ISSN 2286 – 9131

Transactions

on Computer and Information Technology

Volume 11, No. 1, May 2017

ECTI-CIT

Research Article

ECTI Transactions on Computer and Information Technology
ECTI Association
Sirindhon International Institute of Technology
131 Moo 5 Tiwanon Rd.,Bangkadi
Muang, Pathumthani 12000,Thailand
Tel: 02-5012578
Email: editorcit@ecti.or.th
http://www.ecti-thailand.org

An Efficient Search Algorithm for Multi-hop Network Localization in Sparse Unit Disk Graph Model

Phisan Kaewprapha[1], **Thaewa Tansarn**[2], and **Nattakan Puttarak**[3], Non-members

ABSTRACT

We consider a network localization problem by modeling this as a unit disk graph where nodes are randomly placed with uniform distribution in an area.The connectivity between nodes is defined when the distances fall within a unit range. Under a condition that certain nodes know their locations (anchor nodes), this paper proposes a heuristic approach to find a realization for the rest of the network by applying a tree search algorithm in a depth-first-search manner. Our contribution is to put together priori information and constraints such as graph properties in order to speed up the search. An evaluation function is formed and used to prune down searching space. This evaluation function is used to select the order of the unknown nodes to iterate to. This paper also extends the idea further by accommodating various other properties of graph into the evaluation function. The results show that node degrees, node distances and shortest paths to anchor nodes drastically improve the number of iterations required for realizing feasible localization instance both in noise-free and noisy environments. Finally, some preliminary complexity analysis is also given.

Keywords: Localization, Tree Search, Unit Disk Graph, Wireless Sensor Networks

1. INTRODUCTION

In sensor networks deployments and applications, geographical information along with measured data from the sensors is as crucial; for instance, representing sensed data on a map. Measuring data, for example, traffic flow will be less meaningful if not geographically represented. It is common that stationary sensing nodes are placed in pre-known locations. In different circumstances, sensor network deployments may be done in a random manner, for example, adhoc sensor nodes are air dropped into some specific area. Determining locations is not as simple as ones would expect.

Establishing a communication network is certainly a challenging research area, however, finding a geographically deployed location of each individual device is as challenging as the communication part. Although devices could be GPS-equipped, the cost of doing so could add up and prohibit such usage in practical scenarios. Hence, a more realistic assumption is that some of the nodes know their locations while the majority of them do not. Each sensor node may or may not have direct connection with the anchor nodes and often rely on multiple-hops communication to the anchors. This specific instance is known as multi-hop network.

Unlike single-hop localization such as GPS system where location can be computed by lateration technique (finding the intersection of circles drawn around reference nodes), multi-hop network can be modeled and solved in different ways. Researchers have paid attention for this kind of problems for a few decades and this is still one of the active research areas [4]. We will provide some detail in the literature review section. It is also worth mentioning that our studies are based on 2-dimensional space, although 3 or higher dimension can also take advantage from our approaches.

There are several research directions and assumptions regarding the network localization problem. Some approaches are based on real-word scenarios where physical properties such as noise and nature of signal propagation are taken into account. Others work on simplified models so that fundamental understanding and knowledge can be derived [4-6]. Even the problem is formulated in its simplified form where network is represented as graph and distance measured is exact, finding the feasible solution can be difficult [7].

Our approaches to investigate this problem started from an abstract setting where we model the problem as connected graph with its simplest form. Then we are looking into heuristic techniques to tackle this kind of localization problem. Along the line we add more realistic assumptions and constraints to meet the real-world scenarios while maintaining the acceptable average complexity. In this paper, we propose

Manuscript received on September 15, 2016 ; revised on March 6, 2017.

Final manuscript received on March 6, 2017.

[1,2] The authors are with Department of Electrical and Computer Engineering, Faculty of Engineering, Thammasat University, Thailand, E-mail: kphisan@engr.tu.ac.th, debugger_man@hotmail.com

[3] The author is with Telecommunications Engineering department, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang (KMITL), Thailand, E-mail: kpnattak@kmitl.ac.th

a centralized algorithm capable of computationally solving the problem based on the use of simple yet effective tree search algorithm.

