nvarchar	Optional	
Stores Function Table	[Internal number]	nvarchar	Mandatory	
	[Supplier product name]	nvarchar	Optional	
	[Manufacturer]	nvarchar	Mandatory	
	[Country of origin]	nvarchar	Mandatory	
	[Hazard information and labels]	nvarchar	Optional	
	[Safety precautions]	nvarchar	Mandatory	
	[Storage location]	nvarchar	Mandatory	
	[UoM]	nvarchar	Mandatory	
	[Amount held in stock]	numeric	Mandatory	
	[Last work order]	nvarchar	Optional	Last work order where issued
	[Amount issued]	numeric	Optional	
	[Where issued]	nvarchar	Optional	
	[Date issued]	date	Optional	
	[Reorder level]	numeric	Mandatory	
Transportation Function Table	[Internal number]	nvarchar	Mandatory	
	[Customer ID]	nvarchar	Mandatory	
	[Customer product number]	nvarchar	Optional	
	[Packaging ID]	nvarchar	Optional	
	[Packaging Type]	nvarchar	Mandatory	
	[Packaging Material]	nvarchar	Optional	
	[Dimensional Data]	nvarchar	Optional	
	[Packaging weight]	nvarchar	Optional	
	[Labelling data]	nvarchar	Optional	
	[Documentation]	nvarchar	Mandatory	
	[Customer contact ID]	nvarchar	Mandatory	
	[Dispatch method]	nvarchar	Mandatory	
	[Dispatch reference]	nvarchar	Mandatory	
	[Dispatch date]	date	Mandatory	
Transportation Function (GDPR) Table	[Customer ID]	nvarchar	Mandatory	
	[Customer contact ID]	nvarchar	Mandatory	
	[Customer contact name]	nvarchar	Mandatory	
	[Customer contact phone number]	nvarchar	Optional	
	[Customer contact email address]	nvarchar	Optional	
	[Customer address (street)]	nvarchar	Mandatory	
	[Customer address (city)]	nvarchar	Optional	
	[Customer address (state)]	nvarchar	Optional	
	[Customer address (zip / postal code)]	nvarchar	Optional	
	[Quality procedure ID]	nvarchar	Mandatory	

Table	Field Name	Field Type	Mandatory / Optional	Comments
Quality Function Table	[Applicable function name]	nvarchar	Mandatory	
	[Quality procedure name]	nvarchar	Mandatory	
	[Quality procedure details]	nvarchar	Mandatory	
	[Quality procedure status]	nvarchar	Optional	
	[Cross reference standard]	nvarchar	Optional	
	[Quality procedure last reviewed date]	Date	Mandatory	

Numerical Simulation of One Pavement Structure of Polyethylene Terephthalate Submitted to Static Point Loads

Juan Sebastián Gilart Pulido[*], María Camila Zambrano Bello, Ender Jhobany Orduz Duarte

Program of Civil Engineering, Universidad Católica de Colombia, 111311, Colombia

A R T I C L E   I N F O

A B S T R A C T

Plastic is highly polluting nowadays and is found in many disposable elements used for people. Plastic materials are disposed inadequately and large quantities of them become pollution in the ground, water, and air, affecting plants and animals. Most of the plastic is Polyethylene Terephthalate (PET) which is used in many activities. Resistance, support capacity, and flexibility are some characteristics that make PET useful in road engineering. This research shows a numerical simulation of three kinds of structures for pavements in PET. The simulation uses 3D models that were analyzed in the software SAP2000. The material characteristics of the recycling PET were taken from industries and analyzed in the laboratory. The model used in the simulation was composed of vertical loads for traffic type A. Loads were applied in the border, 0.75 meters and 1.5 meters from the border. Three types of structures were analyzed with three, four and five supports respectively. Structure type 2 with four supports displayed the best results regarding vertical displacements in all the elements. Structure type 2 was simulated with six different thicknesses (0.06, 0.08, 0.10, 0.12, 0.14 and 0.16 meters). Optimization of this material and the resistance of the structure showed the 0.10 meters thickness as the best due to its admissible vertical displacements and required the lowest quantity of material for producing it.

1. Introduction

Transporting people and resources are two of the most important elements of economic development in society. Comfort, security, and speed are three parameters required for users of the roads. The concrete asphalt is one of the most popular materials used in pavements in order to have high quality and adequate traffic service nowadays [1]. Environmental damage caused by asphalt production and building roads in concrete asphalt is an important topic studied lately. Production, transport, and storage are activities that produce significant pollution. The Environmental Protection Agency of the United States has pleaded worldwide activities to underwrite production controls, environmental impact reduction, and more policies of environmental care around the world. Researches account more than 500000 kilograms of emissions to the atmosphere in pollution with volatile compositions, organic materials and so on [2]. Environmental damage is an important topic at the present time and society should find new strategies for controlling all different kinds of pollution in it. The traffic and the transport of resources and materials are activities that induce high quantities of pollution, especially into the air. Using synthetic materials and other kinds of additions for making the road`s concrete is one of many development possibilities to build road infrastructure

around the world [3,4]. The addition of polymeric materials to concrete are implemented in order to reduce the volume of waste in landfills and to recycle different kinds of polymeric elements. Sustainable activities have served as a way of making new construction materials and resources for customers [5,6]. Reusing materials, used currently, could be another strategy to reduce humans' impact on the Earth [7].

Polymers are versatile materials and many products are made with it. However, polymers are also highly polluting materials because a large number of disposable elements are used daily and not all of them are recycled or reused in new activities. As a result, those elements are transported to the landfill and increase the contamination problems in land, water, and air; it can attack different species of animals in nature. Synthetic polymers are potentially polluting because their degradation time is so long and the oil is an irreplaceable resource. Also, the chemical composition and the high quantity of waste during the production is a considerable issue [8].

This research shows a strategy proposed for making pavement structures in recycled plastic. The Polyethylene Terephthalate is a material easy to find in the modern world. Some alternatives made in the past have identified materials used for different activities and their characteristics can offer functionality and quality service

[*]Corresponding Author: Juan Gilart, Email: jsgilart@gmail.com

in roads infrastructure as a pavement structure. The process shows three kinds of structures evaluated in the software SAP2000 with the finite element methodology. The simulated material is recycled plastic, a unique resource for the structure. The software seeks to evaluate the model of the structures with the application of static and point loads.

3. Synthetic polymers materials and their recycling

Gaelle Gourmelon research, "Global Plastic Production Rises, Recycling Lags," discusses the behavior of the polymeric industry around the world during the period 1950-2012. This industry produces more than six hundred thousand dollars each year. Transport, construction, health, education, and food industry are some of the sectors of the economy where plastic is highly used. For example, for making one automobile the industry uses 152.41 kilograms of plastic, while in 1960 it only used 9.07 kilograms. Of course, technology advances look for high quality and less cost: the automotive industry uses polymers in important parts of the cars. During 2013, the US automotive industry nearly used 300000 million kilograms of plastic materials in automobile parts [9].

During 2012, Europe recycled 26% of its polymeric materials. Additionally, during the same year, 36% of this material was incinerated. Around 38% of the polymeric materials were taken to landfills. Austria, Belgium, Sweden, and Switzerland governments regulate recycling, and citizens are required by law to recycle plastics and other materials. These countries have the highest rates for recycling plastic materials and most of those materials are transformed or incinerated [9].

Microplastics, little pieces of plastic in millimetric size, are highly polluting. These microplastics have become a major problem because they are scattered easily and are potentially dangerous for animals and the forest [10]. Around 275000 million kilograms of plastic waste were generated in 2010 in 192 countries with coasts. It is estimated that 12700 million kilograms of that garbage polluted the ocean [11]. Some estimations calculate around 268900 kilograms of garbage floats on the ocean's surface, without considering the material below the surface [12].

In Colombia, the DNP (by its acronym in Spanish Departamento Nacional de Planeación) estimates that 11600 million kilograms of garbage is produced annually and just 17% is recycled. Some policies are applied in order to reduce the impact of this waste on the environment. The DNP generates the Integral Policy for Management of Solid Waste with the horizon in 2030. This policy calls for the implementation of one official plan for the integral management of solid waste as a public policy, with interest for all the population, in order to foster an economic and sustainable development, transformation of the society and mitigation on the environmental impact in the country, and of course, for all the planet. Research shows that 2800 people work in Colombia in the recycling business, from the collection of pieces to fiber production in recycling factories [13]. Not only are they factory workers, where the principal resources are the recycled plastic, but most of them are not formal employees. This topic is predominantly studied in socio-economic research because it is an important aspect of a country's economic development. Furthermore, ground pollution is another environmental problem.

Plastic materials degrade over a long period of time and cause ground pollution that affects the growth and development of animals and plants [14]. This prediction of a sanitary emergency in most of Colombia's cities by 2030 highlights that the majority of the waste in Colombia is disposed in landfills, and eventually, those landfills will not have capacity for any more garbage. Considering the DNP`s information in 2016, Colombia needs better waste management, to change its consumption of resources and to transform its industries as soon as possible, in order to promote a productive and sustainable economy.

4. Reused polyethylene Terephthalate as an absolute material for pavement structures

Roads are important for economic development: industries, tourism, commerce, agriculture, and so on, are important sectors in developing countries. High quality roads permit faster traffic, fewer transportation costs and less pollution from emissions into the air. Today, one of the most important elements for defining economic development is the investment in quality road construction, faster traffic and safer roads: the base for increasing productivity [15]. Synthetic polymers are used frequently for making reusable materials in construction. Masonry is one of the most important activities where plastic is used. Recent research has attempted to use bricks made with plastic in building construction in order to reduce the impact of plastic on Earth. Polyethylene Terephthalate is the most prevalent polymer to modify construction materials. Using plastic bricks in buildings curbs challenges in traditional construction methods in developing countries and reduces the environmental impact of garbage in landfills [16]. Polymers are also frequently added to an asphalt mix. The addition of polymers to asphalt concrete changes its characteristics, especially the mechanic behavior, durability and thermal susceptibility [17]. Likewise, polymers are used as concrete aggregates. This kind of mixture changes their resistance and durability. However, this material should be used carefully in big buildings because of changes in their composition [18].

Polyethylene Terephthalate has valuable qualities, especially its durability and resistance to abrasion, it's an adequate answer to chemical attack, resistance to weather conditions and can be used in construction. They show better conditions and higher potential use in the construction activities compared to with other materials [19]. Construction factories around the world use commonly plastics in many applications. New infrastructure projects use lighter materials made of polymers to construct architectural finishes and residential public service networks. Nevertheless, recycling is not a frequent activity in developing countries, despite it's easy management and material characteristics. This research shows a new possibility for recycled plastic to develop road projects in places of light traffic loads. Also, the research as a first approximation in one type of structure developed only in plastic. Using recycled plastic in high quantity helps society in two important ways: first, in development of new materials in pavements for new road projects; second, plastic pavements demand high bulk from plastic materials that reduce the quantity of waste in landfills.

5. Characteristics of the model

Different kinds of structures are used in road pavements. Asphalt concrete is the most used material in developing

countries. Also, rigid concrete made with cement is used frequently. Another different pavement called articulated pavement uses masonry on the surface of roads with light traffic. All of these kinds of structures are constructed by layers. In the bottom of all projects, subgrade material is made up the natural soil. Sometimes, low resistance materials are changed or modified with some additions. Over the subgrade are used arid material, produced by rock's trituration (natural or artificial), builded in layers according designs. The top layer normally is made of concrete or masonry. The research proposes to change classic pavement structures of asphalt concrete or hydraulic concrete to plastic elements over arid layers. The Figures 3, 4 and 5 show some possibilities of structures where the plastic is in the top and in the bottom, but inside the structure is empty and there are ducts that can be used as channels for water or electrical or hydraulic networks.

The model presented was developed in pieces with length and width of 3 meters, a total height of 0.5 meters, and a horizontal and vertical thickness of 0.14 meters. The proportions used were taken from base recommendations for rigid pavements that INVIAS (Instituto Nacional de Vías) proposes for these types of pavements [20].

The Colombian norm NTC 237 recommends tests in density and absorption in plastic materials as Polyethylene Terephthalate. This research tested recycled Polyethylene Terephthalate from plastic bottles. Its density was 1219 kilograms per cubic meter and the absorption 0.28%. Figure 1 shows the material tested in the laboratory at Universidad Católica de Colombia in 2017.
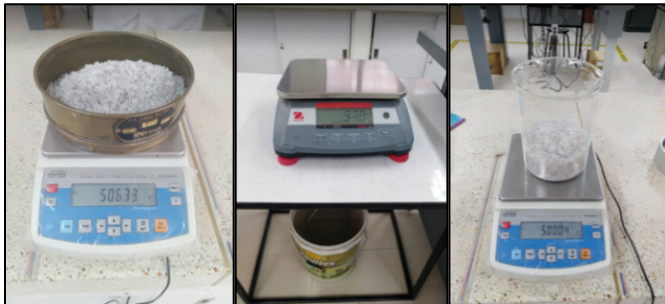

Figure 1. Polyethylene Terephthalate tested in the laboratory.

A preliminary model proposed recycled plastic for the entire structure as a first approximation through numerical simulation. Its specific weight was 1219 kilograms per cubic meter, elasticity 3.70^8 Pascal, Poisson distribution 0.37 and thermal expansion 80^{-6} Kelvin^{-1} [21]. Density and other characteristics were analyzed in the civil engineering laboratories of Universidad Católica de Colombia. Plastic material was tested in flexion in Test Three Points with a prototype in reused polyethylene terephthalate. The dimensions were: length 0.25 meters, width 0.05 meters and 0.01 meters thickness.

Prototype's flexion test was developed in an electronic universal machine and the graphics load-deflection were extracted from its software. Figure 2 shows the graphic load vs displacement of the material. Plastic behavior is in the linear part of the graffic between 0 meters to 0.03 meters deformation. The load applied for this deformation was 380 N.
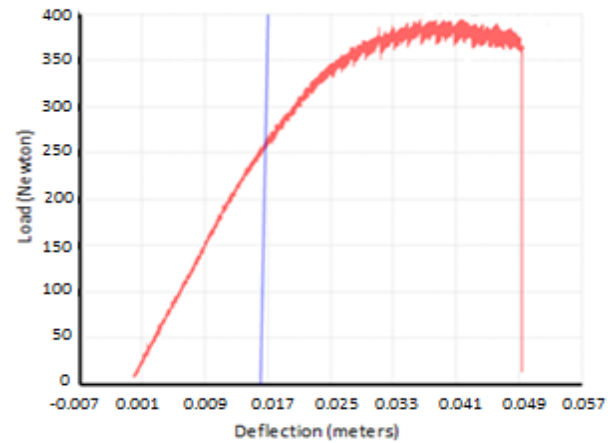

Figure 2. Load vs deflection in flexion test.

The laboratory results support the inference that recycled plastic can be used in the entire structure and the process presupposes the material offers resistance, flexibility, and functionality in structures of pavement.

5.1. Structure type 1.

Figure 3 shows the Structure type 1 model. This included three supports between the top and the bottom. Each support is 0.14 meters wide and as long as the element. Two of the supports are in the borders and the other is in in the middle.
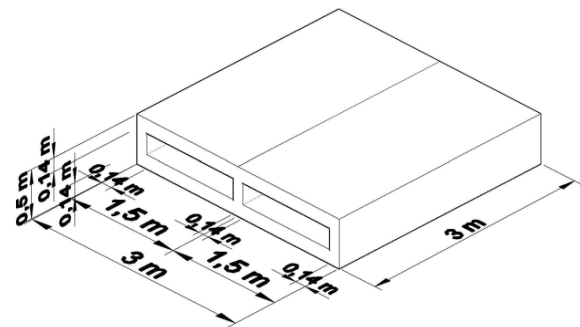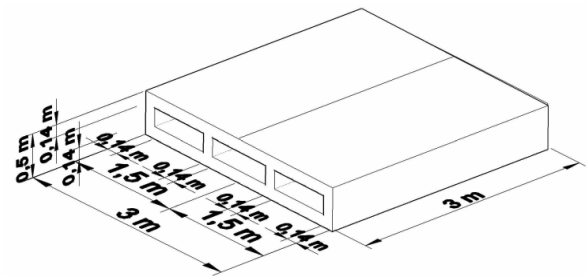

Figure 3. Structure type 1.

Figure 4. Structure type 2.

5.2. Structure type 2.

Figure 4 shows the Structure type 2 model. This included four supports between the top and the bottom elements. Each support is 0.14 meters wide and as long as the element. Two of the supports are in the borders and two are one-third (1.0 meters) to each border.

5.3. Structure type 3.

Figure 5 shows the Structure type 3 model. This included five supports between the element in the top and the bottom. Each support is 0.14 meters wide and as long as the element. Two of the supports are in the borders and the other one is in the middle. Also, two of the supports are in the quarters (0.75 meters) to the borders.
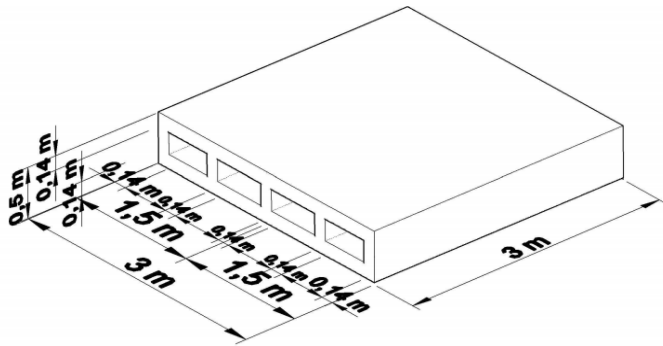

Figure 5. Structure type 3.

6. Simulation.

The process developed in this phase of the research was the simulation in the software SAP2000. This software uses the methodology of limit elements and with a sophisticated, intuitive and versatile tool used in design for engineering in different applications. This software enabled the creation of 3D elements for modeling analysis, including research parameters used to simulate one structure in plastic as real as possible. This is the first approximation for modeling structures without the arid layers of support. The subgrade was assumed to be a non-deformable material; it was an initial hypothesis for modeling, albeit the majority types of soils are deformed under loads. However, due it begin the first model, research analyzed only the element in plastic. The software permitted analysis of displacements in many points of the structure because the model can be divided in a high quantity into little cubic elements. Parameters are included in SAP2000 and the network is developed with 4 nodes.

Once the model is made, all elements can be divided into a network, and each node permits the identification of displacements under loads in different positions in the structure. This is achieved in multiple steps. First of all the element is divided into multiple series of elements. The series are determined by the required precision. The precision was connected by a discrete number of nodal points. Subsequently, a matrix is developed. The matrix relates variables in the nodes for each one of the divisions. Afterwards, the matrix is combined in a set of algebraic equations that describe the behavior of the structure analyzed. Finally, the set of algebraic equations is solved by a matricial method [22].

Elements from the general structure are drawn with shells. In the software, each node of the top shell has five grades of liberty with deformation. Three of them are translations (axes X, Y, and Z) and the other two are rotations (axes X and Y) [23].

Characteristics of shell elements showed the closest for the analysis, so the model was drawn in shells for this research.

Analysis of deformations in Polyethylene Terephthalate with loads over the surface can be easily studied and it is possible in its study to measure reactions, deformations, and strengths in each node of the structure. Furthermore, elements were defined as "thin" based on the relation between thickness and length; the relation showed the length was over 20 times greater than the thickness. This relation is important to determine if elements work thin (more flexible) or thick (more rigid). This definition permits to SAP2000 the highest detail in the calculations developed in numerical simulation [24].