The complexity of the original exhaustive search space for a realized instance of the network can be improved by utilizing the priori knowledge from some properties derived from the graph itself. The results show that computationally-prohibited searching space can be greatly reduced when some side information from graph properties are used to assist the search and eliminate some search paths sooner with high probability. The right order of search paths (or the order of traversal) significantly affects the converging time to the final solution.

Our method can be applied in several real-word scenarios. Since this algorithm works well in sparse wireless network with a few numbers of anchor nodes and limit communication range, localization in ultra-wideband sensor network is one of the deployment scenarios. Sensed data along with measured distances among nodes will be passed along to a gateway which forwards data to a central server for post-processing. The data from sensors and locations will be extracted in a centralized manner.

The paper is organized as follows. We discussed some related work and background in section II. Then in section III we describe our algorithm and the rationale behind the chosen properties and give out the simulation results in section IV and finally the discussion and conclusion are given in section V and VI respectively.

2. RELATED WORKS

Network localization, especially for wireless sensor networks, can be viewed as two-step processes. The first step is to obtained measured data regarding spatial relationship between nodes (their coordinates x, y). The relationship is naturally obtained as polar form (r, θ). If we know both parameters in polar form, the solution is considered found. Hence the challenge is when only one of the parameters is available. For example, omni-directional antennas are unable to sense the direction or angle of a received signal.

Physically, the distances or angles can be obtained by different measurement techniques. It can be measured by the time of arrival of the propagated signal (TOA) or by the received signal strength (RSS) or by the angle of arrival (AOA) or by other means [4]. These measured data solely or along with other side information are the input to different approaches used to infer the actual locations. We often call this process as the network realization. Such techniques include tri-lateration, bearing measurement, etc [4-6].

In a more complex environment of network localization such as multi-hop network localization where multiple hop communication is required to reach the other end, connectivity and graph properties such as hop-count are used as an input to solve this problem, example of this approach can be found in [8,11].

Apart from real-world scenarios, theoretical work has also been established [15-16] where the clean state of the network is studied and some fundamental theories have been established.

In both theoretical and physical problem settings for network localization, it can be universally defined as an optimization problem where the discrepancy between the measured distances and the calculated distances from the realized instance is minimized. Direct approaches such as some variations of non-linear optimization solving techniques are also proposed. The problem is also approximated as linear or convex optimization, which can be tackled by various solving techniques, e.g. maximum likelihood estimation, semi-definite programming, spring mass model, belief-propagation and message passing algorithm [8-14].

Our work, as mentioned earlier, fall into the area of clean state of the network in its simplest form while making use of links and connectivity information to solve the problem similar to [8-24], with a different approaches. Our method is based on tree search algorithm inspired by [17]. Our original proposed version can be found in [2] and that will be briefly explained in the next section. This paper is an extended version from our previously published work [3]. In this version we have add preliminary analysis of the complexity bounds for our algorithm as well as open problems needed to be addressed.

2.1 Theory of Network Localization

We briefly discuss the theory of network localization in this section as it has some implication from the fundamental result that we can use to explain the effect of our heuristic approaches. In [14], the network is modeled as graph, we follow the same settings. Below are some useful results:

- In network localization, the network can be solved if it has a unique realization. The condition is that the graph must be globally rigid. This is the case when we use only edge information. We showed in our previous work that this is not a necessary condition if we use implicit information about non-connected node [1].
- Solving unit disk graph localization is NP-hard, however our approach can reduce the iteration times for the average cases.
- If a graph is dense enough, it will be solved efficiently by tri-lateration method. This implies that our approach is useful for sparse network.

In the next section we explain our method and improvement that increases the speed of finding solution.

3. TREE SEARCH LOCALIZATION

3.1 Problem Formulation

We model a group of communication devices forming a network by establishing communication to neighboring devices as a graph G consisting of a set of vertices (we use node interchangeably) V and a set of edges (we use connection or link) E connecting between two nodes. $V(i)$ refers to an i^{th} node labeled as node i where i is an integer from 1 to n. $E(i,j)$ represents an edge between node i and node j.