The element in the bottom was defined with restrictions of all displacements. For this simulation the soil is non-deformable and the structure does not suffer other effects. Using this assumption, though not real, provides the possibility for the first approximation of simulation. The target with this assumption is to analyze only the structure in recycled plastic and not other materials. Also, SAP2000 is a tool used in materials with different behavior as shown by the soil. Figure 6 shows the structures type 1, 2 and 3 modeled in SAP2000 with restrictions as supports in the bottom's element.
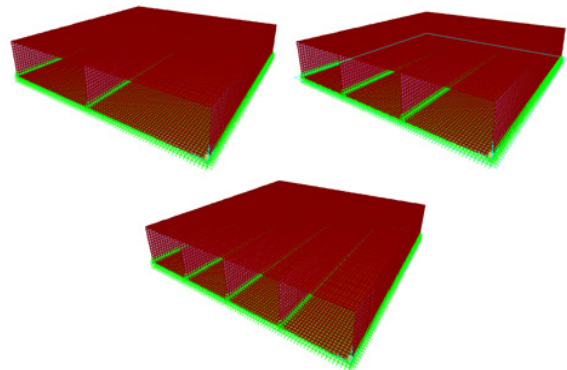

Figure 6. Simulation conditions of structures type 1, 2 and 3.

Loads applied in the simulation were defined in accordance with INVIAS to point load generated by trucks with a single axle and double tire. The norm defines 11000 kilograms for this axle divided into two point loads over the pavement with 58000 Newton each tire. This value is assumed even though it is superior to vehicles type A. The loads in real parameters could be present in tracks for passengers and light loads.

Loads applied simulate fires that produce point loads in each axle over top shell in the structure. Model in SAP2000 permits the location of loads in each node of the mesh. The structure was divided into nodes each 0.15 meters and one load per simulation. Using this mechanism shows the highest detail as possible, in order to identify the place where the structure could be more vulnerable. Furthermore, the load was applied in the border, in the middle (1.5 meters of the border) and each quarter (0.75 meters of the border) in the parallel direction of traffic as is shown in the Figure 7. Also, each load was simulated independently as is shown in Figures 8, 9 and 10 for structures type 1, 2 and 3. Each kind of structure was analyzed with the same parameters of the load in order to do the comparison of vertical displacements in the top for the three types of structures. Logically, it is possible to think that Structure type 3 would have fewer displacements. However, the intention with all simulations is to identify the

structure with permissible deformations using the minor quantity of plastic in its construction.
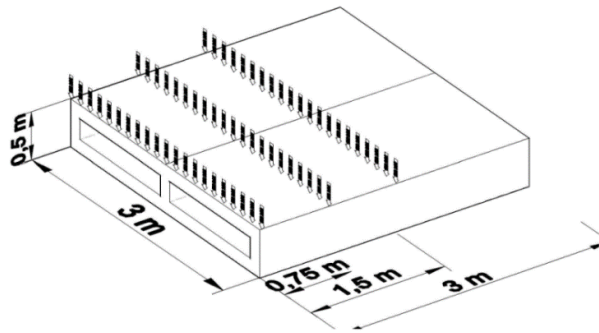

Figure 7. Isometric of application to loads on structure type 1.

Developing the simulations required 189 times with one load. The output information by SAP2000 was saved in data processor Excel and the relevant information was extracted. This information is selected to compare each node network.
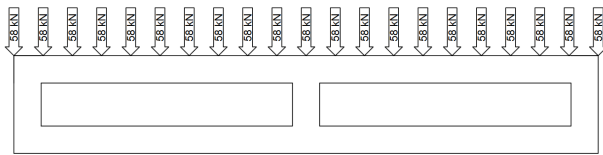

Figure 8. Loads application in cross-sectional view in structure type 1.
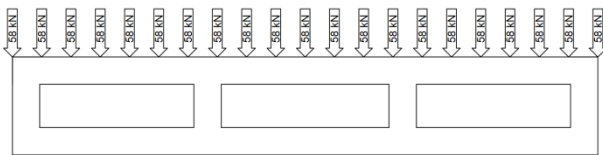

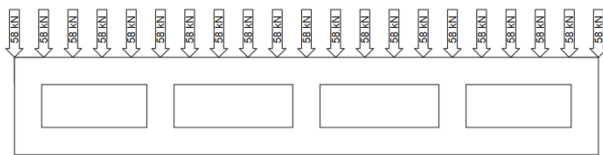Figure 9. Loads application in cross-sectional view in structure type 2.

Figure 10. Loads application in cross-sectional view in structure type 3.

7. Results and analysis

Vertical displacements in the nodes are shown in a discrete network. The point load (58000 N) was applied over the top surface in the structure for each simulation. The information permits the comparison of behavior in all the points in the structure when the load is applied to different points over it. The purpose was to identify the critical point of the structure, or the place where the highest displacement is shown and defined by whether it is possible to develop a plastic structure in these conditions of simulation. Figure 11 shows an example of how the structure was analyzed, each simulation in two axes longitudinally and transversely.

Vertical displacements are typical behavior from plastic materials loaded with gravitational forces. The load perpendicular to the structure and deformations are shown for all of the structure. Some of the points display performance compression and others traction. Deformations are produced by a thermodynamic process.

Energy is accumulated internally and is transformed in elastic potential energy. So, it is important to know the elasticity module and use it correctly in order to find high accuracy.
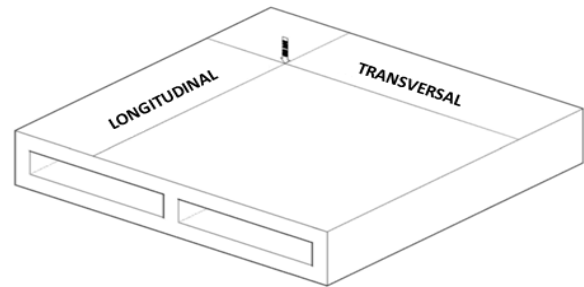

Figure 11. Axes for measuring results.

Figure 12 shows, without a real scale, how one load can affect entire structure. So, it is necessary to analyze all points and to compare displacements at the same point in the structure. This research shows displacements in the top. There are more vulnerable places where the structure could fail.
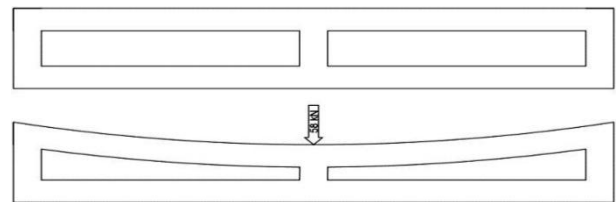

Figure 12. Model of displacements cross-sectional in structure type 1.

The same load reviewed in the direction parallel to the traffic, it has a different behavior in accordance with the point where is applied the load. Figure 13 shows one model without scale about how are the displacements in the structure when the load is applied to the border of the structure.
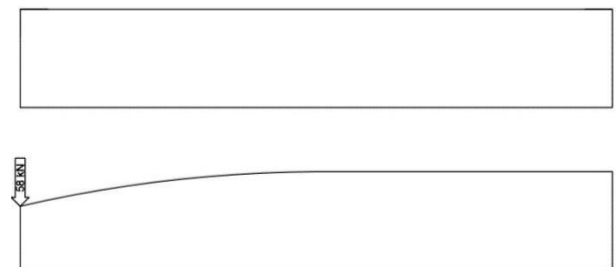

Figure 13. Model of displacements in the parallel direction to the traffic in structure type 1.

Models used were defined in three cases. Case A, when the loads are applied in the border. Case B with load on 0.75 meters of the border and Case C with loads on the middle of the structure (1.5 meters of the border). Vertical displacements in structure type 1 were taken in the three cases and compared in Figure 14. It is possible to identify that in the cases B and C, vertical displacements are shorter in comparison with case A. The results obey to the continuity of the material. For example, in the border (case A) deformations in the middle of the empty space are more than 0.0035 meters. While in the quarters (case B) or in the middle (case C), vertical displacements are around 0.00025 meters. In other words, vertical displacements are between 60% and 73%

less in continue material. In this case, it is important to identify different mechanisms for doing a complete structure in order to transfer the loads adequately and get less deformation in the borders of the elements. It is possible to identify also that on the supports, vertical displacements are depreciable; the solid material shows the high resistance of contraction to continue material.

Consequently, when it analyzed the structure type 2, the behavior is completely similar to the first type of structure. Deformations on supports are very small and in the middle of the empty spaces are highest as is shown in Figure 15. In this case, three cases are analyzed as well. In the border (case A) in a quarter to 0.75 meters (Case B) and in the middle to 1.5 meters of the border (case C) all of them were studied by the same magnitude of load but the difference is the number of supports throughout the structure.

In the Figure 15 can be seen that vertical displacements are significantly smaller than structure type 1. In the border, the highest deformation was around 0.00168 meters. Also, the middle vertical displacements are around 0.00025 meters. The magnitudes permit to infer that this type of structure can have an adequate behavior for difficult conditions with important

applications in roads engineering. Deformations conserve the same relation between 60% to 74% less in the quarter (case B) and the middle (case C) in comparison with the border's vertical displacements.

Clearly, it is easy to identify that where the material is homogeneous as under the supports, the vertical displacements are considerably minors than in the border where it showed the highest deformation. Furthermore, with this results in the border analyzed vertical displacements, it is possible to infer that the behavior is completely defined by the boundaries condition and especially when the material is discontinuous. Making a solid structure is a possibility for fewer deformations. However, it is important in this phase of the research to identify the behavior in all the structure because of the necessity of force's transference in completely necessary in roads engineering.

Structure type 3 was analyzed in the same way. The case A with loads throughout the border, punctual loads on 0.75 meters of the border (case B) and in the middle of the structure (case C). Figure 16 shows the envelope of vertical displacements in each case. The
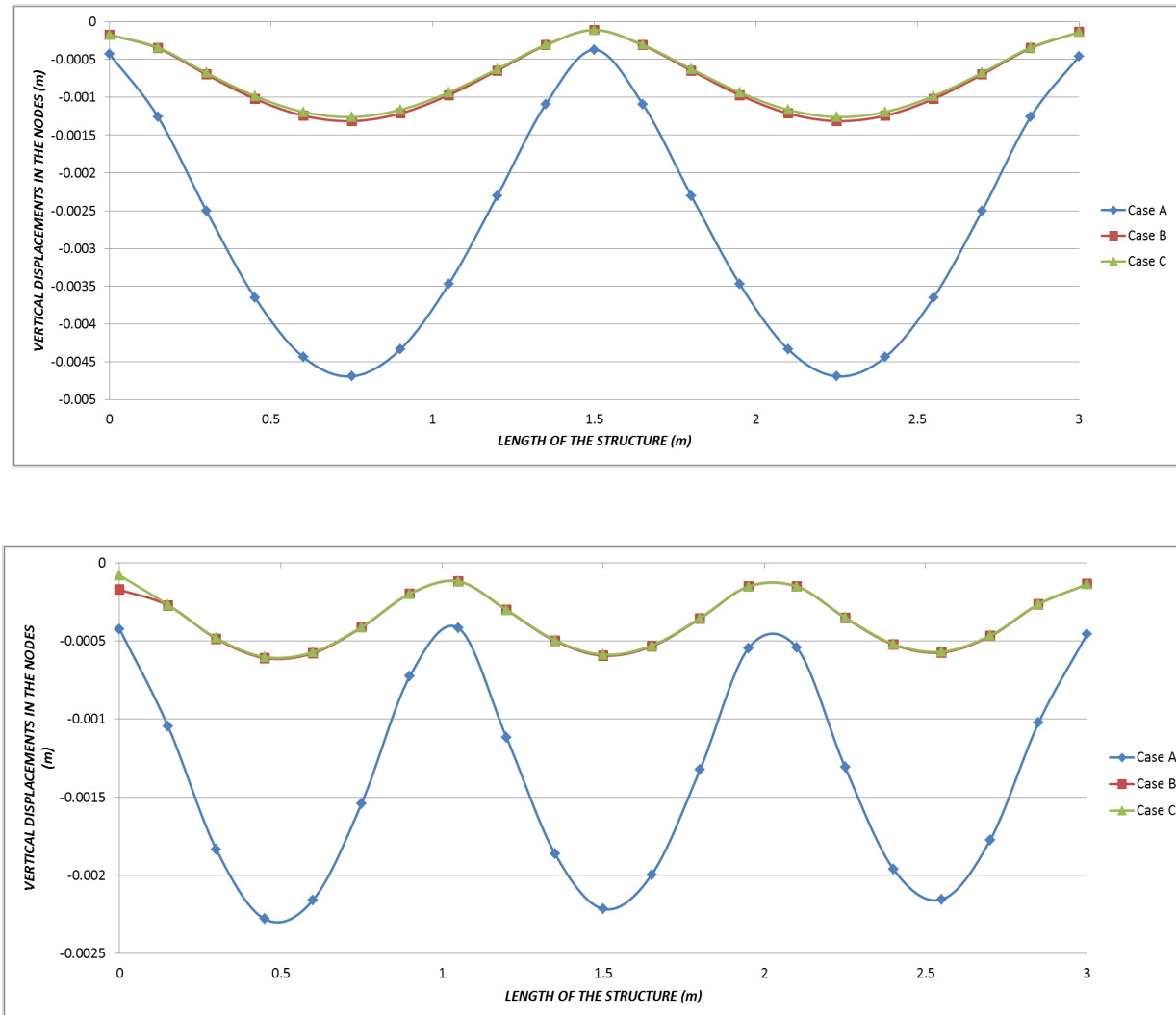

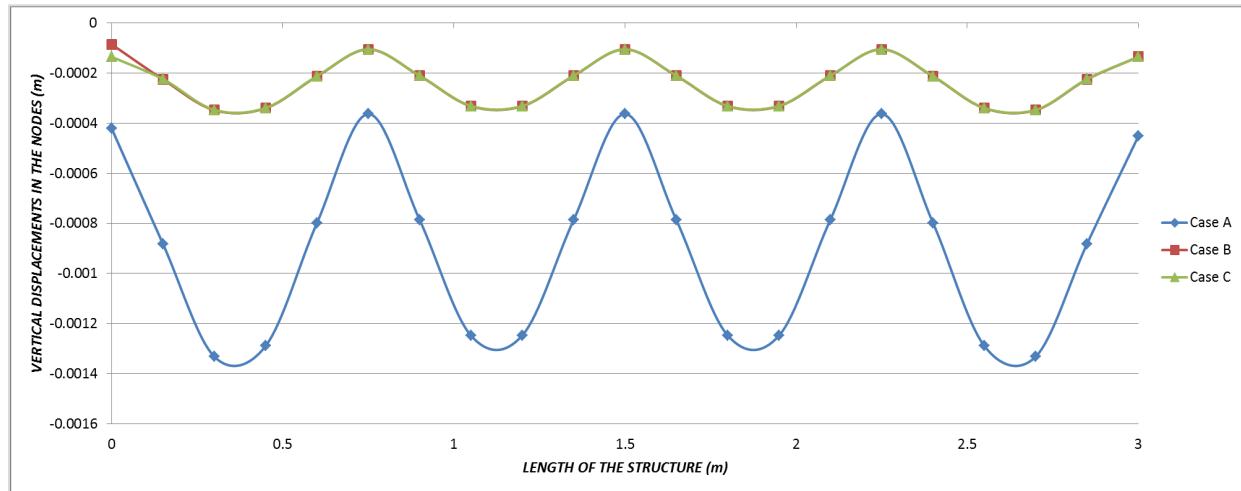Figure 15. The envelope of vertical displacements in structure type 2.

Figure 16. The envelope of vertical displacements in structure type 3.

tendency of its behavior is the same. Negligible deformations in the supports and the highest vertical displacements in the middle of the empty space. Clearly, the quantity of this results shows that vertical displacements descend in the middle until near to 0.001 meters. The deformation is depreciable if is analyzed by traffic loads type A. Structure type 3 could be utile for traffic heavier and could be used in national roads as possible. However, it is necessary for other possibilities, to modify the magnitude of loads and to make the simulation.

It is possible to say that the structure type 3 has the highest efficiency of strengths distribution. The results, showed in the envelope, permit to identify vertical displacements in the critical points in structure type 3 are more than 70% inferiors than the critical points in the structure type 1. The comparison shows the results that were expected because of the number of support the more resistance of all the structure. Clearly, it is important to define the material is plastic and the behavior analyzed was defined as a continuous and homogeneous material.

Structure type 1 shows a behavior with highest deformations in the borders (case A) when the load is applied in the middle of the empty space. Likewise, structure type 2 shows a similar behavior, but in comparison with the type 1 structure its vertical displacements are reduced by around 28%. Structure type 3 demonstrates the best behavior with vertical displacements; it shows a reduction of 41% compared with structure type 2 and 72% compared with structure type 1 in the same conditions of loads.

Furthermore, it is possible to say that is unnecessary to use more supports in the structure, considering the traffic loads type A. It is important to clarify that the material should be homogeneous and the construction should guarantee high quality in the structure in order to get the best behavior in a prototype. Tested in laboratory or in camp tests.
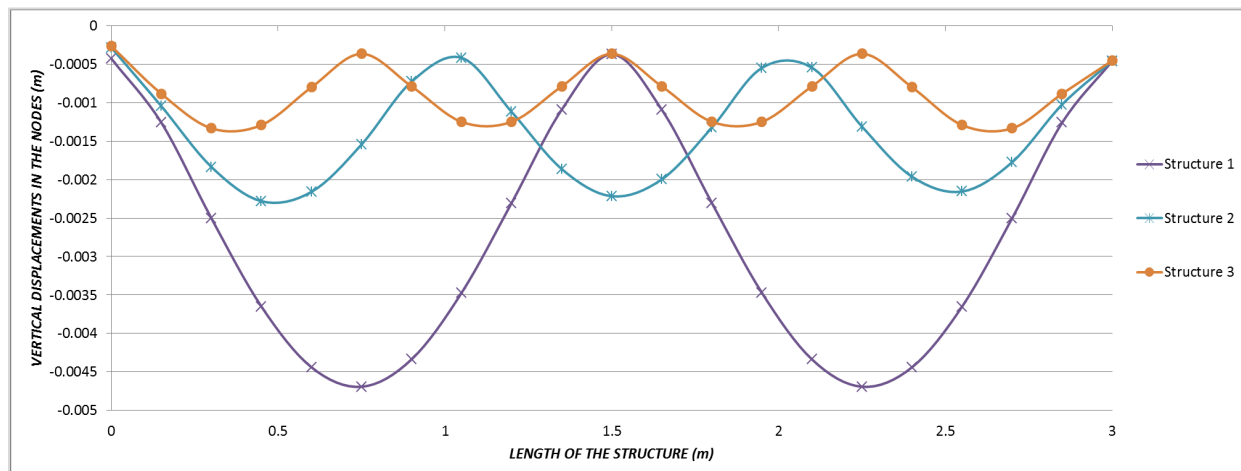

Figure 17. Comparison of envelopes of vertical displacements in structure type 1, 2 and 3.

8. Model of optimization

During the process of simulations, it was possible to identify that the structure type 2 displays the necessary resistance for supporting these loads with vertical displacements in the critical points that is a considerably low value. After the analysis of this material in the laboratory, it is possible to infer that Polyethylene Terephthalate can be deformed 0.03 meters and does not lose its plastic characteristics.
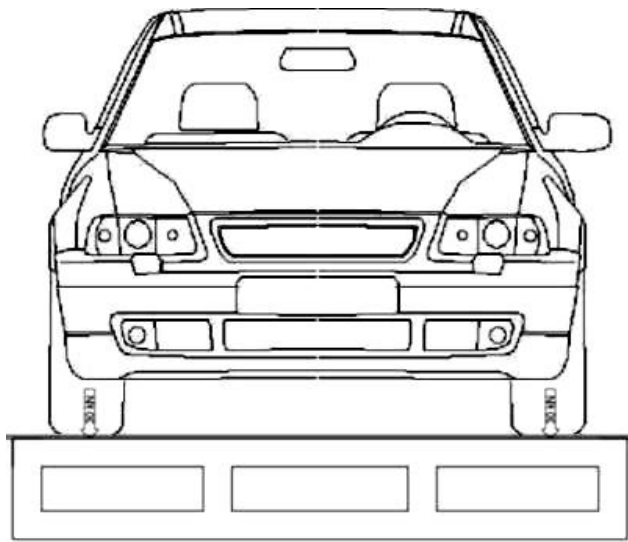

Figure 18. Model of charge for traffic type A.