Additionally, we assume that locations are aligned in 2-dimensional integer coordinate system where nodes positions are integer numbers. Hence the distance square will also be integer. This assumption enables us to apply search algorithm with finite number of branches. Granularity and complexity are the trade-off that needs to be considered. So far small to medium size network with 200 or fewer participating nodes are studied. Nodes are placed in a uniformly distributed randomly in a unit square area. Two nodes establish connectivity if and only if they are within a certain distance r apart from each other. We define $D(i,j) = |(E(i,j))|$ to be the distance between node i and j obtained from the measurement and let $X(i)$ denote the 2-dimensional coordinate for a node i. Then $|X(i) - X(j)|$ represents the calculated Euclidian distance according to the instance of the realization. We can establish a realization of a graph as an optimization problem with constraints as follows:

Minimize:

$$\Sigma \| X(i) - X(j)| - |E(i,j)\|, \; E(i,j) \in E \tag{1}$$

Subject to:

$$X(i) = x, y \tag{2}$$

for some nodes i that are in a set of anchor nodes.

Our search algorithm is searching into exact solution where $||X(i)-X(j)|-|E(i,j)|| = 0$; in the other words, $|X(i)-X(j)| = |E(i,j)|$. Fig.1 depicts an example of our network instance. The grid size is 20x20 unit with 40 nodes are randomly placed. The radius is 5 unit distance.

3.2 Tree Search Algorithm

Proposed in [1], the algorithm can be briefly described as follows: Starting from a set of anchor nodes, this set can be called "realized network instance", while the other nodes that are not placed into this realized instance are kept in another set. The search tree starts from this instance. The algorithm traverses through the search tree by adding a new node that is not in the set and checks if a new node is compatible with the objective function in Eq.(1), e.g. $|X(i) - X(j)| - |E(i,j)| = 0$ if they are neighbors and $|X(i) - X(j)| > 1$ if they are not neighboring nodes.

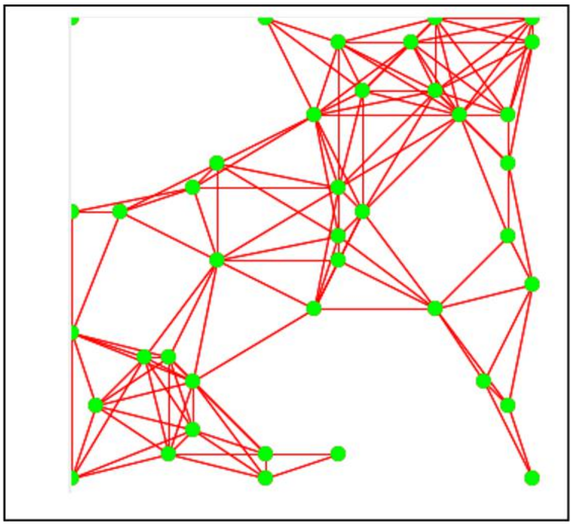


Fig.1: *An instance of random unit disk graph.*

If the new node is not compatible, then the algorithm discards the instance and moves on to the next path in the tree in a depth-first-search style. If the node is successfully placed without conflicting with any constraint, the new node will be added to the current instance of the realized graph. The algorithm resumes until all nodes have been localized. Table 1. summarizes the method. Starting with a set of anchor nodes and let $R\{V\}$ denote a set of nodes that are realized. $R\{V\}$ is empty initially (or containing only anchor nodes) and realized nodes will be added to the set as the algorithm tries to place nodes into feasible locations that are consistent with measured data. Another set called $I\{V\}$ represents unrealized nodes. During each iteration, the algorithm picks a node from $I\{V\}$ according to $Eva(u)$ function, which is used to select a node with a better chance of constructing the final feasible solution as the search algorithm continues.

The algorithm relies on the probability that a new node selected will be more likely to converge to the final solution faster, while discarding the wrong search paths sooner. Our algorithm lies on the same concept like decision tree used in computer chess games [17]. We iterate through search tree and eliminate infeasible branches while applying evaluation function to increase the chance of eliminating infeasible branches sooner. Hence this speeds up the search. It is similar to chess or board games where evaluation function is used to select the best move to win the game. However, our evaluation function is used to select a move that will likely result in reaching the solution faster (e.g. fewer iterations or walk-through). Another difference is that our algorithm is depth-first-search which is efficient in term of memory usage while games may require breadth-first-search approach.

Our contribution in this paper is mainly exploiting intrinsic/implicit properties of graph. Then we use these properties as part of an evaluation function to determine the order of node being added or placed during the process of constructing the graph. Properties of graph such as connectivity, node-to-node distance, connectivity count from a node to its neighbors (node's degree) as the mean for further enhancement/speeding up the search algorithm. We also give out preliminary analysis of the complexity of this algorithm.

Table 1: *Tree-search algorithm [1].*

1.	Start with all known nodes in a set $R\{V\}$ which represents realized nodes and $I\{V\}$ contains nodes which are waiting to be realized.
2.	Sort nodes in $I\{V\}$ in dependence order according to the $Eva(u)$ and choose a node u with largest value of $Eva(u)$.
3.	Generate all feasible realization of u centered at a neighboring node v in $R\{V\}$ and add all the instances to the search tree and put u in $R\{V\}$
4.	Repeat step 2 until all nodes in $I\{V\}$ are realized (empty).

3.3 Search Constraints

In the conventional problem formulation for network localization, the only constraint used for realizing the network is the connectivity and distances such as $||X(i)-X(j)|-|E(i,j)||=0$. We include an assumption that the implicit information such as no connection between nodes can imply that the nodes are not within the range and this can be used to discard iteration into graph instance that conflicts with this constraint, e.g. $|X(i)-X(j)|>1$. Similar approach is done in [13] by transforming this information into virtual links.This priori constraint gives significant improvement to the size of the searching space, especially for sparse networks. Similar concept is applied to urban outdoor localization of mobile devices where buildings are excluded from feasible locations for the devices.

3.4 Graph Properties

We look into graph properties that lead to the selection of a new node with high probability of eliminating unsuccessful paths. We examine and apply those properties, individually. Those properties are defined as, for a given node that is being placed:

- Nc: Number of links whose end nodes locations are known.We consider a new node that has more links to the known/placed nodes by counting neighboring nodes that location is realized.
- D: Degree of a new node: similar to Nc but consider all neighboring nodes regardless of their realization statuses by counting the number of neighbor nodes, including those that are not realized.
- Sd: Sum of distances from the realized nodes.
- Ss: Sum of shortest paths between a new node and all anchor nodes.

Fig. 2, 3, and 4 demonstrate that the appearance of those parameters. Fig. 2 depicts the case where node A is chosen to be placed because it has more connectivity to the instance of realized network rather than node B that has more total connectivity. Fig. 3 shows that node degree of B is greater than A. Hence B will be examined first. Fig. 4 shows that the distance from realized graph from A is greater than B. Thus A is chosen first.

There are other possible properties such as sum of the shortest path to all nodes or to all realized nodes. Max/min or average or count of those parameters can also be studied. Those parameters will be studied extensively as our future work.

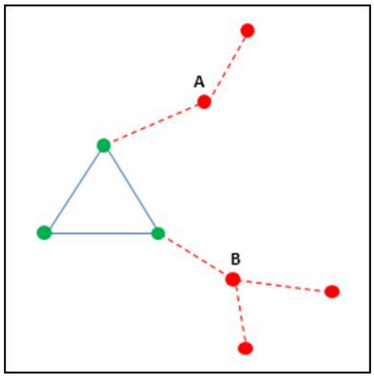


Fig.2: *Example of most connected to a realized graph.*

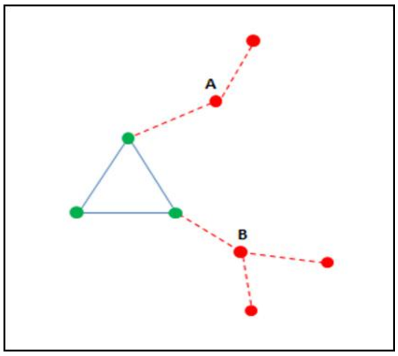


Fig.3: *Example of node degree preference.*

3.5 Evaluation Function

Our evaluation function is in a very general form as the sum of the function of each individual property.