Furthermore, the comparison of the supports of all vertical displacements show that both types of structures have similar behavior. Consequently, the type 2 structure was selected for developing another simulation, changing the thickness of the structure and reviewing its behavior. Models of simulation proposed one model of charge with one automobile, which induces punctual vertical loads over the structure as is shown in Figure 18. Each load is defined as 30000 N and separated 2.25 meters, according to the description of cars dimensions in the INVIAS manual.

The analysis was developed in the border of the structure (case A). The point where the critical conditions for loads were found. The new simulation was defined with six widths of structure: 0.06, 0.08, 0.10, 0.12, 0.14 and 0.16 meters respectively. Figure 19 shows vertical displacements in the critical points of the structure for the six versions of the model. As was identified in the simulation of critical conditions, it is possible to determine that the structure shows the highest deformation under the point where the loads are applied. This model applied two loads simulating the fires of a car. The central space of the structure showed positive displacements; in that segment the structure permits flector moments to transfer strengths, and the displacements change to the other size of reference ax.

A structure with 0.06 meters of width shows the highest vertical displacement 0.01303 meters. Even though this deformation is high in the simulation, if it is analyzed in real conditions, deformations from 0.013 meters are not very significant; it is possible that one structure has irreversible deformations for the frequency of the load's application.

The structure with 0.08 meters shows deformations of 0.0058 meters. For the 0.10 meters of width, it showed 0.00291 meters. With the structure, 0.12 meters is found 0.00117 meters of vertical displacement. The structure with 0.14 meters of width 0.001001 meters. And finally, for 0.16 meters in simulation, the structure showed 0.000759 meters of deformation in the borders.

Figure 20 contains critical deformations in all the type 2 structures with different widths. The loads were analyzed parallel to traffic direction. It is possible to infer that for applications on a real scale, it is necessary to identify one mechanism for transferring strains.
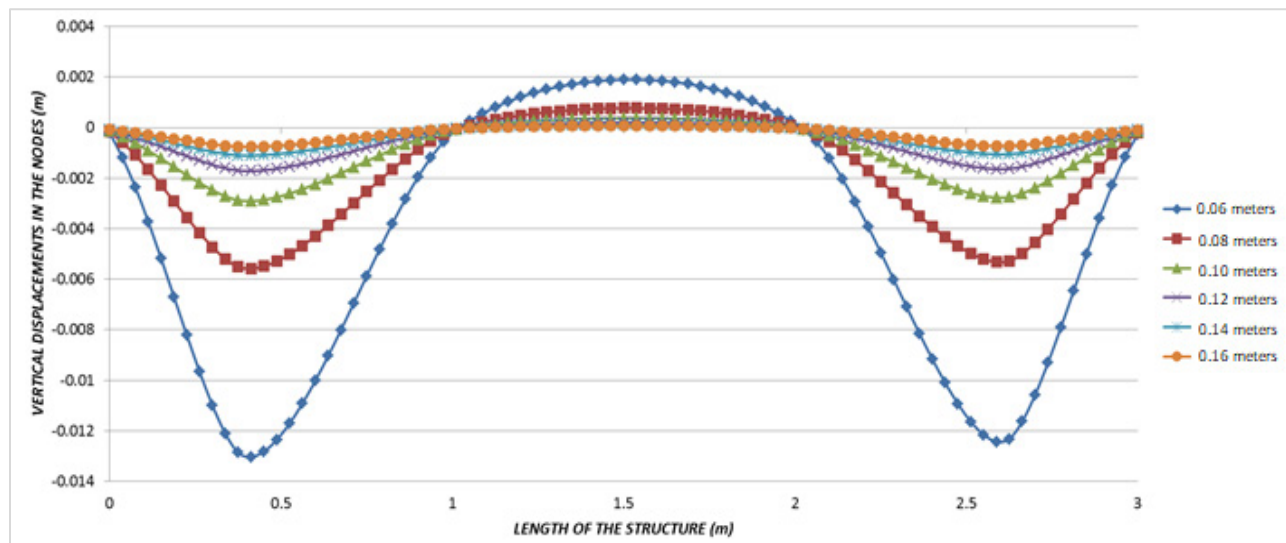

Figure 19. Enveloped of vertical displacements for structure type 2 with different width cross-direction to traffic.
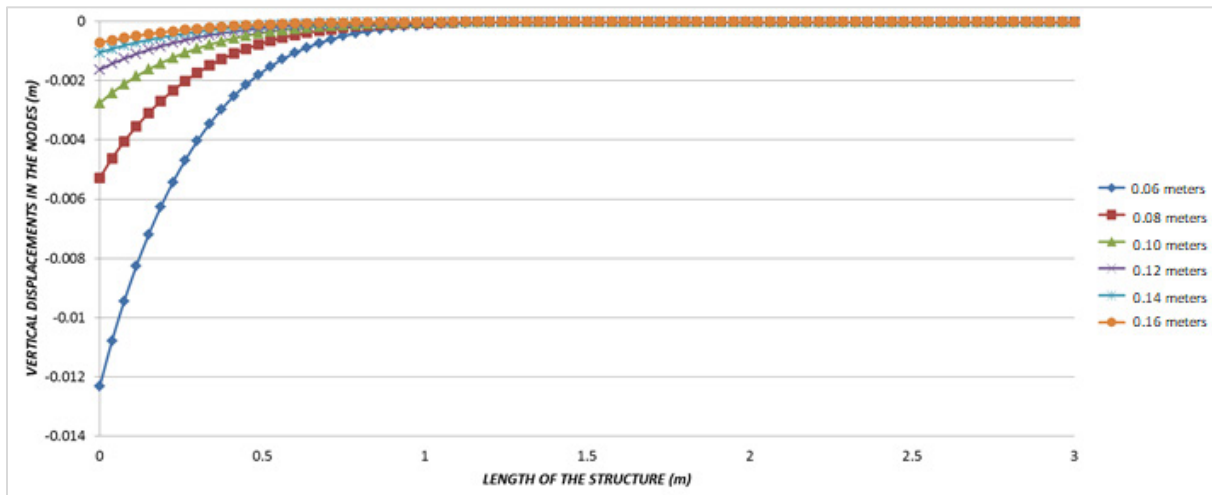
Figure 20. Enveloped of vertical displacements for structure type 2 with different width parallel to traffic direction.

Conclusion

Polyethylene Terephthalate highly pollutes the environment, and is frequently disposed without control in landfills or incinerated. Because of its characteristics this material can be recycled and transformed for a variety of uses while reducing its polluting potential. Polyethylene Terephthalate`s high resistance and flexibility inspire studies into its potential uses. The research proposes developing rolling surfaces in pavement structures.

The simulation offers considerable results according to the identification of different types of structures and different width in one of those structures. Structure type 2 showed important results because the vertical displacements are permissible and it permits optimization of the quantity of material required for making each element. Different widths show the structure supports without difficulties of traffic loads type A and would support loads type B in some cases.

Admissible deformations were the basis for changing the model of structures and the width in the last exercise where it was possible to allow for making test prototypes. The critical case shows potential use in cycle paths, walk paths and small streets for automobile vehicles. However, it is necessary to do more studies in resistance, prototypes and especially in protection against fire.

Empty spaces inside the structures are functional. It is possible to use these spaces for electrical or hydraulic ducts or for another kind of service in the structure.

Case A was the most critical situation in all simulations during the research. These results obey the discontinuity of the element in the borders. The results open new ways for research in how transfer charges from one element to the next. The process, even though it would be for construction, helps research to identify different possibilities for making the structure as a prototype or on a real scale.

Another research possibility is the characteristics of this behavior when the support of all the structure is not rigid. Simulation was developed using soil deformations under the structure. For traffic type A, these deformations are non-considerable, but in a real process of analysis, it is necessary to introduce the most variables as possible.

References.

[1] A. Gutiérrez, "¿Qué es la movilidad?. elementos para (re) construir las definiciones básicas del campo del transporte", Bitácora Urbano Territorial, Universidad Nacional de Colombia, 21(2), 61-74, 2012.

[2] E. Arencibia, J. Junco, M. Perera, "Análisis de la contaminación atmosférica de la planta de asfalto caliente "Abel Santamaria", de coliseo", Revista Avanzada Científica, 16(3), 2013.

[3] F. Reyes, M. Madrid, S. Salas, "Mezclas asfálticas modificadas con un elastómero (caucho) y un plastómero (tiras de bolsas de leche con asfalto 80-100)", Infraestructura Vial, Pontifica Universidad Javeriana, 1(17), 25-34, 2007.

[4] N. Kalantar, R. Karim, A. Mahrez, "A review of using waste and virgin polymer in pavement", Construction and Building Materials, 33, 55-62, 2012.

[5] R. Siddique, J. Khatib, I. Kaur, "Used of recycled plastic in concrete: A review", Waste Management, 28(10), 1835-1852, 2008.

[6] D. Martínez, L. Cote, "Diseño y fabricación de ladrillo reutilizando materiales a base de PET", INGE CUC, 10(2), 76-80, 2014.

[7] J. Carswell, F. Cruz, "Misión y ventajas de los betunes modificados con polimeros", Revista de la Asociación Técnica de carreteras, 1(63), 33-39, 1997.

[8] A. Tellez, "La complejidad de la problemática ambiental de los residuos plásticos: una aproximación al análisis narrativo de política pública en Bogotá", MSc Thesis, Universidad Nacional de Colombia, 2012.

[9] G. Gourmelon, "Global plastic production rises, recycling lags", Vital Signs Online, World Watch Institute, Washington, DC, United States, 2015.

[10] R. Sarria, G. J, "La gran problemática ambiental de los residuos plásticos: microplasticos", Journal de Ciencia e Ingeniería, 8(1), 21-27, 2016.

[11] J. Jambeck, R. Geyer, C. Wilcos, T. Siegler, M. Perryman, A. Andrady, K. Lavender, R. Natayan, K. Lavender, "Plastic waste inputs from land into the ocean", Science, 347(6223), 768-771, 2015.

[12] M. Friksen, L. Lebreton, H. Carson, M. Thiel, C. Moore, L. Borrero, J. Reisser, "Plastic polution in the world's oceans: more than 5 trillon plastic pieces weighing over 250,000 tons afloat at sea", PLoS ONE, 9(12), 2014.

[13] Departamento Nacional de Planeación, "Rellenos sanitarios de 321 municipios colapsaran en cinco años, advierte el DNP", Bogotá, Colombia, 2016.

[14] A. Melro, P. Camanho, F. Andrade, S. Pinho, "Micromechanical analysis of polymer composites reinforced by unidirectional fibres: Part I-Constitutive modelling", International Journal of Solids and Structures, 50(11), 1897-1905, 2013.

[15] A. Black, Urban mass transportation planning, McGraw-Hill, New York, United States, 1995.

[16] S. Al-Salem, S. Lettieri, J. Baeyens, "Recycling and recovery routes of plastic solid waste (PSW): a review", Waste Management, 29(10), 2625-2643, 2009.

[17] C. Maharaj, R. Maharaj, J. Maynard, "The effect of Polyethylene Terephthalate particle size and concentration on the properties of asphalt and bitumen as an additive", Progress in Rubber, Plastics and Recycling Technology, 31(1), 1-23, 2015.

[18] C. Swaptik, A. Tashkant, O. Suganya, "Polyethylene Terephthalate (PET) waste as building solution", International Journal of Chemical, Environment and Biological Sciences, 1(2), 308-312, 2013.

[19] H. Kosmatka, B. Kerkhoff, W. Panarese, J. Tanesi, " Design and Control of Concrete Mixtures", Portland Cement Association, Illinois, United States, 2004.

[20] Instituto Nacional de Vías, "Manual de diseño de pavimentos asfalticos para vias con bajos volúmenes de tránsito", Bogotá, Colombia, 2007.

[21] L. Avila, G. Martinez, C. Barrera, F. Ureña, A. Loza, "Effects on mechanical properties of recycled PET in cement-based composites", International Journal of Polymer Science, 1-6, 2013.

[22] Singiresu, Rao, "The finite element method in engineering", Butterworth-Heinemann, Chennai, India, 2017.

[23] E. Wilson, "Three-Dimensional static and dinamic analysis of structures", Computer and Structure Inc, Berkeley, United States, 2011.

[24] S. Ahmad, B. Irons, O. Zienkiewicz, "Analysis of thick and thin shell structures by curved finite elements", International Journal for Numerical Methods in Engineering, 2(3), 419-451, 1970.

Difference in Speech Analysis Results by Coding

Yasuhiro Omiya[*,1], Naoki Hagiwara[2], Takeshi Takano[2], Shuji Shinohara[3], Mitsuteru Nakamura[4], Masakazu Higuchi[4], Shunji Mitsuyoshi[3], Hiroyuki Toda[5], Shinichi Tokuno[4]

[1]*PST Inc., Industry & Trade Center Building 905, 2 Yamashita-cho, Naka-ku, Yokohama, Kanagawa 231-0023 Japan*

[2]*Research and Product Development, PST Inc., 231-0023, Japan*

[3]*Mathematical Engineering of Morality Emotions Graduate School of Engineering, The University of Tokyo, 113-0033, Japan*

[4]*Verbal Analysis of Pathophysiology Graduate School of Medicine, The University of Tokyo, 113-0033, Japan*

[5]*Department of Psychiatry, National Defense Medical College, 359-8513, Japan*

A R T I C L E   I N F O

A B S T R A C T

Mental health disorder is becoming a social problem, and there is a need for technology that can easily check for states of stress and depression as a countermeasure. Conventional methods of diagnostic support and screening include self-administered psychological tests and use of biomarkers. However, there are problems such as burden on subjects, examination costs, dedicated reagents and equipment required for examinations, and reporting bias. On the other hand, voice-based evaluations are advantageous in terms of providing diagnostic support for physicians. They are non-invasive, do not require special and exclusive equipment, and can therefore be easily conducted remotely. We are pursuing the research and development of the Mind Monitoring System (MIMOSYS), which estimates the state of mental health from voice. Recorded audio is often compressed for efficient storage and transmission. However, there are concerns regarding the effects of deterioration of sound quality on analysis by MIMOSYS. Therefore, this study aims to verify the influence of the deterioration of voice quality due to coding on MIMOSYS analysis. As a verification experiment, coding was applied on the recording of 704 subjects reading 17 fixed phrases, assuming compression for transmission and storage. Then, the results of MIMOSYS analysis before and after encoding were compared. A strong correlation was observed before and after encoding, suggesting that MIMOSYS analysis is also valid for voice to which coding was applied.

1. Introduction

This paper is an extended paper from IEEE ICIIBMS 2017 (The International Conference on Intelligent Informatics and Biomedical Sciences) [1].

Mental health disorder is becoming a social problem, and there is a need for technology that can easily check for states of stress and depression as a countermeasure.

Conventional methods of diagnostic support and screening include self-administered psychological tests and use of biomarkers. However, there are problems such as burden on subjects, examination costs, dedicated reagents and equipment required for examinations, and reporting bias.

On the other hand, voice-based evaluations are advantageous in terms of providing diagnostic support for physicians. They are non-invasive, do not require special and exclusive equipment, and can therefore be easily conducted remotely.

We are pursuing the research and development of the Mind Monitoring System (MIMOSYS), which estimates the state of mental health from voice [2]. Recorded audio is often compressed for efficient storage and transmission. However, there are concerns regarding the effects of deterioration of sound quality on analysis by MIMOSYS. Therefore, this study aims to verify the influence

*Yasuhiro Omiya, , +81-45-263-9346 & omiya@medical-pst.com

of the deterioration of voice quality due to coding on MIMOSYS analysis.

2. Materials and Methods

2.1. Vocal Analysis by MIMOSYS

Generally, a healthy person expresses emotions such as joy, anger, sadness, or pleasure depending on changes in their surroundings. On the other hand, when stress accumulates and the person becomes depressed, the expression of emotions declines. MIMOSYS is based on the sensibility technology ST (AGI Inc.) [3] which measures emotion from voice. Short-term mental health index ("vitality") and long-term mental health index ("mental activity") are analyzed from the balance of five combined emotional indices, which are the four emotions "calmness", "anger", "joy", and "sorrow" obtained by voice analysis, and "excitement" represented by the strength of brain emotional activity. ST software detects voice activity based on volume as shown in Figure 1. For each utterance, feature values such as pitch, power, and deviation in power are calculated by a robust fundamental frequency and intonation detection method. Based on these feature values, they are classified into emotional elements of "calmness", "anger", "joy", and "sorrow" through decision tree logic. The degree of "excitement" is also analyzed as the strength of brain emotional activity. Each emotion detected from the input voice is an integer value from 0 to 10, and the strength of "excitement" is output as an integer value from 1 to 10.
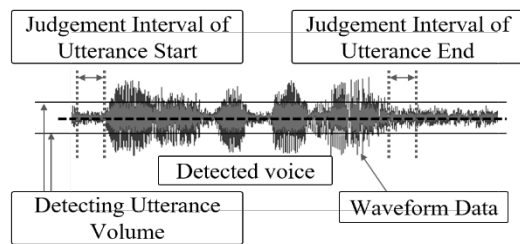

Figure 1. Voice activity detection

MIMOSYS calculates "vivacity" and "relaxation" as intermediate feature values, in the course of its analysis. Vivacity is calculated from "joy" and "sorrow", and relaxation is calculated from "calmness" and "excitement". The short-term mental health index "vitality" is then calculated from these, and the mid-long term mental health index "mental activity" is calculated from the "vitality" accumulated over two weeks. The flow of MIMOSYS processing is shown in Figure 2.
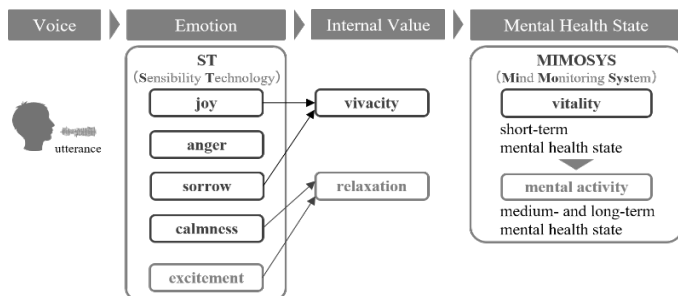

Figure 2. The flow of MIMOSYS processing

2.2. Beck Depression Inventory

The Beck Depression Inventory (BDI) test is widely used for depression screening and consists of 21 questions on how the subject has been feeling in recent days, including on the day. Each question has four options, and the total score determines the severity of depression. Table 1 shows cutoff values according to reference [4]. In this study, since only healthy subjects were targeted, only the voices of subjects with a score of 9 or less were used.

Table 1. cutoff scores for BDI.

Total BDI score	*Level of depression*
0 – 9	indicates minimal depression
10 – 18	indicates mild depression
19 – 29	indicates moderate depression
30 – 63	indicates severe depression

2.3. Voice Collection

We prepared voice (96 kHz, 24 bit, wav file) collection which collected using the Portable Recorder R-26 (Roland, Japan) and a pin microphone ME52W (OLYMPUS, Japan) from male and female subjects (n=986) in the assembled soundproof shelter S-909 (Starlight, Japan). The voices were collected when subjects reading 17 fixed phrases. Additionally, BDI test was conducted before the experiment. Then we excluded abnormal participants using BDI score. There were 704 subjects with screened data used in the analysis. Then we down-converted to 11,025 Hz, 16 bits, which is the input format of MIMOSYS and is used as original voice in this research. Hereafter, audio processing was done using "SoundExchange" [5] and "FFmpeg" [6] software with default coding settings.

This study was approved by the Research Ethics Review Committee. Informed consent was obtained from the subjects.

2.4. Experiment

We prepared data collection of recording format corresponded to linear PCM with a sampling frequency of 11,025 Hz, and a quantized bit number of 16 bits as "Base voice(11k)".

Next, "Base voice(11k)" was converted to a sampling frequency of 8,000 Hz and a quantized bit number of 16 bits simulating transmission across a telephone line. Then, coding methods "G.723.1(MP-MLQ)", "G.723.1(ACELP)", and "G.729", which are based on ITU-T recommendations [7,8], were applied to "Base voice(8k)". Additionally, coding methods "AAC", "MP3", and "WMA", were applied to "Base voice", simulating recording and/or archiving. "Base voice" and each coded voice were decoded and assigned an analyzed "vitality" score through MIMOSYS.

Finally, the "vitality" score before and after coding using each coding method was compared by correlation evaluation. We also compared the analyzed "vitality" score of "Base(11k)" and "Base(8k)" in order to confirm the influence of resampling (down-

sampling and up-sampling). We analyzed the data using Microsoft Excel 365 (Microsoft Corp.).

3.   Results and Discussion

Figure 3-5 is a scatter diagram of "vitality" before and after encoding when "G.723.1(MP-MLQ)", "G.723.1(ACELP)", and "G.729" is used as the voice coding method, assuming transmission over a telephone line. Figure 6-8 is a scatter diagram of "vitality" before and after encoding when "AAC", "MP3", and "WMA" is used, assuming recording and archiving. Figure 9 is a scatter diagram of "vitality" before and after resampling. In addition, Table 2 shows correlation coefficients of "vitality" before and after encoding.

Table 2. correlation coefficients of "vitality" before and after encoding.

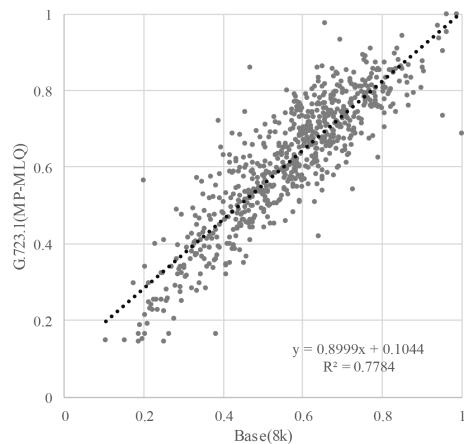| Compared	Coding method	Coefficient of correlation
Base(8k)	"G.723.1(MP-MLQ)"	0.882
Base(8k)	"G.723.1(ACELP)"	0.740
Base(8k)	"G.729"	0.802
Base(11k)	"AAC"	0.959
Base(11k)	"MP3"	0.972
Base(11k)	"WMA"	0.962
Base(11k)	Base(8k)	0.780

Figure 3. Scatter diagram of "vitality" encoding with "G.723.1(MP-MLQ)".
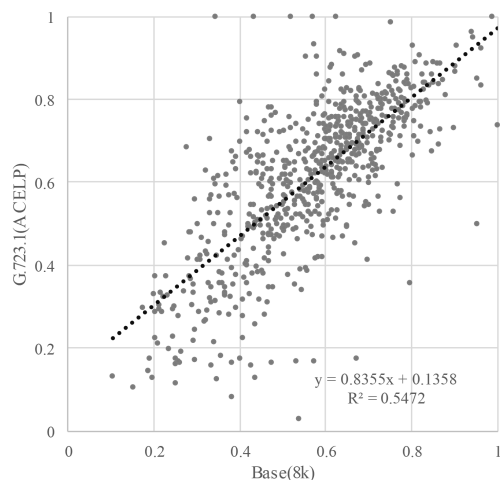

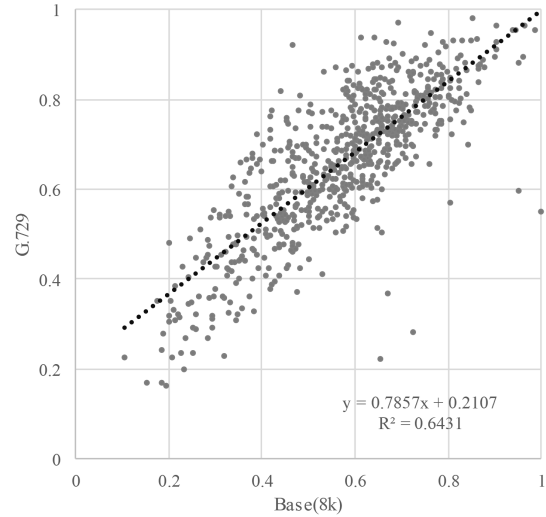Figure 4. Scatter diagram of "vitality" encoding with "G.723.1(ACELP)".

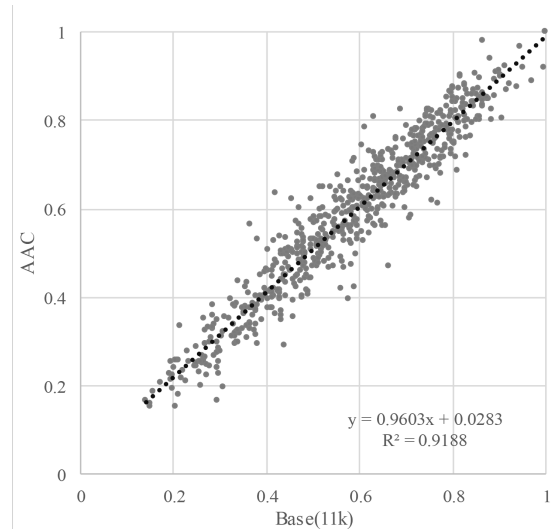Figure 5. Scatter diagram of "vitality" encoding with "G.729".

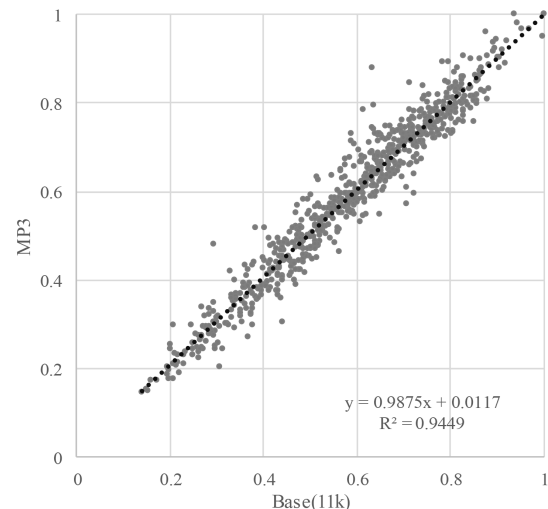Figure 6. Scatter diagram of "vitality" encoding with "AAC".

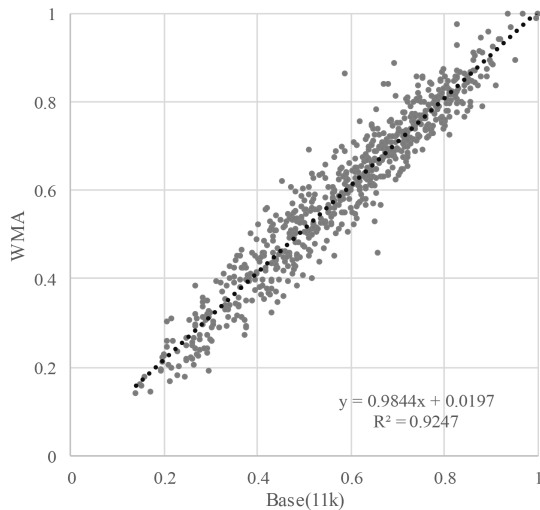Figure 7. Scatter diagram of "vitality" encoding with "MP3".

490

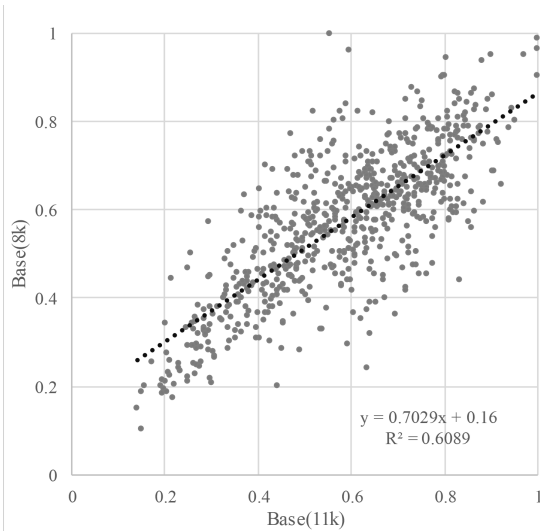Figure 8. Scatter diagram of "vitality" encoding with "WMA".

$$y = 0.9844x + 0.0197$$
$$R^2 = 0.9247$$

Figure 9. Scatter diagram of "vitality" before and after resampling.

$$y = 0.7029x + 0.16$$
$$R^2 = 0.6089$$

Table 3. Average value of difference in "vitality" before and after coding.

Coding method	Ave.$\pm$ S.D.	Base(11k)	Base(8k)
"G.723.1(MP-MLQ)"	0.619$\pm$0.166	0.033	0.047
"G.723.1(ACELP)"	0.613$\pm$0.184	0.028	0.042
"G.729"	0.660$\pm$0.160	0.074	0.088
"AAC"	0.590$\pm$0.181	0.005	0.019
"MP3"	0.589$\pm$0.184	0.004	0.018
"WMA"	0.596$\pm$0.185	0.011	0.024
Base(8k)	0.571$\pm$0.163	0.014	-

The coefficients of correlation before and after coding is extremely high at about 0.74. Coding methods of recording and/or archiving such as "AAC", "MP3", and "WMA" are extremely high at about 0.95. Further, as shown in Table 3, the results suggested that the influence of sampling frequency conversion on "vitality" tends to be higher after coding. This result show that the influence on MIMOSYS analysis of sound degradation, either due to coding for transmission across a telephone line or coding for recording or archiving, is minor and is therefore a useful method. In Figure 4, there are multiple anomaly points where "vitality" after coding by

"G.723.1(ACELP)" is observed around 1 and less than 0.2. This is a peculiar feature of this coding method and use of other coding methods might be preferred. On the other hand, the coefficients of correlation before and after a sampling frequency conversion without coding causes deterioration. This deterioration is inevitable because the input format of MIMOSYS is 11,025 Hz while the input format of the general telephone line is 8,000 Hz.

Prior to this research, we conducted an examination of simulating transmission across a telephone line using coding methods of "u-law", "A-law", "G.726", and "G.723.1" [9]. In this study and previous research, results show other coding methods such as compression methods are useful after coding.

4. Conclusion

In this study, in order to verify the influence of deterioration of voice quality on the estimation of health condition using voice by MIMOSYS, the results of MIMOSYS analysis before and after applying coding to voice were compared and verified. The correlation coefficients of mental activity before and after the application of voice coding were high, and it was suggested that even voice compression coded for transmission via a telephone line or storage could be used. Furthermore, it was also suggested that the influence of sampling frequency conversion on vitality tends to be larger than applying coding.

In the future, we would like to further improve accuracy by considering sampling frequency conversion methods and value correction methods with less influence on MIMOSYS. We also want to improve the accuracy of the MIMOSYS algorithm itself.

Conflict of Interest

The authors have no conflicts of interest to declare.

References

[1] Yasuhiro Omiya, Naoki Hagiwara, Takeshi Takano, Shuji Shinohara, Mitsuteru Nakamura, Masakazu Higuchi, Shunji Mitsuyoshi, and Shinichi Tokuno, "Difference in Speech Analysis Results by Compression," in International Conference on Intelligent Informatics and BioMedical Sciences (ICIIBMS) 2017, Okinawa, Japan, 2017.

[2] Shinichi Tokuno, "Pathophysiological Voice Analysis for Diagnosis and Monitoring of Depression," Understanding Depression, Springer, Singapore, 83-95, 2018.

[3] Shunji Mitsuyoshi, Fuji Ren, Yasuto Tanaka, and Shingo Kuroiwa, "Non-verbal voice emotion analysis system," Int. J. Innov. Comput. I., 2(4), 819–830, 2006.

[4] Beck, AT, Steer RA, Garbin MG J.: Psychometric properties of the Beck Depression Inventory Twenty-five years of evaluation. Clinical Psychology Review 8(1), 77–100, 1988.

[5] Chris Bagwell, SoX – Sound eXchange, Available at: <http://www.soundexchange.com/> (Accessed 12 Jul. 2018)

[6] FFmpeg Developers, "ffmpeg tool (Version 20170503-a75ef15) [software]," Available from: <https://www.ffmpeg.org/> (Accessed 28 Sep. 2018)

[7] ITU-T Recommendation G.723.1 (05/2006), Available at: < https://www.itu.int/rec/T-REC-G.723.1-200605-I/en> (Accessed 28 Sep. 2018)

[8] ITU-T Recommendation "G.729" (06/2012), Available at: <https://www.itu.int/rec/T-REC-"G.729"-201206-I/en> (Accessed 28 Sep. 2018)

[9] Naoki Hagiwara, Yasuhiro Omiya, Shuji Shinohara, Mitsuteru Nakamura, Masakazu Higuchi, Shunji Mitsuyoshi, and Shinichi Tokuno, "Difference in Voice Analysis Result by Pre- and Post- Processing of Telephone Line," IEEE EMBC 2017, Jeju, Korea, 2017.

Towards Adoption of Authentication and Authorization in Identity Management and Single Sign On

Ujjwal Joshi[*], Sangwhan Cha, Saeed Esmaili-Sardari

Computer Science, Harrisburg University of Science and Technology, Pennsylvania, USA

A R T I C L E I N F O

A B S T R A C T

Identity and Access Management (IAM) and Single Sign on (SSO) are two security concepts that are related to each other. IAM governs the user access in an organization whereas SSO facilitates the user by authenticating to one centralized application and not having to re-authenticate when trying to access other applications. This paper addresses the different benefits that an IAM and SSO tool can provide to reduce the security risk within an organization. Since, authentication and authorization are one of the major concerns in the cyber security; this paper analyzes common problems that are faced during authentication and authorization. We have also analyzed prior researches that have been done in the IAM and SSO space along with conducting a survey to understand the different issues and benefits of IAM and SSO. From the survey that has been conducted, we have addressed different issues when implementing IAM and SSO solutions along with understanding the architecture, in which these solutions have been deployed into. The surveys conducted have been compared with prior researches done in IAM and SSO space to understand the benefits that the solution provides. The results from the survey have been analyzed to provide the best practices when implementing IAM and SSO solutions along with the benefits provided by the solution.

1. Introduction

Identity Management also commonly known Identity and Access Management (IAM) refers to the set of processes that can be applied to grant right access to the right people within any corporate enterprise environment [1]. Businesses today, implement IAM solutions so that it provides them with the framework that can manage user IDs and passwords along with solving problems related to the challenges associated with managing accesses and permissions of multiple user IDs [2]. IAM solutions also provide different auditing capabilities that enable the managers and higher officials within organization to keep track of the different accesses that their employees have [2]. A terminated employee still having access to the organization resources can cause certain damage to the organization so having an IAM system can reduce the risk associated while manually removing or adding users to different systems. Today, architectures within an organization can be categorized based on whether their applications are deployed in the cloud or within their own network also known as on-premises [1]. This increases the complexity of managing users and granting accesses to them due

to increased privacy and security risk [3]. Although IAM systems can remove and create access, the user still needs to enter their user name and password to access the systems that they were provisioned to by the identity management systems. Remembering passwords and usernames can lead to issues where the user can forget their credentials or lose their credentials for certain system, which might fall into the hands of an intruder. Single Sign On (SSO) provides a better functionality for managing authentication and authorizations to applications, by not having the user re-enter the credential every time they need to login [4]. SSO provides a single source of authentication mechanism which provides the user a single platform to enter their credentials to access different applications and also discourages the need to maintain multiple credentials to access different systems [5]. IAM and SSO solutions provide a secure way for businesses to manage and authenticate their users. This research explores the authentication and authorization mechanism that IAM and SSO provides with the view of providing better security while implementing these solutions. The research also explores the different models of architectures that an enterprise environment can have to recommend the best practices that can be followed while implementing IAM and SSO solutions.

[*]Ujjwal Joshi, 696 Fox Ave, Lewisville, TX, USA, 469-451-9885
ujoshi@my.harrisburgu.edu

492

Security is one of the major concerns that an organization faces today, as risk from cyber-attacks can damage an organization [6]. As business today are growing bigger and faster than anticipated, they are looking for new technologies that can provide a simplified solution for them to manage their user's accesses and resources. A user within an organization needs to have the right set of access to do their work and manually managing user accesses is not a suitable practice for an organization since there can be gaps such as assigning wrong or incorrect accesses [1]. Moreover, manually adding and removing accesses can lead to issues such as forgetting to remove accesses when users leave organization or adding accesses when users join organization.

Since authentication and authorization are one of the major components of IAM and SSO, the need to provide access and manage user authentication are highly in demand. Corporate environments today are complex because they have users and data scattered in different applications and servers. Therefore, there is always a need to manage users efficiently so that the organization is less vulnerable to intrusions and outside attacks. Most organizations today need the security knowledge that can help them to investigate the security solutions that are available. Business always makes mistakes by believing the vendors or service providers rather than understanding their own infrastructure for their security purposes.

Lack of security knowledge and security principles are one of the common problems that organizations face today. By tradition software applications within an organization were supposed to be deployed within the organization boundaries [1]. However, today an organization cannot restrict itself to the traditional views and are deploying applications outside an organization boundary. Any application that is outside the organization boundary is not within the organization trusts zones and organizations do not have complete control over it. Since some applications are within the organization boundary and some are maintained by service providers, the complexity to sync users between application increases. As technologies are rapidly changing today, IAM is gaining a popular momentum as it is one of the key components when managing users on applications deployed within organization boundaries and outside organizations boundaries [3].The problem still lies on how business and organizations can securely manage users and their credentials for applications that are inside the organization and outside the organization boundaries. Since service provider's applications are managed by service provider's resources and application within organization boundaries are managed by organizational resources, manually adding the users and creating credentials can lead to gaps where the user may not have the right access. The complexity also increases when users try to access service provider's applications with different credentials which can lead to forgetting or losing passwords. Such complex situations can be managed by using SSO for authentication and authorization along with IAM for managing user accesses. However, not all the SSO and IAM may be feasible for an organization. Depending of the complexity of the organization infrastructure there is a need to do research which provides guidelines to business so that they know the check and balances that is required when implementing IAM and SSO solutions.

In this paper, we evaluate different criteria that an organization can have from the infrastructure and architecture perspective. The architecture depends on the number of applications that are hosted within and outside the organizational boundaries. The infrastructure is based on the number of employees that an organization has and the different kinds of policies that is applicable to each employee. Then, we categorize the organization based on architecture and infrastructure that the organization has. Depending on the categories that the organization fits into we analyze the different process that an organization can undergo to select the suitable IAM and SSO solutions. We also provide the guidelines on some of the best practices that the organization can follow to implement those solutions. However, this research does not recommend any solution that exists in the market today but provide more of a guideline that the organization needs to take into consideration while implementing IAM and SSO solutions.

The remaining of this paper is organized as follows. Section 2 presents related work. Section 3 describes the proposed approach. The experiments for the proposed approach are described in Section 4. Section 5 provides the concluding remarks.

2. Related Work

Authentication and authorization are one of the mechanisms used in computer security to validate if the user has the right credentials and permission to access a system. However, authenticating and authorization also known as identification, is one of the common problems in computer security [4]. Although many researches have been done to improve the authentication standards, passwords is still common authentication mechanism [4]. Using passwords as an authentication mechanism increases complexity in an enterprise environment since forgetting passwords from time and again can lead to inefficiency while accessing systems. Organizations today are also demanding systems that help them manage users so that end users can manage their private information more efficiently [7]. Identity management is one of the key components of an organization since it plays an important role in authentication and authorization [3]. Identity management refers to the process of managing identities so that each identity has the right set of accesses that is needed for them to do the required work in an enterprise environment [1]. Identity management consists of four major modules which are Authentication, Authorization, User Management and Central User Repository [1]. The authentication module deals with validating user credentials where the user can authenticate themselves by providing username and password or through other different authentication mechanisms such a Kerberos, SAML and so on [1]. The authorization modules provide the functionality of validating a user access to different systems. The user access can range from the group of permissions or roles that a user has to access the system [1]. The user management module provides functionality such as provisioning and de-provisioning of users, life cycle management to manage the user accesses, when the user joins the company till the time the user leaves the company. The central repository module delivers a one stop place for reviewing users and different access they carry along with different systems that IAM is integrated to and reading information from different other managed systems [1].The proper deployment of the four modules of identity

management can result in building robust enterprise environment where processes can be automated and limited manual intervention is required.