$$Eva(n) = \Sigma Fi(Pi) \tag{3}$$

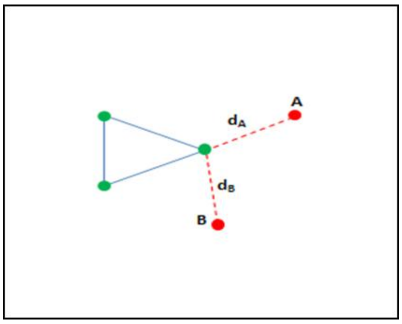


Fig.4: *Distances between unrealized nodes and realized graph.*

where $Eva(n)$ is the evaluation function of a node n. $Fi()$ is a function of a parameter Pi. We currently use the simplest from by using linear weight to each parameter. By putting in the parameters studied in this paper, we obtain the following evaluation function.

$$Eva(n) = wl * Nc + w2 * D + w3 * Sd + w4 * Ss \quad (4)$$

where $w1, w2, w3$ and $w3$ are the coefficient for those parameters. Any node n with higher $Eva(n)$ will be placed into a realized graph first.

3.6 Tree-search Algorithm in Noisy Measurement

We also consider the case where noise is part of the measurement. Thus, the measured distances are contaminated by noise. The original search algorithm, unfortunately, will not be able to find the exact solution because the added distance causes the disagreement in the measured data and the distance obtained while placing a node into the graph during the runtime of the tree-search algorithm.

We modify the search algorithm by relaxing the constraint. As noise varies, we vary a threshold t so that the magnitude of the disagreement between the measured distance and the calculated distance is within this threshold t. Any search result that is compatible with this new relaxed constraint will be regarded as a feasible solution.

Previously, under the ideal assumption that noise-free distance measurement is obtained, it is a matter of how much complexity is needed to find the exact solution (if a unique solution exists) or the first feasible solution found (in case there are ambiguities as the graph is not rigid or side information is insufficient). In the latter case when noise is present, even though the feasible solution is obtained after performing the algorithm, there is no guaranteed of the exact solution.

Fig. 5 depicts the extended bound t where the red-dot circles represent the original possible search boundary according to Eq. (1) and (2) the solid-blue circles from inner to outer ones represent an area in which the integer coordinate will be searched. The relaxation can be added into the equation as follows:

$$\Sigma||X(u) - X(v)| - d(u, v)| \ \leq \ t \quad (5)$$

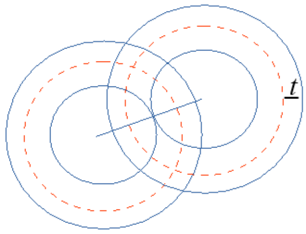


Fig.5: *Extended Search Boundaries with a threshold t.*

4. SIMULATION SETUPS AND RESULTS

We prepare and perform the simulation for both noise-free and noisy environment as follows:

4.1 Simulation Setup

Since the searching space of any random network can be huge and there is no known analytical closed-form results regarding the performance of this approach at this moment. Thus we demonstrate the idea that the algorithm is able to operate efficiently in real-word at moderate network size. We put the algorithm into the test by implementing the search in parallel using multi-thread on multiple processing cores available on modern CPUs. We simulate our algorithm on a system equipped with an Intel i7-5960X consisting of 8 computing cores and 16 logical threads running at 4.0 Ghz with 32GB of RAM. The algorithm is implemented using Java JDK 1.8. Each of the simulation setting is run multiple times. Each time, a new random instance of graph is used.