An enterprise environment today, is not just restricted to application deployed within a particular network, but also includes various other applications on different vendor systems. Enterprises now are also moving their resources to cloud and are exploring the different cloud models. There are three cloud concepts which are gaining popularity such as Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS) [8]. The concept of SaaS, PaaS and IaaS has also changed the way different authorization and authentication mechanism worked. The mechanism to store identities across different systems using SaaS, PaaS and IaaS is also known as federated identity management [4]. With cloud-based applications and the need to access applications anywhere at any time has brought in changes in the traditional authentication mechanism and developed different standards that are also known as SSO standards. However, enterprise still use username and passwords as a common authentication mechanism and these can lead to certain problems. In password-based authentication the user passes in a username and password to prove they have the right credentials to access the system. So every time a user needs to access a system they need to enter the password. The use of password-based authentication and classical authorization techniques are described below

- Alpha numeric passwords are used which sometimes can be easy to guess so an unauthorized user may also access the system [4].
- Since different system may have different passwords the user has to memorize all the password or maintain a list somewhere else which can be stolen or lost
- Some system with password-based authentication allows grammars in passwords which provides a way for attacks like dictionary attacks [4]
- Two factor authentications can be a better approach but not all the systems follow two factors authentications.
- Once the user is able to authenticate the next process that happens is authorization which allows the user to access certain resources in the system.
- The authorization techniques for classical system are assigning permissions directly which is difficult to manage [4].
- RBAC (Role Based Access Control) is one of the mechanisms used in various systems to grant access to users for authorization [4].
- The RBAC solves the issues with permissions since permissions are assigned within the roles and roles are assigned to the users [4]. This reduces management risk since they only need to be aware of the roles that the user has.

SSO provides enterprises with a mechanism to authenticate and authorize users with a single password. Once the user logs into one system and authentication happens the users does not need to login again to access other applications [4]. This prevents user from having to remember multiple passwords for different applications. However, since there are many different standards that provide single sign on for users, there is a need to understand and analyze how different single sign on systems works and suits for a particular enterprise environment [4].

Today, there are different SSO and IAM solutions that are available in the market. Organizations need to understand how these solutions work for them based on their architecture and infrastructures design. Some of the commonly used identity management systems that exist today are Sailpoint IdentityIQ, Oracle Identity Manager, RSA lifecycle and governance and so on. SSO tools that exist today are Okta, IdentityNow which have the capability of creating and removing users and also providing single sign on for the enterprise users. Rather than focusing on the tools that exists, it is necessary for the organization to understand the protocols that these tools use and choose the right one that fits their organization. Some of the protocols that are widely use and accepted are OAUTH, OpenID, and SAML etc.

Organizations today have also started a centralized authentication mechanism where the user does not have to remember their entire password and maintain a sticky note or list, attached to their computer [5]. This authentication mechanism when configured properly provides security where users do not have to worry about maintaining sticky note of all their passwords to access different systems [5]. To attain a centralized authentication mechanism different standard can be followed and one such standard is Security Assertion Markup Language (SAML) which uses Extensible Markup Language (XML) based solution for exchanging information between service providers [5]. Another authentication mechanism that can be used is OpenID where an OpenID provider acts as an authentication provider to authenticate relying parties [9]. SSO systems such as SAML, OpenID are the first step in reducing the problem of using different password to log in different systems [4]. Aside from SAML and OpenID another SSO standard that is quite popular is Kerberos authentication. Kerberos provides single sign on through the generation of tickets [10]. These systems are also known as federated authentication standards and are used to authenticate users across different platform [11].SSO protocols like SAML, OAUTH and OpenID replaces traditionalauthentication by using a single source of authentication at the identity provider (IdP) [12]. However, each of these protocols also has their own advantages and disadvantages and based on the literature review there are different ways to attack these single sign on protocols. One of the attacks that is possible which using SAML based single sign on is by constructing a false SAML token [13]. Since SAML uses xml to authenticate the users the attackers can modify the xml to attack the system. One of the major concerns for organization today is how to remediate certain attacks that can occur within the organization. No systems today are free from attacks by attackers despite of the fact whether the organization is using identity management and single sign on solutions. Knowing and having information about different attacks can help organization to plan ahead to prevent those attacks. Some of the vulnerabilities and ways that an attacker can attack an organization involve:

- Injecting malicious endpoints: In this attack the attacker creates a malicious end which is used to force the user to enter their user name and password so that the attacker can retrieve the username and password. Once the credentials are

retrieved the attacker uses the known credential to authenticate themselves [12].

- Redirect Attack: In this attack the attacker learns about the user credentials when the identity provider redirects that user by using wrong redirection code [14]. This is possible since neither OAUTH nor OpenID connect specify how the redirection works [14].

- IdP Mix up Attack: In this attack the attacker confuses the resource provider about the identity provider that the user chooses at the beginning of authorization or authentication process to impersonate as the user to access the data [14].

- State Leak Attack: In this attack the attacker forces the browser to be logged under the attacker name at the resource provider [14]. This attack is also called session swapping since it breaks the session integrity property [14].

One of the challenges that many organizations face today is to understand how single sign on works and the best practices to follow while implementing any single sign on protocols. Since most of the organizations are moving their application to cloud and each application are supported by different vendors the traditional authentication mechanism is not suitable. Also, since not every system is threat proof and analysis need to be done when using different authentication and authorization protocols. The literature and journals available do point out the different standards available for single sign on but do not take initiative on predicting how one relates in terms to other. Also understanding how authorization and authentication works with single sign is one of the major concerns that will be reviewed in this paper. The literature review provides information regarding different SSO standards such as SAML, Kerberos, and OpenID but analyzing how authentication and authorization works with the SSO standards is one of the major aspects that needs to be researched more on.

3. The Proposed Approach

In order to understand the challenges faced by different organizations while implementing IAM and SSO solutions it is necessary to interact, understand and analyze the insights of people working in that platform. Proceeding further, reviewing previous researches done in IAM and SSO provides great insights into the data gathering method and analysis. The research will use qualitative method for data gathering. Qualitative method of data gathering involves individual interviews, observations and research. This research will conduct interviews in phone and through online survey with the IAM and SSO resources that are working with different organizations. The research screening is not only limited to a particular organization within a certain region but will cover different regions and organizations inside United States. Aside from interviews we will also review different researches that have been done prior to this research and analyze the results and outcomes of those researches. The interviews to be conducted will be focused more on organizations that have recently implemented IAM solutions and SSO solutions. The interview methods will also look at organizations that are recently planning to implement some form of SSO and IAM solutions since those interviews can provide more insights into the challenges that the organization had to go through in picking the right solution.

Since interviews are the primary source of data for the proposed research, the interviews will help us to build recommendations and guidelines that can be followed while implementing identity management and single sign on solutions. The interview questions are to be more focused on the steps that were taken before finalizing the IAM and SSO solutions and the challenges faced with those solutions. We tend to use these questions to analyze the different circumstances within the organization prior to building the recommendations. The interviews to be conducted will also have questions that are related to the architecture and infrastructure of the organization and the effectiveness of IAM and SSO solutions to provide enterprise security. The response received from interviews is than compared with the different concepts and researches that have been done prior to this research. Different concepts regarding SSO and IAM are also explored by reviewing journal articles, books and prior researches to build a foundation on the existing solutions. The research will also do a comparative study on the prior researches to fill in the gaps that has not yet been covered. The interviews related to the research have questions related to the impact of SSO and IAM in risk management. The risk management section of the questionnaire will be directed towards understanding how SSO and IAM solutions helped in mitigating risk and vulnerabilities within an enterprise environment.

3.1. Methodology Design

Interviews are the primary source of data for this research topic. The interviews questions are designed to help us understand how IAM and SSO solutions are functioning within an organization. The questions presented are within the understanding of people working in different IAM and SSO solutions or people who have some cyber security knowledge. The research questions are designed to understand the different systems of environment an organization has, different SSO protocols used, examine the IAM and SSO concepts within the cyber security space, number of resources that organization has for IAM and SSO and understand the easiness or efficiency provided by using different IAM and SSO solutions. Visual representation of the data collected and gathered is presented in the experiment results section to analyze and understand the feedback from the survey conducted.

In this paper, we classify the data collected into three different architecture models that an organization can have. The models are based on the different applications and users that the organization has. The models are categorized as small, medium and big and analysis is done projecting the different models and evaluating the differences that each model has in compared to another.

The result obtained from data gathering and reviewing different articles is reviewed to understand the impact and various factors that are involved within an organization. It is expected that the research methodology used above will provide us with the information that is needed to understand the organizations limits and their issues in following the best practices that are required. The result obtained from the research is used to build a set of best practices that is needed for any organization while implementing IAM and SSO solutions in cyber security space. This research

evaluates how IAM and SSO solutions can help to mitigate risk in an enterprise environment when best practices are followed.

3.2. Data Rubrics Classification

The data rubrics classification is used to classify organization based on the users, applications, size of the team that an organization has. The recommended practices to be followed can vary based on each organization architecture and model.

3.2.1. Architecture Design Model

The organization architecture is based on small, medium and big based on the number of users that the organization has

Small
- Number of users less than 10000
- Number of applications less than 50

Medium
- Number of users less than 25000 and more than 10000
- Number of applications less than 150 and more than 50

Big
- Number of users more than 25000
- Number of applications more than 150

Size of IAM and SSO team (Team Classification)

The team classification is based on the number of employees and contractors that are directly and indirectly involved in IAM. Indirectly involvement may include employee and contractors that are aware of the IAM system and need to coordinate with the IAM team for different enhancements that needs to be done.

Small
- Less than 5 employees and Contractors directly involved in IAM and SSO
- Less than 10 employees involved indirectly involved in IAM and SSO

Medium
- More than 5 and less than 10 employees and contractors involved in IAM and SSO
- Less than 20 and more than 10 employees and contractors indirectly involved in IAM and SSO

Big
- More than 10 employees and contractors involved in IAM and SSO
- More than 20 employees indirectly involved in IAM and SSO

Complexities with IAM and SSO solutions

5 The system is reliable
4 The system has issues but has helped organization to function better
3 The system is fairly designed and has space of improvement
2 The system is poorly designed but can be improved
1 The system is not reliable and needs to be replaced

3.3. Population and Sampling

The data is collected based on the survey that is conducted and evaluated based on the data rubrics classification mentioned in the previous section. The data analysis process is as described below

- Understand the current architecture of the organizations surveyed and classify them based on the classification rubrics
- Analyze the problems and issues that can be potentially faced by different organization who have implemented IAM and SSO solutions
- Analyze the security issues reduced with IAM and SSO solutions
- Analyze the recommended practices that can be followed with IAM and SSO solutions through survey and prior researches
- Compare and contrast the feedback from the survey and other past researches that has been done
- Build set of recommended practices based on the survey or prior researches

User Population Sampling

The interview process involves 20 different people who are working in IAM and SSO and the analysis is based on their feedback and literature reviews. The user population consists of different people with different background on IAM and SSO.

4. Experiment Results

The experiment results are based on the survey that was conducted with 20 participants that are working in the IAM and SSO security space. The experiment results presented below has been analyzed and compared with prior researches to reflect IAM and SSO practices that are being followed. It also looks at the complexity involved with IAM and SSO solutions and the role they play in mitigating security risk.
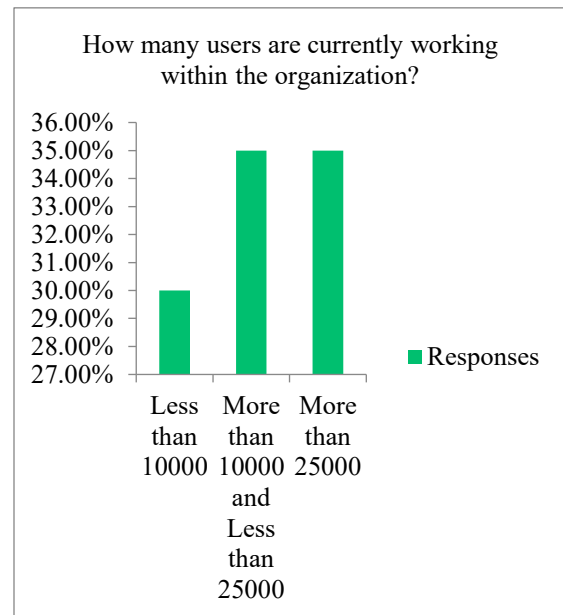
Figure 1: Users currently working within the Organization

4.1. Organization size and Architecture

Based on the survey conducted it can be concluded that most of the organizations that were surveyed had more than 10000 users that was needed to be managed by SSO and IAM solutions. Figure 1 and Figure 2 describes the numbers of users and applications currently managed or needed to be managed by SSO and IAM solutions. 40 percent of the respondents admitted that

the number of applications that has been integrated with IAM and SSO systems is more than 150. Based on this result it can be predicted that the survey data fall under the medium or big tier. From the survey conducted it can also be concluded that most of the organizations engaging with IAM and SSO have a bigger infrastructure and architecture to manage.
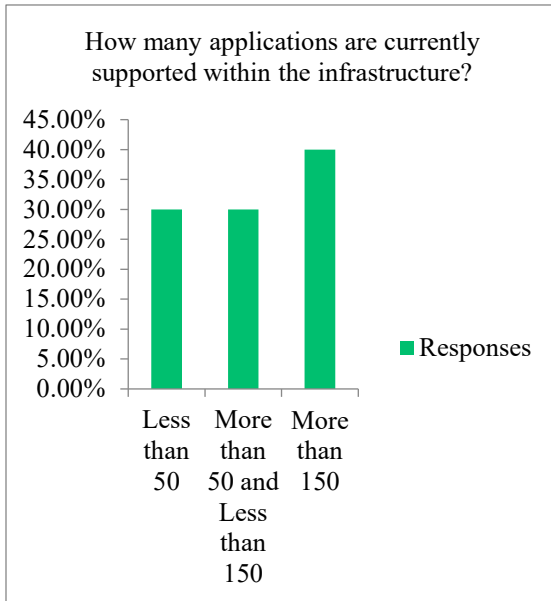

Figure 2: Applications currently supported within the Infrastructure

4.2. Compliance and Auditing Policies

Compliance and Auditing can be described as one of the major reasons that require organization to purchase IAM and SSO solutions. The compliance and auditing policies may differ based on the different domain that the organization is associated with, however the two most popular of those are SOX and HIPAA [15]. Based on the survey gathered majority of the respondents agreed that the organization has some compliance and auditing policies that they need to follow. Figure 3 shows the responses from the participants that were involved in the survey.
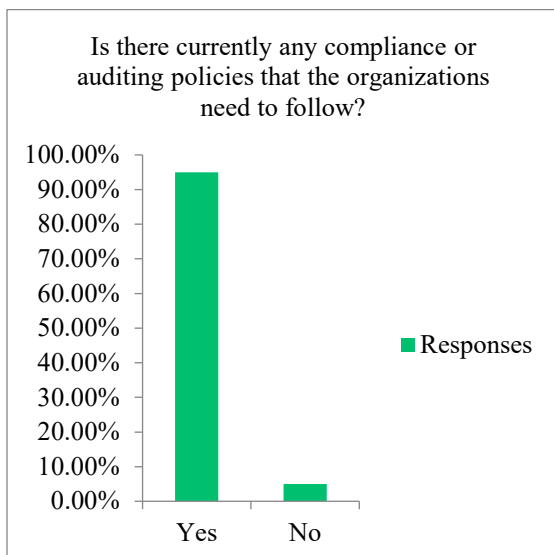

Figure 3: Compliance and Auditing Policies Enforced

4.3. Complexity

Based on Figure 4 most of the respondents responded with already having an IAM and SSO solution in place. Based on this data finding it can be argued that the responses provided to other questions in the survey are more inclined with organizations that have implemented IAM and SSO.
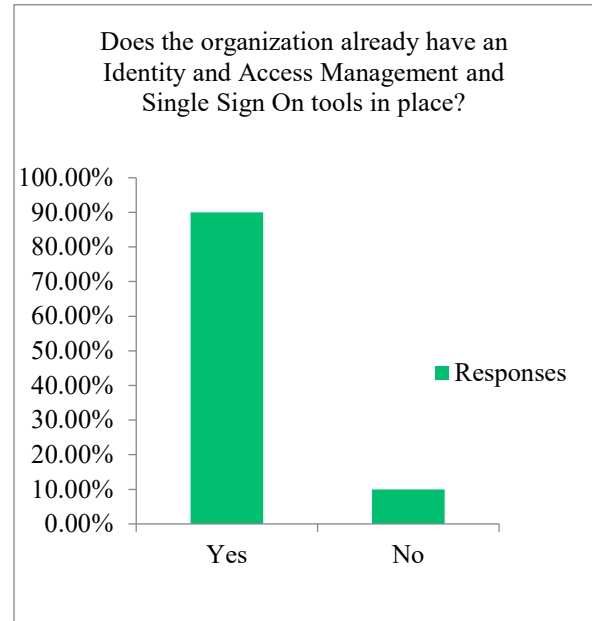
Figure 4: Already having IAM and SSO in place.

The data gathered from the survey also identifies that organizations had reviewed the IAM and SSO solutions before implementing them. The bar graph below shows that most of the respondents responded with the organizations efficiently studying or analyzing the IAM and SSO solutions prior to implementation as shown in Figure 5.

Figure 5: Review of IAM and SSO solutions

Some of the other restriction that was related with the IAM and SSO solutions is shown in the Figure 6.

The Figure 6 defines some of the restriction that is there when using the IAM and SSO tools. Some of the restrictions that are there with the tool are mentioned below from top to bottom according to the data collected are as below:

1. More customization needed with the tool
2. Provisioning capabilities are limited
3. Compliance and Auditing capabilities not as needed by the organization

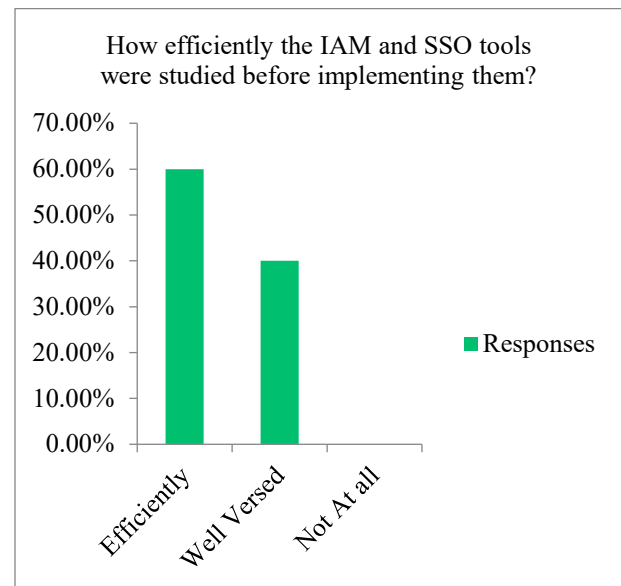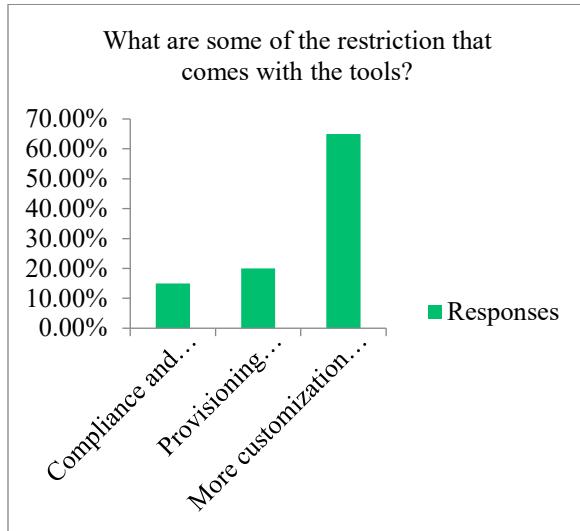**Figure 6: Restrictions with the Tools**

4.4. Benefits provided By SSO and IAM solutions

According to the survey conducted and by reviewing the prior researches that have been done in IAM and SSO, we define the different benefits that are provided by Single Sign On and Identity and access management solutions.