We simulate through 1 million samples of graphs per one setting. The number of 1 million experiments is obtained in a Monte Carlo fashion. The average running time for each experiment ranges from 10 minutes to 21 hours.

To extend the effect of our algorithm, we choose to work with a sparse network similar to the one depicted in Fig. 1. Assuming that the unit length is meter (m), we simulate the network in an area of 15×15 m containing 37 nodes with each node has communication range of 1 m and we put 4 anchor nodes at the corners of the graph. Firstly, we run our algorithm with randomly placed nodes; the average performance is 30,000 iterations. This is due to the fact that the graph is sparse.

The simulation is done for each individual parameter as well as the combined evaluation parameters.

The theoretical lower bound for the iterations is 33 (as we have 33 nodes). We count a placement of a new node as 1 iteration.

4.2 Simulation Setup in Noisy Environments

Similarly, in noisy environment we perform simulation on the same setting with node radius of 4. There are 4 anchors nodes at the corners and we also add 2 anchor nodes randomly along with other ordinary nodes. We also modify our evaluation function as in equation (4) by normalizing each parameter of the evaluation function by the maximum value of each parameter of each instance of the random network.

$$Eva(n)=w1\frac{Nc}{|Nc|\max}+w2\frac{D}{|D|\max}+w3\frac{Sd}{|Sd|\max}+w4\frac{Ss}{|Ss|\max} \quad (6)$$

Under this new equation, we then find the new set of weight and parameters and choose the one that performs the best before applying this in the noisy environment.

4.3 Performance of Individual Parameter

By running each individual parameter separately, we obtained the result as shown in Fig. 6. The parameter Nc effectively reduces the number of iterations. The rationale behind this is that a node having more connected links to the already-realized nodes has smaller valid search space. Other parameters may not be obvious; for example, the sum of the distance can be interpreted as follows; placing a new node that is father away allows the algorithm to expand and cover the entire graph faster. This helps to eliminate the infeasible moves quicker. The Shortest Path parameter (Ss) is not performing as good as expected when it is used alone. For the parameter Node Degree (D), it performs better than the Ss, however if D is used alone may cause the algorithm to pick a new node with higher node degree but not relevant to the current instance of the graph that is being realized. This is because D also counts links that are connected to unrealized nodes.

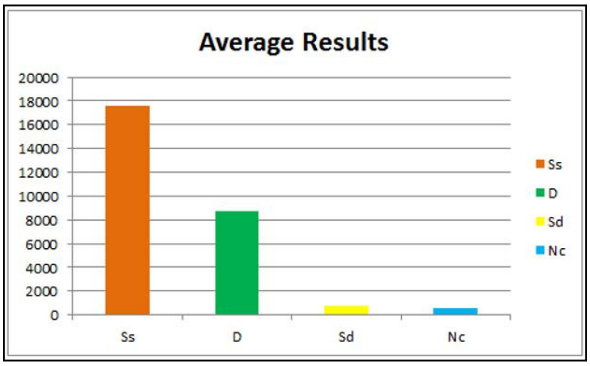


Fig.6: *Average Iteration for Individual Parameter.*

4.4 Performance of Combined Evaluation Function

Next, we use combined evaluation function. The coefficient is assigned from the observed facts according to the previous simulation. Since the parameter Nc performs well, we assign more weight to it followed by the other parameters D, Sd and Ss accordingly. We assign the weight $w1, w2, w3$ and $w4$ to be 4, 2, 1 and -0.1 respectively.

The parameter Ss (sum of the shortest paths) has negative weight because the shorter path implies that it is closer to anchor nodes and that helps to eliminate search path better. The results show that combining parameters does improve the overall performance though the gain is diminishing as shown in Fig.7. Moreover, Nc alone performs on average at 530 iterations. When adding D (node degree), the performance improves and the number of iterations reduces to 390. Similarly, adding Sd (sum of distance) brings the total iterations down to 298 whereas Ss (sum of shortest paths) takes it further to 290. Hence the complexity is reduced by 46% comparing to using individual parameter.