4.4.1. Security

One of the major benefits that a SSO and IAM solution are able to provide is the extra layer of security that it adds to the existing infrastructure. IAM solutions ensures that each user in the organization has the right access to do their work so it decreases the risk of employees having more access than that is needed for them. SSO ensures that user do not have to remember their authentication credentials for every application that they need to login in. Removing the necessity of having to remember multiple credentials also reduces the risk of being easily targeted by losing their credentials to the intruders. SSO protocols make it tougher for attackers to easily extract credentials that are caused by server-side vulnerabilities and trying to trick the system to access a particular application which are managed by SSO solutions [16].

4.4.2. Privacy

An SSO solution also helps to maintain privacy within an organization, since employees do not have to re-enter their credentials for every application. Also SSO authentications techniques such as SAML and OAuth depends upon Identity Provider (Idp) and Service Provider (SP) for authentication mechanism the Idp does not have knowledge about the SP that the user is authenticating to which is also known as private browsing [16].

4.4.3. Automation

IAM and SSO solutions are able to provision users to other application without manual intervention or little manual intervention. This functionality helps organization to automate different provisioning, compliance and auditing activities that are needed which helps organizations to grant and remove employees' access faster and with reliability. An organization not using IAM and SSO solutions will need to rely on manual resources for provisioning, compliance and auditing activities which takes more time and effort.

4.4.4. Regulatory Compliance

Organizations within the United States are required to perform compliance activities based on the domain that they operate within [15]. The Sarbanes-Oxley (SOX) and Health Insurance Portability and Accountability (HIPAA) act are among the most popular act that organizations need to abide by. According to Sarbanes-Oxley act of 2002, the SOX act protects the interest of the investors and general public by diminishing fraudulent practices with the enterprise and improving the accuracy of the disclosures. The SOX act applies to all the organizations in US and the organizations are required to comply with the SOX [15]. The HIPAA act which provides data privacy and security within the health and medical domain. One of the major reasons of using IAM solutions at least in USA is also because of the Sarbox or SOX and HIPAA since organizations need to comply with that act [15].

4.4.5. B2B Collaboration

Most of the organizations today do not work alone but rather collaborate with other organizations as partners. This brings in the need for IT systems in one organization to collaborate with other organizations IT systems [17]. Collaboration with other organization needs resources of one organization to be able to access the systems available in other organizations. For this particular reason an employee within a particular intranet of one organization needs to access applications that are in a different intranet. An SSO and IAM tool makes this possible by centralizing their authentication and authorization mechanism and allowing their users login once and be able to access shared resources across multiple organizations [17].

4.4.6. Simple Administration

Having a centralized system makes it possible for enterprises to have a single source where different access that the user has can be defined [17]. The IAM and SSO solutions provide interactive UI which makes it possible to be a one stop shop to view all the users and their accesses. The ease of being able to navigate through different users and revoke/remove certain access that the user should not be having provides organization with greater flexibility of managing their users.

4.5. Challenges Faced by IAM and SSO solutions

The challenges faced by IAM and SSO solutions mentioned below are derived based on the survey that was conducted and prior researches that describe the issues related to SSO and IAM.

Some of the common challenges are as follows:

- Lack of research done on the tool to study whether it meets the requirements for the IAM and SSO solutions
- Complexity of the architecture and applications that needed to be integrated with the IAM and SSO solutions
- Lack of proper documentation of the product and the tool
- Having incident management and change control in place.
- Understanding the vulnerabilities of the tools and how it can impact the infrastructure
- Sufficient and Knowledgeable resources needed to support the tool.

4.6. Best Practices to Follow

After analyzing the survey some of the best practices that can be followed during IAM and SSO implementation are as follows

- The organization should review the tool that they are implementing and understand the complexities associated with the tool
- The organization should have sufficient and trained resources to manage their tool
- The organization should maintain proper documentation for any changes or customization that is done to their IAM and SSO tool
- The organization should have incident management and change control activities in place for any changes that are done to the IAM and SSO tool
- The organization should understand the vulnerabilities that are with the tool and have teams and resources to mitigate the vulnerabilities with the tool
- There should be 24/7 support for monitoring and managing the tool like having an incident response team available
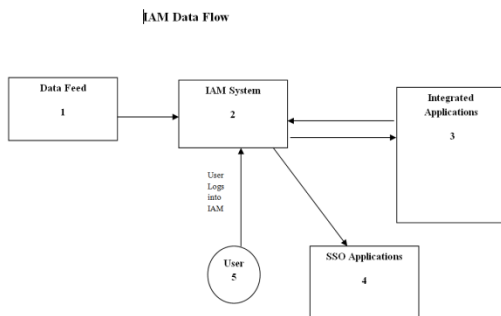
4.7. IAM Data Flow and Architecture

Figure 7: IAM Data Flow

Figure 7 represents the data flow of an IAM system. The bi-directional arrow between 2 and 3 represents the data flow from both the system where one system is the IAM and the other is the applications integrated with the IAM system. The IAM system reads in the user accounts, accesses that each user has in the integrated system but also has the capability to provision new accounts and access associated with users to the integrated system. The data-feed acts as a source of truth for the IAM system from where the IAM system reads in its primary data. The user in the figure represents the user who can log into the IAM system and make specific request for different accesses they need using the IAM system access request framework. If the IAM system

supports SSO capabilities the logged in users into the IAM system can access applications without being prompted for username and password. The applications supporting SSO should be integrated into the IAM system to allow SSO functionality to the users. The SSO mechanism is handled by one or more of the SSO protocols that the IAM system is capable of working with such as SAML, OpenID etc.
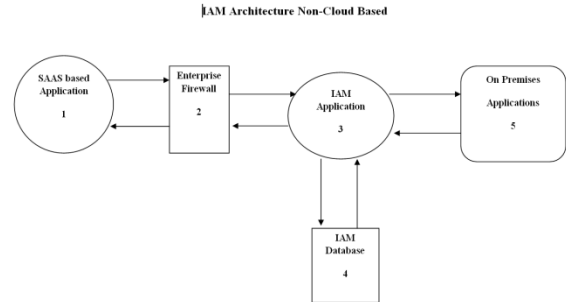

Figure 8: IAM Architecture Non-Cloud Based

Figure 8 represents the non-cloud based architecture of an IAM system. IAM architecture can be categorized as either cloud based or non cloud based. In the non-cloud based architecture the organization is responsible for maintaining the infrastructure needed to maintain the IAM system. The non-cloud based model can be integrated with both the on-premises applications along with cloud based SAAS applications. The architecture consists of an IAM database which is responsible for storing the data that the IAM system reads from integrated applications along with any specific capabilities or data modeling that the IAM system needs to function. On-premises applications are applications that are inside the enterprise network perimeter and the arrows between 3 and 4 represents that the IAM system can read and write data to on-premises application. Since SAAS based applications are not within an enterprise network an enterprise firewall might be present between the SAAS based applications and the IAM system. The firewall should be able to allow request to flow "to and from" SAAS applications to IAM system.
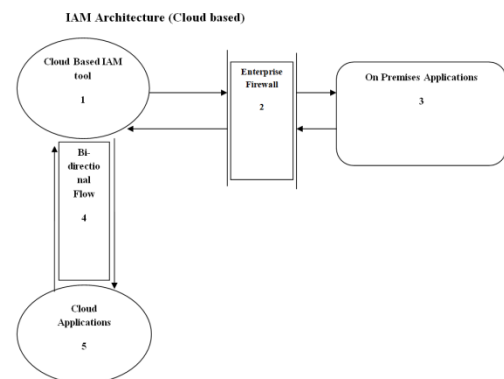

Figure 9: IAM Architecture Cloud Based

Figure 9 represent the cloud based architecture model of an IAM system. In this architecture the IAM system itself is a SAAS application that provides IAM services to the organization. The SAAS based IAM system also provides the functionality to read and write to both on-premises and other cloud based applications.

The arrow between 1,5 and 1,3 represent the flow between IAM system and the integrated applications.

5. Conclusion

In this paper, we analyzed how different IAM and SSO practices are being followed within several organizations. The survey looked at different architectures that the organizations has along with different practices that the organizations followed. In this paper we also analyzed different challenges that the organizations have while implementing IAM and SSO solutions along with the different benefits that an IAM and SSO solution provide. The research also reviewed other research that was done in IAM and SSO to analyze the best practices and comparing it with the survey conducted to understand IAM and SSO practice in organization based on their infrastructure and architecture.

However, we limited our scope to only twenty participants and prior researches to understand the importance of IAM and SSO in the cyber security space. The survey conducted was also limited due to the time and scope of the paper. The participants were only interviewed based on the survey questions and no other possible feedback or inquiry was done aside from the survey questions. Although the research and survey was limited, the responses from the participants helped to analyze the different aspects of IAM and SSO solutions.

Since, our research was only limited to 20 participants, the research scope can be expanded by increasing the number of questions on the survey and the number of participants for the survey. The survey conducted was more focused on understating the importance and benefits of IAM and SSO in the cyber security space, but it did not go in depth to really understand and analyze the IAM practice in different organizations. Since the method of conducting survey was online there was also a limitation on the questions that could be asked and responses that was received. Future work can also look into IAM and SSO practice more in depth from different risk management aspect and cyber security vulnerabilities that can exists. In this research we did not analyze into different cyber security vulnerabilities that can exists with an organization and how IAM and SSO solution can help in reduce each of those vulnerabilities that exists. Future, work can also take into consideration in focusing depth analysis on different SSO protocols that can be used within an organization considering their infrastructure and architecture.

References

[1] Manguic, D. M. (2012). CLOUD IDENTITY AND ACCESS MANAGEMENT–A MODEL PROPOSAL. Accounting and Management Information Systems , 11 (3), pp. 484-500.

[2] Peterson, B. H., Smedegaard, P., Heninger, W. G., & Romney, M. ß. (2008). Managing Multiple Identities. *Journal Of Accountancy* , 1-6.

[3] Nunez, D., & Agudo, I. (2014, 3 6). BlindIdM: A privacy-preserving approach for identity. International Journal Of Information Security

[4] Catuogno, L., & Galdi, C. (2014). Achieving interoperability between federated identity management systems:A case of study. Journal of High Speed Networks 20 , 209-221.

[5] Lewis, D. L., & Lewis, J. E. (2009). Web Single Sign-On Authentication using SAML. IJCSI International Journal of Computer Science Issues , 2.

[6] Andre, T. (2017). Cybersecurity: An Enterprise Risk Issue. Hfm Healthcare Financial Management , pp. 1-6.

[7] Tormo, G. D., Millán, G. L., & Pérez, G. M. (2012, 12 27). Definition of an advanced identity management infrastructure. pp. 173-200.

[8] Alston, A. (n.d.). Attribute-based Encryption for Attribute-based Authentication, Authorization, Storage, and Transmission in Distributed Storage Systems. pp. 1-20.

[9] Hsu, F., Chen, H., & Machiraju, S. (2011). WebCallerID: Leveraging cellular networks for Web authentication. Journal of Computer Security , 862-899.

[10] Pérez-Méndez, A., Pereñíguez-García, F., Marín-López, R., & López-Millán, G. (2012). A cross-layer SSO solution for federating access to kerberized services in the eduroam/DAMe network. International Journal of Information Security , 365-388.

[11] Shitamichi, T., & Sasaki, R. (2014). TECHNOLOGY OF FEDERATED IDENTITY AND SECURE LOGGINGS IN CLOUD COMPUTING. International Journal of Electronic Commerce Studies , 5, 39-62.

[12] Mainka, Christian; Mladenov, Vladislav; Schwenk, Jörg. (n.d.). On the security of modern Single Sign-On Protocols –Second-Order Vulnerabilities in OpenID Connect.

[13] Krawczyk, P. Secure SAML validation to prevent XML signature wrapping attacks.

[14] Fett, D., Küsters, R., & Schmitz, G. (2016). A Comprehensive Formal Security Analysis of OAuth 2.0.

[15] Heino, E. (2011). Evaluating financial benefits of an identity management solution CASE Logica. Retrieved July 9, 2018, from http://epub.lib.aalto.fi/en/ethesis/pdf/12500/hse_ethesis_12500.pdf

[16] Alaca, F., & Oorschot., P. C. (2018). Comparative Analysis and Framework Evaluating Web Single Sign-On Systems.

[17] Bazaz, T., & Khalique, A. (2016). A Review on Single Sign on Enabling Technologies and Protocols. International Journal of Computer Applications, 15 (11), 18-25.

Emotional Impact of Suicide on Active Witnesses: Predicting with Machine Learning

Richard Osei Agjei[1], Emmanuel Awuni Kolog[2,3,*], Daniel Dei[4], Juliet Yayra Tengey[5]

[1]*Texila America University, Department of Public Health, India*

[2]*University of Ghana, Department of Operations and Management Information Systems, LG 78, Ghana*

[3]*University of Eastern Finland, Department of Computer Science, School of Computing, FI-8001, Finland*

[4]*Valley View University, School of Graduate Studies, Ghana*

[5]*University of South Africa, Department of Psychology, South Africa*

A R T I C L E   I N F O

A B S T R A C T

Predicting the impact of suicide on incidental witnesses at an early stage helps to avert the possible side effect. When suicide is committed in public, incidental observers are left to grapple with it. In many cases, these incidental witnesses tend to experience the emotional side effect with time. In this study, we employed a Machine learning algorithms to predict the impact of suicide and suicidal attempt on incidental witnesses. This prediction was based on the accounts of suicide given by selected participants who have witnessed the act. The accounts, which was pre-processed into a corpus, were manually annotated with predefined emotion categories. While sadness emerged as the most salient emotional impact on the witnesses, fear was found as the lowest of the emotional impact on the witnesses. However, the machine learning prediction yielded highest in predicting depression with insignificant variations in the other emotional categories. This nonetheless shows that people who have witnessed suicide or suicidal attempts are inherently affected by some form of emotions that may require urgent attention to alleviate. By evaluating the performance of the Machine learning algorithms, the Support Vector Machine was superior, in terms its prediction, then the Multinomial Naïve Bayes algorithm. The outcome of the study contributes to the pool of research that sought to advocate the use of Machine Learning for predicting social phenomenon.

1. Introduction

Suicide is the termination of an individual's own life purposefully due to several factors ranging from work pressure, family pressure, peer pressure, financial pressure, academic pressure, relationship or family disharmony, social pressure, alcoholism among other known and unknown circumstances. These behaviours have become a global phenomenon where a considerable research has proven its association with cognitive and affect disorder. In recent times, the problem of suicide has become a common occurrence in both developed and under-developed countries leading to psychosocial effect and evocation of negative emotions such anger, sadness, depression, fear, worry and anxiety. Among the

global violent deaths, suicide accounts for 50% for men and 71% for women [1]. The impact of suicide on second and third parties in the society has become a social issue that requires utmost attention. Globally, the female and male ratio for suicide is 1: 4. This ratio amounts to the seventh leading cause of death among men [2]. However, in developed countries, suicide ratio of women to men is 1:3 while low and middle-income countries have the ratio of 1: 5.

Suicidal behaviour is any action that could cause a person to die. Before committing suicide, people exhibit suicidal behaviour that could be averted if the behaviour is detected early enough. Individual's with suicidal behaviour possess some kind of physical, psychological and economical threat to their families, friends and the society at large due to the rippling effect that

*Dr. Emmanuel Awuni Kolog, Box LG 78, +233248271443 & eakolog@ug.edu.gh

suicide or suicidal behaviour pose. Generally, research has revealed that suicide rate is predominant in the age of 70 years and above while the second leading cause of death is among those in the age range of 15 to 29 years [1]. People commit suicide in several ways. Notable ways are Jumping from a height, drowning, going under a train, hanging, use of firearms, poisoning, self-cutting, self-immolation, and vehicular suicide, [3,4,5,6,7,8,9]. Adequate knowledge of these behaviours can inform several interventional methods and techniques.

When suicide is committed, the emotional impact of the act on witnesses is often overlooked. When the act is committed in public, incidental observers are left to grapple with it. In many cases, these incidental witnesses tend to experience the emotional side effect, which in some cases result in more severe damage. Predicting the impact of suicide of witnesses at an early stage helps to avert the possible side effect. With this in mind, we aimed at, from the point of view of using computational Machine learning techniques, tracking the emotional impact of suicide or suicidal attempt on witnesses from their accounts on witnessing suicides or attempts on suicide. The accounts were analysed with Machine learning classification techniques.

Many researchers have used machine learning applications for the detection of emotions and sentiments in text from diverse domains. In the health sector for instance, where clinical situations or assessment appear ambiguous, machine learning can be employed to validate, ascertain and strengthen clinical prediction [10, 11]. Machine learning can be applied to clinical data and/or data on the measurable indicator of the severity of suicidal behaviour. Thus, machine learning used in conjunction with clinical data or biomarkers will help reduce the error margin in suicide risk and suicide predictions. In the arena of psychology, the use of machine learning is used predominantly to predict suicide risk among patients with mood and depression disorders [10].

2. Theoretical Framework

In this section, these researchers discuss related literature and theoretical foundation of emotions and their impact on suicide. Additionally, the role of using machine learning to predict suicide is discussed in this section.

2.1. Emotion and Suicide behaviour

Humans are a highly social species [12], and that imply that emotion prioritize response to threats and opportunities in an environment in which individuals seek to survive. When friends and family relations witness suicidal behaviours in the environment, where they live, it poses psychosocial effect and negative urgency on their emotional behaviour.

Emotion is considered to be inherently social [13] which implies that the environment and the interacting partners are key players in the evocation of positive and negative emotions. Emotions can trigger heuristic thought or swift action. It is hypothesized that emotions will bring a direct effect on quality of relationship [14], but will also moderate the relationship between its components. Thus emotions exert causal effects on the quality of our relationships [15]. Emotions represent key underpinnings for interpreting not only human decision making but also much of human behaviour as a whole. Since affective researchers have given much attention to the impact of the environment on

individual's emotions [16, 17, 18], it is important to consider the impact of suicide on people who incidentally witness suicide. It is therefore crucial to understand how human emotions are affected by constant interaction with their immediate families, friends and the society as a whole [19]. The social functions and impact of society on individual's emotion have seen a considerable research, yet the general agreement among affective researchers as to whether they represent emotion is a matter of constant debate [19].

Decision serves as a conduit for both positive and negative emotion while decision making is largely influenced by emotions [21, 22, 23]. Emotion can influence suicidal thought and the depth of information which can trigger physiological, behavioural, experimental and communicational responses that enable an individual to take action quickly based on the encountered problem [23]. There is a connection between general mood and the emotional situation and happenings in an individual's environment [24]. When individual witness the incident of suicide, the emotions that are evoked are incidental emotions, which often produce influences that are unwanted and unconscious.