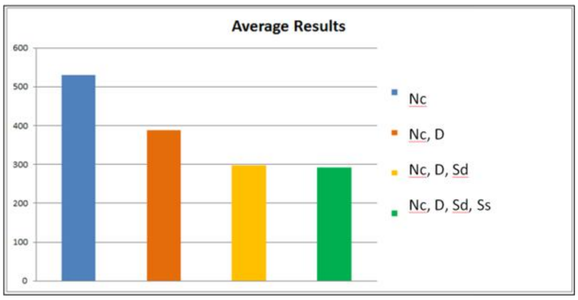


Fig.7: *Average Iteration for Combined Parameters.*

It is interesting to see that when the parameter is used individually, even though it performs better than randomly select node, it performs poorer than using them as a combined evaluation function.

4.5 Performance under Noisy Measurement

As shown in [18], with the evaluation function and search algorithm presented in Table 1, we vary the weight of the function to find the optimal weight that provides the best yield of the average iteration. In Table 2, the optimal iterations obtained is 85 with $w1$=0.8, $w2$=1.4 and $w4$= -0.06. The combined evaluation function significantly performs better than each individual parameter. This weight will be used as we investigate the effect of noise over the algorithm.

We look into two performance metrics that affected by the additive noise. That is the increases of the average iterations and the ability of finding a feasible solution.

For simplicity, we assume that the noise during the measurement follows normal distribution so we apply

Table 2: *Average iteration at various weight ratios of Eva(u).*

Nc	***Sd***	***SS***	Iterations			
			Mean	Min	Max	SD
1	0	0	141	34	22000	446
0	1	0	7745	34	138000	32700
0	0	1	14000	34	75600	621000
0.8	1.4	-0.06	85	34	127	1511

additive Gaussian noise. Eq. (5) can be written as follows:

$$d(u,v) = |E(u,v)| + N(0,\sigma^2) \tag{7}$$

We choose Gaussian distributed noise because the measurement is done over wireless communication channel and the Gaussian is common. Even though other channel's characteristics such as fading maybe considered, we plan to investigate this in the future.

We use the optimum weight from Table 2 as noise varies, we then vary the relaxing threshold t against σ^2 and t is proportional to the measured distanced (e.g. σ^2 varies from 1% to 6% of the distance d). The results are shown in table 3 [18].

Table 3: *Average Iteration at various noise levels.*

***t* (threshold)**	σ^2	Average iterations	% Solution found
0.1	0.01-0.03	96	100
0.1	0.04	1154	19.6
0.1	0.05	103	0.4
0.2	0.04	175	99.8
0.2	0.05	186	95.2
0.2	0.06	214	73.5
0.2	0.07	205	37.3
0.2	0.08	186	11.2
0.2	0.09	104	1.9
0.2	0.10	-	0
0.3	0.09	1148	57.3
0.3	0.10	1054	31.9

According to the result, when the noise level is relatively small i.e. $\sigma^2[0.01, 0.03]$, the relative threshold, for example, $t = 0.1$ is sufficient. t allows the algorithm to tolerate up to some level of uncertainty. Thus, the algorithm reaches a feasible solution eventually. There is a trade-off as we notice that there is a light increase in the average iterations. As noise increased, however, the success rate of finding consistent realization drops drastically. The increase in the complexity is also noticeable when $\sigma^2 = 0.5, t = 0.1$. If we increase the threshold to support the increase of noise, for example, $t = 0.2$, this does allow the solution to be found. The patterns follow with $t = 0.3$, which can tolerate more noise with significant increase of the complexity.

5. DISCUSSION

In our experimental results, even though the algorithm is capable of finding solutions. It should be noted that we are using the assumption that the coordinate of the location is in an integer domain; hence allowing us to discretize the search space and running search algorithm. If the coordinate is in fact embedded in real number coordinate, then this algorithm can still be mapped into our settings by assuming that the quantized from real value location causes some additional error due to the truncation of real coordinate to integer coordinate system. A trade-off between granularity and complexity will have to be taken into account.