Suicide is a complex neurobiological phenomenon, and there is great variability in the validity of clinical assessment tools for predicting suicidal risk. Despite reaction of uncertainty surrounding witnesses' accounts of events, recent scholarship on violence has highlighted the credibility of testimonies by witnesses. Exposure to violent behaviour has the same traumatic effects on witnesses as it does on victims [25]. Like violence of all forms, witnessing suicidal behaviours affect witnesses. Extant studies have revealed that there is a high chance that witness of suicidal behaviours or suicidal attempts can develop maladaptive lifestyles [26]. In some situations, witnesses show symptoms of withdrawal in addition to maladaptibility [27]. When individuals witness suicidal behaviour the rippling effect is the retardation of the normal mental functioning of witnesses [28]. Several research techniques have been applied in determining the effect of suicide on incidental witnesses, notable is from the point of view of qualitative approach. To add is that, less studies have considered the use of predictive techniques, such as machine learning, in this domain of research. This partly informed the reason for this work, in order to ascertain if the use of machine learning can predict the emotional effect of suicide on incidental witnesses of suicide.

Considerable studies have reported that three out of every five suicide deaths yielded no clue of the suicidal communication, threats, ideation/suicide attempts or any life time history of suicidality [29, 30]. More so, there are several illnesses that precipitate suicidal ideation among patients, and these illnesses differ from patients to patients which makes it more difficult to identify a specific biomarker(s) to predict suicide risk [31]. The ability for health professionals and providers to predict an individual's suicide risk with high certainty is limited by lack of clinical prediction rules [31]. The concept of suicidal intent/behaviour in connection with emotional dysregulation is ambiguous and contradictory while examining several studies [32, 33, 34].

Several theoretical models have been proposed in literature to provide strong theoretical and factual information that concerns the progression of puzzling behavioural and emotional outcomes. Notable of these theories of suicidal behaviour is the Interpersonal

Theory of Suicide [35, 36, 37]. The Interpersonal theory of suicide was earlier suggested by Joiner (2005), and later elaborated by Van-Orden *et al.* [36]. According to this theory, "the most dangerous form of suicidal desire is caused by the simultaneous presence of two interpersonal constructs—thwarted belongingness and perceived burdensomeness" [36: 1]. The theory further elaborates that the capability to engage in suicidal behaviour is separated from the desire to engage in suicidal behaviour. The Interpersonal Theory of suicide presents a workable construction to further improve the quality of our understanding of suicidal behaviours. This theory has proven to be one of the best lenses through which suicidal behaviour can be understood [38]. It explains the suicidal intent or ideation-death gap and provide more detail understanding of suicidal behaviour in general. The concept reveals that the interplay between perceived burdensomeness and thwarted belongingness brings about the need for an individual to die by suicide. The theory suggests that the riskiest form of suicidal desire is introduced and provoked by the interplay between perceived burdensomeness and thwarted belongingness. In addition, Christensen *et al.* [39] found that the interplay between perceived burdensomeness and the thwarted belongingness predicts suicidal ideation. It is interesting to note that in their findings interpersonal theory of suicide explained more variance than the epidemiological models using demographic factors, including depression, anxiety, traumatic and stressful events as predictors [39].

2.2. Effect of Negative Emotions on Suicide

Lerner *et al.* [40] view emotion as a "coordinated reactions to survival-relevant event, including cognitive and biological changes, facial and/or bodily expressions, subjective feelings, and action tendencies such as approach or withdrawal". Suicide has a close association with the influence of negative emotions. While emotion is considered as one of the factors to influence suicide, Loyo *et al.* [60] have reported failures in interference control, cognitive rigidity, and alterations in verbal fluidity and decision-making as other influences of suicide or suicidal attempts.

The emotional aspect of the Interpersonal Theory of suicide is important, in that, the three component of the theory are dangerous for any one. This danger is "significantly amplified in individuals with high level of negative urgency, as they may be more reactive to less severe environmental stressors and more highly motivated to respond to negative behaviours capable of impacting suicidal risk" [41]. Depression, anxiety and trauma comes with negative urgency. This tendency to act rashly in an attempt to reduce feelings of negative Affect: unspecified feelings; the superordinate umbrella of constructs involving emotion, mood, and emotion-related traits [42]. Individuals with high levels of negative urgency might experience quicker, more sever increase in suicidal desire as a result of the devastating nature of their negative emotions [41]. Anestis *et al.* [41] argued that negative urgency does not only impact the desire for suicide, but also intensify the development of the acquired capability.

Problems associated with emotions are classified as a significant risk factor for suicide behaviour. Lees and Stimpson [43] explain that the concept of emotion, in reference to suicide, is often neglected due to the misunderstanding of emotional behaviours. Theories that bother on the state of the mind,

emotional states and actions of suicidal person assimilated affect by theorising suicidality as self-harm or a form of hostility to self. Researchers have found that people participate in deadly self-harm (suicide) as a way of avoiding undesirable emotions and prevent unwanted pain. For instance, extant study suggests that being conscious of one's shortfalls leads to higher destructive emotion and for that matter, individuals cultivate a longing to end this consciousness [44]. With this effect, individuals try to lessen this consciousness and control emotion thus resulting in irrationality and worthlessness [44]. This state of reduced consciousness and rejection of emotion makes extreme actions, such as suicide, a suitable way to escape from one's self and the world.

Additionally, flashbacks, trauma, anxiety and fear of those who witnessed, experienced or experiencing suicidality is connected with problems in controlling emotions and heightened rates of suicide. Though Thomas [45] maintains that individuals with unstable emotions have inadequate knowledge of their personality or health status, thereby engage in suicidal behaviour as they recognised as the only means of coping. Research investigating emotion susceptibility to suicidal behaviour have shown that there is no single pathway or pattern to follow or track individuals who are the first time suicide attempters. This however suggests that the non-rejection of emotional reactions and poor access to professional emotional treatments could be the major means to foretell modern behaviours that are connected with suicide [46].

Anestis *et al.* [41] maintain that research examining the role of emotion regulation in Interpersonal Psychology Theory of Suicide has been inadequate. This is because the main focus of suicidal behaviour is lack of emotion regulation rather than behaviour. The researchers further reiterated that since emotions affect a number of psychiatric disorders with high suicide rates, it would be important to examine its relationship to suicide behaviour. Fredrickson [47] offers some theoretical and empirical support by indicating that individuals undergoing positive affect show flexible, imaginative, resourceful thought and positive emotions get rid of persistent negative emotions.

3. Research Methodology

This section elaborates on the methodological process of achieving the research objectives. The context of this study and the data collection procedure are as well outlined in this section. Pre-processing of the subjective responses of the participants is coded according to a pre-defined emotion category, which are presented in this section. Given the recent interest in advancing inter-disciplinary research in ICT, these researchers have further analysed the collected data from the point of view of using machine learning classification techniques to predict the emotional impact of suicide on witnesses based on their accounts of suicides or suicide attempts.

3.1. Data collection

Data for this study was collected in Ghana. Nine hundred and seventy-eight (N = 978) instances of emotional antecedents, regarding witness of suicides or suicidal attempts, were collected from randomly selected participants, drawing mainly from the health and other institutions of higher education. The basic rationale for the diverse sources of the data was to help understand the impact that suicide or suicide attempts have on people. The

collection of the data was mainly through digitally-developed questionnaires. Based on the objectives of this study, the selected participants, for this study, were tasked to respond to the questions subjectively, asking for their experience on suicide or suicidal attempts by themselves or others. The questions were purely an open-ended one, except for their demographic information. The open-ended responses from the participants were developed into a corpus that required further pre-processing for easy coding or annotation for the prediction.

3.2. Ethical Considerations

Given that human emotions are a sensitive phenomenon, we took into account the ethical considerations of the selected institutions and participants who responded to the questionnaires. An official letter requesting for the right to collect data was forwarded to the ministry of education and the selected institutions. The major content of the letter was to seek ethical clearance before the data was collected. This process was in compliance with the Ministry's policy of data collection.

What is more, an informed consent was obtained from each of the participants. These researchers took turn to educate the participants on the importance of this study and why their data is relevant for this study. The participants were given the opportunity to opt out at any point should they find any ethical breach or discrepancies in the process. Consequent to this, some of the participants opted out without recourse to their action. Nonetheless, the participants were assured that their data or responses will be treated confidentially and that only group findings will be reported, which will not have a direct connection to individual participants.

3.3. Pre-processing

These authors developed basic emotional themes (categories) that align Plutchik's eight basic emotions [59]. The development of the emotion categories was based on the manual analysis of the sample text instances. In line with Plutchik's eight basic emotions, however, we sampled 10% of the corpus and analysed to determine the common emotional themes. In the end, we deduced seven basic emotions which were found as the various emotions that the participants expressed in the sampled data. These emotion categories are *worry, anger, fear, sadness, depression, surprise* and *no emotion.* The corpus was pre-processed for the selected coders to code each instances of the corpus with the predefined emotion categories.

Three human coders with background in human psychology were selected and trained to label the instances of the corpus with the predefined emotion categories. The reason for the selection of the three coders is backed by Kolog *et al.* [49]. Kolog *et al.* [49] revealed that the emotional state of text coders influences their perception when analysing emotions in text, especially when the coders are given limited period to carry out such task. What is more, the coding process is time consuming and the outcome may not be satisfactory which could result in bad decisions on the side of the coders.

Furthermore, these researchers are of the view that people are more susceptible to the influence of their environment which has the potential to influence their perception of emotions. For example, one's domestic challenge, at a point, could influence one's decisions while analysing other people's emotional behaviour. This could, nonetheless, be attributed to the subjective nature of emotions when especially expressed in text. For this reason, the selected coders were given ample and free time to report their findings. Although, some of the instances of the data could yield or contain more than one emotion category, the coders were asked to choose the emotion category based on the most dominant emotional keywords. Below are two instances of the coded text instances of sample corpus:

> *"After hearing that I felt very sad and wondered why one will take his or her life no matter what one is going through. I feel the devil has gotten it way too much into our society because I know there is a spiritual backing to suicide which is taking much control over many especially the youth"* **[Surprise].**

> *"Hmmm! it very sad to hear that a fellow age mate has taken their own live. Sometimes I ask what at all they could be going through that they can share with anyone but only thought of suicide. I got depressed for couple of days before coming back to normal. Hmm only God knows Best"* **[Depression].**

3.4. Machine Prediction

Having labelled the text instances by the selected coders, we further employed a machine learning technique to classify the corpus according to the pre-defined emotion categories. With this, we used WEKA machine learning software [50] for the prediction (classification) of the emotional impact of suicide on the witnesses. WEKA contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization [50]. It is well-suited for developing new machine learning schemes. WEKA was developed by Machine Learning Group at the University of Waikato in the New Zealand [50]. WEKA is a collection of machine learning algorithms for data mining tasks. WEKA machine learning tool was adopted because of its versatility to analyse unstructured text. The tool is a widely known data mining tool that has generally been accepted in this research domain. Additionally, WEKA makes available source code for some of the packages, such as the classifiers, for importation into personal project. This makes it flexible to makes changes to achieve the desired outcome.

Two of the Supervised machine learning classifiers in WEKA were used in this study - Support Vector Machine (SVM) and Multinomial Naïve-Bayes (MNB) classifiers. The choice to use these classifiers is rooted in the fact that considerable studies have found these classifiers efficient and robust for text classification [51, 52]. In WEKA package, we employed 10-folds cross validation for training and predicting of the emotions in the text. This was carried out after the corpus had been tokenised and the features extracted for the classifiers. With this in mind, the feature words were extracted as bag-of-words for the classifiers to learn and create a model for the prediction. The 10-folds cross validation means the coded corpus was randomly partitioned into 10 equal sub-sample sizes. This suggests that the cross-validation process is repeated 10 times (the folds), with each of the 10 subsamples used exactly once as the validation data. In the subsequent paragraphs, these authors explained the various classifiers used in this study. Figure 1 illustrates the various stages in the prediction of the emotions from the witness-based accounts using supervised machine learning.

To ascertain the efficacy of the machine prediction, we evaluated the performance of the classifiers against the human

coded corpus. With this in mind, we computed for the recall, precision, F-measure and the accuracy of the various classifiers. While there are other evaluation measures, precision, recall and accuracy have widely been used in the evaluation of text classifiers. Recall is the proportion of the labelled instances of the gold standard that are identified and extracted by a classifier while the fraction of the automatically extracted data that was labelled correctly as the gold standard by the classifier is Precision. The F-measure, as referred to as the F-score, is the harmonic mean of the recall and precision. The Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations [53].
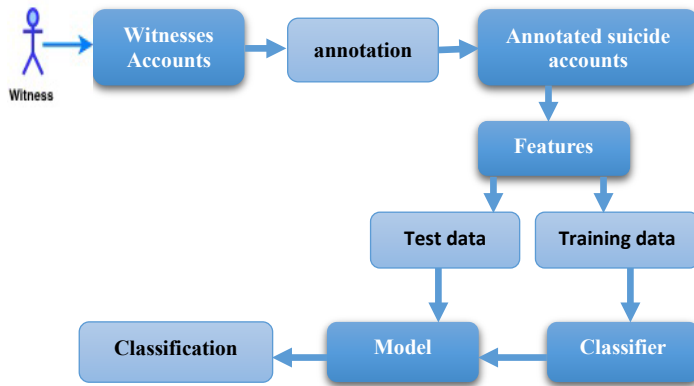

Figure 1. Conceptual classification process

3.4.1. Support Vector Machine

SVM is a supervised machine learning algorithm which can be used for both regression and classification. SVM is a discriminative algorithm in a sense that it is defined by constructing a hyperplane or a set of hyperplanes in high dimensional vector space. Before SVM performs classification task, feature words from the text instances is expected to be extracted through several proposed approaches such as bag-of-words, part of speech tagging among other new advances. After constructing the hyperplane according to how the data is labelled, a model is created for predicting unseen data. Intuitively, "a good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general, the larger the margin, the lower the generalisation error of the classifier" [51]. In this work, we used Sequential Minimal Optimisation for SVM in WEKA [50]. Research has indicated that SVM performs more accurately if the training data is very large [54]. Figure 2 shows a typical separation of data points into two categories, thus binary classification. This can, for example, be the detection and classification of input data into positive and negative polarity.

Detecting and separating the data points in a high dimensional vector space is transform by a *kernel*. The kernel in SVM is used to recognise patterns in the data points. So far four different kernels are available in SVM: *linear kernel, polynomial kernel, Gaussian basis radial basic function* (RBF) kernel and *sigmoid kernel*. However, in this study, all the various kernels were used to determine which of the them would yield the best performance. In the end, the linear kernel performed better, of which the results are reported in this work.
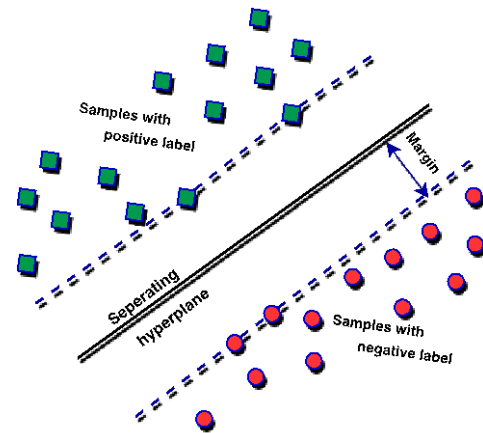

Figure 2. Maximum margin and separation of the hyperplane in SVM (Adapted from Kolog, 2017)

3.4.2. Multinomial Naïve-Bayes

Unlike the SVM, NB is a text classifier that is based on the Bayesian theorem with strong and independence assumption in the model rather than a particular distribution of each feature. A Naive Bayes model assumes that each of the features it uses are conditionally independent of one another given some class. Naïve-Bayes is actually suited when the input data is in high dimensional vector space. What is however different is that NB can classify text instance more accurately even if the training data is small.

4. Results and Discussion

This section reports the findings arising from the analysis and classification of emotions in the witnesses' accounts. This considers the demographic information of the participants, and the emotional impact of the participants from the point of view of their accounts on suicidal antecedents. The results are accompanied with the discussion and backed by literature.

4.1. Participants demography

Of the 978 instances of the text corpus, 49% (n = 483) of them were reported by male participants while the remaining 51% (n = 495) were reported by the female participants. As research has proven that *age* is an important moderating factor in determining suicidal intent or behaviour [55], we considered the age brackets of the participants in this study. With this in mind, 4% (n = 38) of the participants were in the age range of 4 to 24, 49% of the participants were in the range of 25 to 44, 47 % of them were in the range of 45 to 64 and none of the participants were above the age of 65years. Figure 3 shows the gender and age representation of the participants in this study.

Having established the gender and age of the participants, we further prod the participants regarding the demographic information of the supposed people they have witnessed committing or attempting suicide. Although this was not the central focus of this research, the demographic information is equally vital in ascertaining the calibre of people who have committed or attempted suicide. From the responses, we deduced that 43% (n = 419) were female victims of suicide as reported by the participants while the remaining 57% (n = 559) were male victims. This is consistent with extant findings that suicide is

prevalent among men than women. For example, the American foundation prevention for suicide have reported in 2018, that men die by suicide 3.53 times more than women [56]. Additionally, in 2017, WHO reported 1, 129 of suicide cases in men as against 729 suicide cases in female [57]. In Ghana, WHO have reported that 1,500 Ghanaians commit suicide every year where men record the highest number of suicides [57]. However, female has the highest number of attempted suicides and broken hearts in the various psychiatric hospitals in Ghana [57]. Conversely, the participants reported higher number of victims of suicide or suicidal attempts in men than women as in all the age groups.
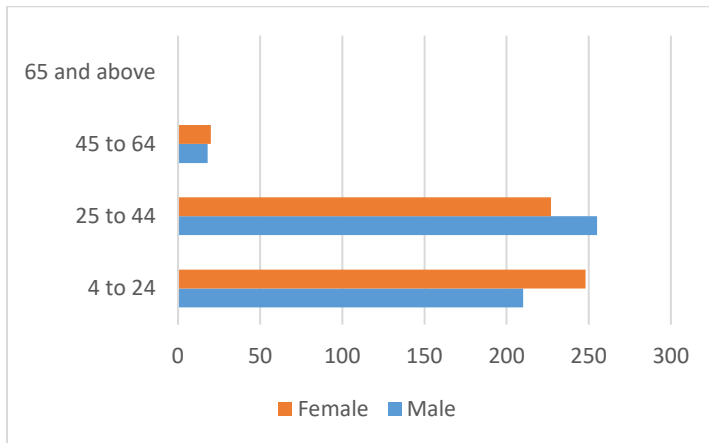

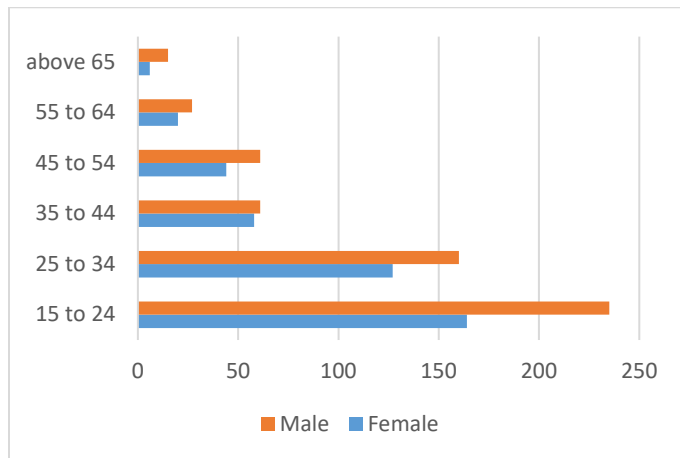Figure 3. Gender and age representations of the participants in this study

Figure 4. Age and gender representation of suicide or attempted victims by reported by the participants

Further, we collected data on the age range of the victims of suicide as reported by the participants. With this data, Figure 4 represents the age ranges of the suicide victims reported by the participants in this study. In the figure, it can be deduced that suicide is prevalent in the age of 15 to 24 years. In line with these findings, WHO in 2015 reported suicide to be prevalent among people in the age group of 15 to 29 [57]. Osafo *et al.* [58] found, in Ghana, that suicide behaviour is motivated by social taunting, hopelessness, and partner's infidelity. The American foundation for suicide prevention has, in real time, rated suicide as the 10th recorded deaths in the United State of America, while in Ghana suicide is the second highest deaths [57]. The difference in suicide deaths with respect to these countries could be attributed to the diverse

characteristics of each context where culture, social, economic and political colouring accounts for the variations.

4.2. Emotional impact of suicide on witnesses

In this section, we present the findings arising from the inter-coders' agreements of emotions in the corpus. In addition, these researchers report the findings from the machine learning prediction. The inter-coder agreement *kappa* score was computed from all the three coders. As indicated in Table 1, a kappa score of 0.75 **(75%)** was obtained, which therefore indicates the variations in the coding patterns among the coders. The results from the coding is illustrated in Figure 5 and this is based on the predefined emotion categories *where 1- worry, 2- anger, 3 – fear, 4 – sadness, 5 – depression, 6 – surprise and 7- no emotion categories.* After obtaining the *kappa* score, the disagreements of the various instances of the data were subjected to further review and scrutiny. The deliberations among these researchers and the coders were meticulously carried out. In the end, we managed to agree on the instances of the data that were disagreed among the coders during the initial coding. The modified inter-coders' agreement of the emotions is shown in Table 1 while the initial coding by the coders is illustrated in Figure 5. The modified coded data was used as a *gold standard* for comparing with the machine learning prediction, as earlier established.
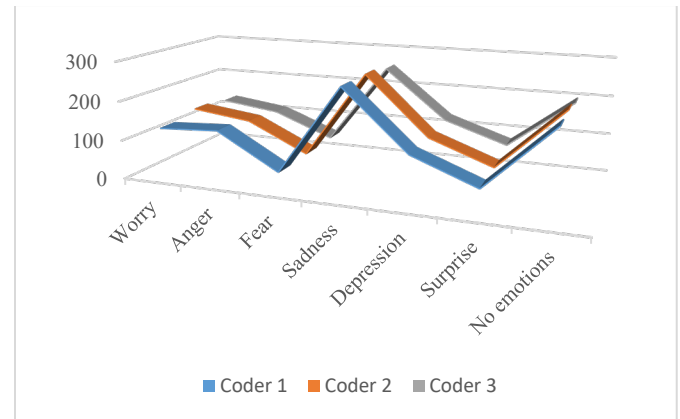

Figure 5. Coding patterns of the participants according to the emotion categories

Table 1: Inter-coded instance of emotions in the corpus

#	Emotion	# of coded instance
1	Worry	120 (12%)
2	Anger	138 (14%)
3	Fear	61 (6%)
4	**Sadness**	**251 (26%)**
5	Depression	131 (13%)
6	Surprise	70 (7%)
7	**No emotion**	**206 (21%)**
	TOTAL	**978 (100%)**

Young *et al.* [61] deduced that grief, sadness, and disbelief typical of all grief, overwhelming guilt, confusion, rejection, shame, and anger are also often prominent in suicide. From Table 1, the highest coded emotion is *sadness* (26% of the 978 instances*)*. In each case of the coding, *sadness* appeared as the dominant emotional impact on the participants (witnesses). This is followed by *no emotion* category. The implication is that the majority of the witnesses felt *sad* after they had witnessed suicide or suicide attempts. However, the *no emotion* category, which appeared as the second highest, implies that the witnesses did not feel any

emotional impact after witnessing suicide and suicide attempts. This could also mean that the participants declined to share their emotional antecedents as requested during the data collection. The *no emotion* indicates 21% of the entire text instances. This implies that of the 978 participants who accounted for witnessing suicide and suicide attempts, 206 of them did not express any emotions.

From Table 1, the lowest of the emotion categories is *fear*. Fear accounted for 6% in the participants' responses regarding the expression of emotions. The implication of this is that only 61 out of the 978 of the participants expressed *fear* after they had witnessed suicides or suicide attempts. The *fear* emotion category may have been aroused on the grounds that the participants assumed to have been those committing or attempting the suicide. This situation is likely to trigger fear, and could eventually lead to another form of emotion category, such as depression.

4.3. Classifiers performance

Having predicted the emotional impact of suicide from the witnesses' accounts, there was the need to ascertain the performance of the various classifiers used in this work. Table 2 shows the Precision, Recall and F-measure of the various classifiers as elaborated in the earlier section. From the table, the performances of the machine learning classifiers for each of the emotion categories are reported, including the overall performance of the classifiers. Aside the *no emotion* category, the highest predicted emotion by the SVM classifier is *depression*. To add is that, the SVM classifier yielded 90% accuracy for predicting depression. With this accuracy, the SVM produced 92% Recall and 87% Precision from predicting *depression* in the text corpus. However, *worry* was the lowest of the emotion categories which was predicted by the SVM classifier though this exceeded the general acceptable threshold of 70%. This implies that, of the 131 instances that expressed *depression*, 92% of them were identified and tagged, however only 87% of the identified proportion of the instances were accurately tagged and matched with the *gold standard*. This score is higher than the threshold of the inter-coder agreement of the 75% reported earlier.

In the same vein, the highest individual emotion category predicted by the MNB classifier is *depression*. From Table 2, the MNB classifier performed satisfactorily as the Recall, Precision and the F-measure yielded 88%, 90% and 89% respectively. These scores are beyond the generally acceptable 70% threshold. The implication is that while 88% of the 131 instances of the data was identified and recognised by the classifier, 90% was accurately predicted as the *gold standard* instance of the corpus. Just like the SVM, *worry* was found to be the lowest in terms of predicting emotions in the text.

In the nutshell, the best performing classifier is the SVM (accuracy = 78%), of which the Precision, Recall and F-measure exceeded the acceptable threshold of 70%. Although the difference in the performance is not significant, both classifiers can be concluded as suitable for predicting emotions in the witnesses' accounts with regards to suicide or suicide attempts.

Although the two classifiers had scores greater than the threshold, the SVM classifier was superior over the MNB. We conclude that machine learning, as compared with humans, can predict emotions in text that can be likened to humans.

Table 2: Evaluation results of the various machine learning predictions

Emotion	SVM			MNB		
	R	P	F	R	P	F
Worry	73	67	70	67	69	68
Anger	82	75	78	73	80	76
Fear	76	89	72	72	71	72
Sadness	78	72	75	68	65	67
Depression	**92**	**87**	**90**	**88**	**90**	**89**
Surprise	74	77	75	78	74	76
No emotion	**93**	**94**	**94**	89	90	90
TOTAL	**81**	**80**	**79**	**66**	**67**	**67**

5. Conclusion

We investigated the emotional impact of suicide on incidental witnesses using Machine Learning classification techniques. Usually, this domain of research is overlooked as many researchers focus on the cause and impact of suicide behaviour. To do this, we collected subjective data about some suicidal antecedents of people from a random setting. The collected data was pre-processed into corpus for suitability and easy annotation with predefined emotion categories adopted from Plutchik. The corpus, which was coded with the predefined emotion categories, was further analysed using Machine Learning classifiers – SVM and MNB. We employed 10-folds cross validation strategy in the classification of emotions in the corpus.

With regard to the manual analysis of the emotions in the corpus by well-trained coders, *sadness* emerged as the most salient emotion expressed by the participants (witnesses) while *fear* was found to be the lowest emotion category expressed by the participants (witnesses). However, the machine learning prediction yielded highest in predicting depression with insignificant variations in the other emotional categories. This nonetheless shows that people who have witnessed suicide or suicidal attempts are inherently affected by some form of emotions that may require urgent attention to alleviate. In addition, through the witnesses' accounts, the study revealed that suicide and suicidal attempts occur in the age of 15 to 24, with men being the highest victim.

By evaluating the performance of the classifiers, the SVM classifier emerged as superior in terms of predicting the emotional impact of suicide on the witnesses (participants). By comparing the performance of the various classifiers with humans, we conclude that the Machine Learning algorithms can predict emotions in text as humans. The outcome of the study contributes to the pool of research that sought to advocate for the use of machine learning for predicting social phenomenon, especially in text, for complementing decisions. Additionally, the finding has demonstrated the emotional impact of suicide on incidental witnesses.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] World Health Organization. World health statistics 2015. World Health Organization, 2015.
[2] UNICEF. Study of prevalence and dynamics of suicide among children and young people (12-24 years of age) in SUGHD REGION, TAJIKISTAN. 2013, pp. 1-88.

[3] R.M. Berman, A. Cappiello, A. Anand, D.A. Oren, G.R. Heninger, D.S. Charney & J.H. Krystal. "Antidepressant effects of ketamine in depressed patients." Biol Psychiatry. 47(4):351-4, 2000.

[4] W. Eisele, D.T. Reay & A. Cook. "Sites of Suicidal Gunshot Wounds", J Forensic Science, 26 (3), 1981.

[5] D. Klonsky & M.M Alexis. "The Three-Step Theory (3ST): A New Theory of Suicide Rooted in the Ideation-to-Action. Framework". International Journal of Cognitive Therapy, 8(2), 114–129, 2015.

[6] HKJC Centre for Suicide Research and Prevention, University of Hong Kong. Method Used in Completed Suicide, 2006. Available at http://csrp.hku.hk/WEB/eng/statistics.asp#3.

[7] J.J. Card. "Lethality of suicidal methods and suicide risk: Two distinct concepts", Omega 5, 1974.

[8] V. Admiraal,. "Guide to a Humane Self-Chosen Death Alt Suicide Holiday". Accessed on 24ᵗʰ April 2018 on http://ash2.wikkii.com/wiki/Helium, with their information taken from a range of sources but including, 2006.

[9] P. Nitschke, & F. Stewart. "The Peaceful Pill eHandbook", 2009.

[10] P. Passos & J.Y Chow, "Interpersonal coordination in team sports," in *Interpersonal Coordination and Performance in Social Systems*, ed. J. Y. Chow (New York, NY: Routledge), 154–164, 2016.

[11] M.J. Just, L. Pan, V. L. Cherkassky, D.L. McMakin,, C. Cha., M.K. Nock & D. Brent. "Machine learning of neural representations of suicide and emotion concepts identifies suicidal youth". Nature Human Behavior, 2017. October 30. doi:10.1038/s41562-017-0234-y

[12] E.O. Wilson. "The Meaning of Human Existence". 1ˢᵀ ed. W. W. Norton & Company, 2014, 192pages.

[13] G.A. Van-Kleef, A. Cheshin, A. H. Fischer & I.K Schneider.. "Editorial: The social nature of emotions". *Frontiers in Psychology*, 7(896), 1–5, 2016.

[14] D. Keltner & J. Haidt. "Social functions of emotions at four levels of analysis". *Cognition and Emotion* 13: 505-21, 1999.

[15] P.M. Niedenthal & M. Brauer. "Social Functionality of Human Emotion" Annual Review of Psychology, Vol. 63, pp. 259-285, 2012.

[16] A.S. Morris, J.S. Silk, L. Steinberg, S.S. Myers & L.R. Robinson. "The Role of the Family Context in the Development of Emotion Regulation". Social Development. Author manuscript; Published in final edited form as: Soc Dev. 16(2): 361–38, 2007.

[17] D. Keltner & J.S Lerner. "Emotion." In The handbook of social psychology, ed. DT Gilbert, ST Fiske, G Lindzey, pp. 317-52. New York, NY: Wiley, 2010.

[18] P. Ekman. "Emotions revealed: Recognizing faces and feelings to improve communication and emotional life". New York, NY: Holt, 2007.

[19] E.A. Kolog, C.S. Montero & M. Tukiainen. "Development and Evaluation of automated e-counselling system for emotion and sentiment analysis". The electronic journal of information systems evaluation. 21(1), 1-19, 2018.

[20] D. Keltner, K. Oatley, J.M. Jenkins. "Understanding emotions". Hoboken, NJ: Wiley, 2004.

[21] R. Coughlan & T. Connolly. "Predicting affective responses to unexpected outcomes." Organizational Behavior and Human Decision Processes, 85(2), 211-225, 2001.

[22] B.A. Mellers, "Choice and the relative pleasure of consequences". *Psychological Bulletin* 126: 910-24, 2001.

[23] N.H. Frijda & B. Mesquita. "The social roles and functions of emotions". In S. Kitayama & H. R. Markus (Eds.), *Emotion and culture: Empirical studies of mutual influence.* Washington, DC: American Psychological Association, 1994.

[24] B. Parkinson. "Emotions are social". *Br. J. Psychol.* 1996, 87, 663–683. doi: 10.1111/j.2044-8295.1996.tb02615.x

[25] N.E. Dowd, D.G. Singer & R.F. Wilson, eds. "Handbook of Children, Culture, and Violence". California: Sage, 2006.

[26] D.A.Brent, J. Perper, G. Moritz, A. Friend, J. Schweers, C. Allman, L. McQuiston,M.B. Boylan, C. Roth, & L. Balach. "Adolescent witnesses to a peer suicide". Journal of American Academy of Child and Adolescent Psychiatry, 32(6): 1184-1188, 1993.

[27] I.T. Young, A. Iglewicz, D. Glorioso, N. Lanouette, K. Seay, M. Ilapakurti, & S. Zisook. Suicide Bereavement and Complicated Grief. *Dialogues in Clinical Neuroscience*, 14(2): 177–186, 2012.

[28] P. Goldblum, D. L. Espelage, J. Chu, & B. Bongar eds.. "Youth Suicide and Bullying: Challenges and Strategies for Prevention and Intervention". Oxford: Oxford University Press. 2015.

[29] F.B. Ribeiro, E.A.T. Lanna, M.A.D Bomfim, J.L. Donzele, M. Quadros, P. Cunha & S.L de. "True and apparent digestibility of protein and amino acids of feed in Nile tilapia". Rev. Bras. Zootec., 40 (5): 939-946, 2011

[30] K. Posner, G.K. Brown, B. Stanley, D.A. Brent, K.V. Yershova, M.A. Oquendo, G.W. Currier, G.A. Melvin, L. Greenhill, S. Shen & J.J. Mann. "The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults". Am J Psychiatry,, 168(12):1266-77, 2011.

[31] M.A. Oquendo, G.M. Sullivan, S. Katherin , E. Baca-Garcia, B.H. Stanley, M.E. Sublette & J.J. Mann. "Toward a Biosignature for Suicide". The American Journal of Psychaitry , 171 (12,), pp. 1259-1277, 2014.

[32] E.A. Selby, M.A. Anestis, W. Theodore & E. J. Thomas E. J. "An Exploration of the Emotional Cascade Model in Borderline Personality Disorder" J Abnorm Psychol., 118(2): 375–38, 2009.

[33] N. Garnefski, V. Kraaij, & P. Spinhoven. Negative Life Events Cognitive Emotional Regulation and Emotional Problems. Personality and Individual Differences, 30, 1311-1327, 2001.

[34] R.L. Tamas, D. Menkes & R.S. El-Mallakh. S"timulating research: a prospective, randomized, double-blind, sham-controlled study of slow transcranial magnetic stimulation in depressed bipolar patients." Journal of Neuropsychiatry Clin Neurosci;19:198–199, 2007.

[35] T. Joiner. "Why people die by suicide". Cambridge, MA, US: Harvard University Press; 2005

[36] K.A. Van Orden, T.K. Witte, K.C. Cukrowicz, S.R. Braithwaite, E.A. Selby & T.E. Joiner Jr. "The interpersonal theory of suicide". Psychological review, 117(2), 575, 2010.

[37] C. Zlotnick, T.M. Shea, K. Rosen, E. Simpson, K. Mulrenin, A. Begin, *et al.* "An affect-management group for women with posttraumatic stress disorder and histories of childhood sexual abuse". Journal of Traumatic Stress, 10, pp. 425-436, 1997.

[38] T.J. Lynch, D.W. Bell & R. Sordella. "Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib". *Engl Journal of Med.; 350: 2129–2139, 2004.*

[39] J.O. Christensen & S. Knardahl. "Work and headache: a prospective study of psychological, social, and mechanical predictors of headache severity". Pain, 153(10):2119–2132, 2012.

[40] I. Lerner, B.C. Armstrong, & R. Frost. "What can we learn from learning models about sensitivity to letter-order in visual word recognition?" Journal of Memory and Language, 77, 40–58, 2014.

[41] M.D. Anestis, C.L. Bagge, M.T. Tull, & T.E. Joiner. "Clarifying the role of Emotion deregulation in the interpersonal-psychological theory of suicidal behaviour in an undergraduate sample". Journal of Psychiatric Research*, 45*, 603-611, 2011.

[42] S.P. Whiteside & D.R. Lynam. "The Five Factor Model and impulsivity: using a structural model of personality to understand impulsivity." Personality and Individual Differences*,, 30*(4), 669-689. 2001. doi: 10.1016/S0191-8869(00)00064-7

[43] M.D. Anestis & T.E. Joiner. "Examining the role of emotion in suicidality: Negative urgency as an amplifier of the relationship between components of the interpersonal–psychological theory of suicidal behavior and lifetime number of suicide attempts". Journal of affective disorders, *129*(1), 261-269, 2011.

[44] Baumeister, R. F. "Suicide as escape from self". Psychological Review*, 97*(1), 90-113, 1990.

[45] A.L. Thomas. "Suicidal Thoughts and Behaviour among Black College Students: Examining the Impact of Distress Tolerance and Social Support on Suicidality." Masters Theses & Specialist Projects, 2015.

[46] K. Rajappa, M. Gallagher, & R. Miranda. "Emotion dysregulation and vulnerability to suicidal ideation and attempts." Cognitive Therapy Res, 2012*, 36*-833-839. DOI:10.1007/s10608-011-9419-2

[47] B.L Fredrickson. "The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions". American Psychologist*,, 56,* 218-226. 2011.

[48] E.A. Kolog, C. Suero Montero, & e. Sutinen. "Annotation agreement of emotions in text: The influence of counsellors emotional state on their emotion perception". In proceedings of international conference on advance learning technologies 2016.

[49] E.A. Kolog, & C. Suero Montero. "Towards automated e-counselling system based on counsellors' emotion perception". Education and information technologies, pp. 1-23, 2018.

[50] M. Hall, E. Frank., G. Holmes, B. Pfahringer, P. Reutemann, & I.H. Witten, "The WEKA Data Mining Software: An Update. SIGKDD Explorations", 2009.

[51] E.A. Kolog, C.S. Montero & T. Toivonen. "Using machine learning for sentiment and social influence analysis in text". In: Rocha, A., Guarda, T. (eds.) Proceedings of the International Conference on Information Technology & Systems (ICITS 2018). AISC, vol. 721, pp. 453–463. Springer, Cham (2018).

[52] A. Tripathy, A. Anand & S.K. Rath. "Document-level sentiment classification using hybrid machine learning approach". Knowledge and Information Systems. 2017; 53(3):805-31.

[53] Cohen, J. "Coefficient of agreement for nominal scales." Educational and Psychological *Measurement, 20*(1), 37–46, 1960.

[54] E. A. Kolog. Detecting Emotions in Students' Generated Content: An Evaluation of EmoTect System. In *International Conference on Technology in Education*, 2018, pp. 235-248. Springer, Singapore, 2018.

[55] N. Ibrahim, N. Amit, N.C. Din & H.C. Ong. Gender differences and psychological factors associated with suicidal ideation among youth in Malaysia. Psychology research and behaviour management, *10*, 129, 2017.

[56] American foundation for suicide prevention (AFSP). "Suicide statistics". Available on <https://afsp.org/about-suicide/suicide-statistics/> Accessed on the 13th April 2018.

[57] World Health Organization. "World health statistics 2016: monitoring health for the SDGs sustainable development goals". World Health Organization, 2016.

[58] J. Osafo, C.S. Akotia, J. Andoh-Arthur & E.N.B. Quarshie, "Attempted suicide in Ghana: Motivation, stigma, and coping." Death studies, 39(5), 274-280. 2015.

[59] R. Plutchik. Emotion: Theory, research, and experience: Theories of emotion. New York: Academic, 1, 399, 1980.

[60] L. Loyo, E. Martínez-Velázquez & J. Ramos-Loyo. "Influence of emotions on executive functions in suicide attempters." Suicidal Online, *4*(1), 42-55, 2013.