It is worth noting that our algorithm is based on centralized processing, the communication overhead is minimal as the information regarding location can be transmitted along with measured data.

We also like to note that graph parameters used as evaluation function are chosen based on the idea that node with more connectivity, larger distances and closer to anchor nodes are more likely to be placed correctly and infeasible choices can be eliminated earlier when evaluated against $Eva(u)$. We found that our current method is useful mainly for sparse graphs. In dense graph, the problem can be solved more efficiently by other algorithms such as the conventional tri-lateration method.

In this section, preliminary approximation of its complexity is given. As we know previously that the lower bound of this algorithm is equal to the number of nodes, in the other words, we place one node per iteration. The upper bound is when we consider that if the first node has k possible coordinate points to be placed without considering any constraints. Then the worst case is expressed in the order of $O(k^n)$ where k is the number of integer points lying inside the outer and inner boundary t shown in Fig.5. k can be estimated by the following relation between integer points inside a circle [18].

$$N(r) = \pi r^2 + E(r) \tag{8}$$

Where $N(r)$ is the number of integer points inside a circle with radius r. And E(r) is small error with upper and lower bound.

$$k = N(r+t) - N(r-t) = 4\pi rt + E(r+t) - E(r-t) \tag{9}$$

This equation (9) represents an estimation of a loose upper bound of the complexity of the algorithm.

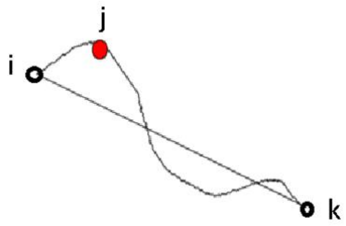


Fig.8: *Node is a member in a shortest path between 2 ends.*

Deriving the upper bound from the radius alone is still a loose bound. We would like to give a conceptual sketch to develop a better upper bound. Consider any node in a network; it is possible to find a shortest path between one end from its adjacent node to an anchor node at the other end as depicted in Fig. 8. Let $s(i, j, k)$ be shortest path along node i, j and k. Normally $s(i, j, k) \leq d(i, k)$. The actual path i to k via node j is usually longer than the direct distance between i and k. The parameter $d(i, k)$ and $s(i, j, k)$ provide another bound to the search space of node j. Then this particular node j can no longer be placed freely within the vicinity of one of its anchor node. It will be bounded inside an area under the triangles formulated from $d(i, k)$ and $s(i, j, k)$ as depicted in Fig. 9.

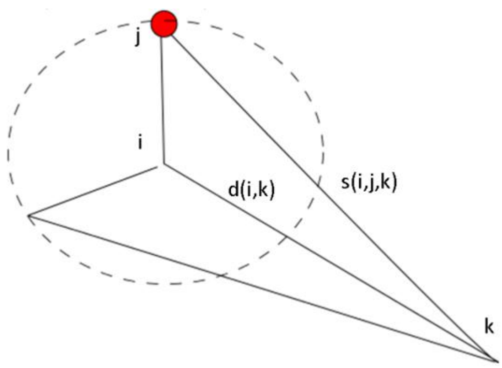


Fig.9: *Max stretched distance while placing node.*

The expression for the upper bound developed based on this idea should be further investigated in our future work. Our primary finding is that the average distance between two random nodes in a unit square is approximately 0.368 [20]. And there is a bound regarding the maximum shortest path (diameter) of a unit disk graph. This diameter is bounded by [21]

$$(1 + 0((\ln \ln n/ \ln n)^{1/d}))/\lambda \qquad (10)$$

where λ is maximum radius of the unit disk graph. Once the 3 sides of the triangle are known or estimated, the bound could be formulated. Nevertheless, the simulation results show that the average complexity is far better than the worst case. Hence finding tighter bounds is also an open problem, which is another interesting problem to be investigated in the future.

One of our directions for the future work consists of utilizing rigidity sub graph where localization can be done efficiently and to build a graph in a bottom up manner. Another direction is to find a tighter analysis of the current complexity of this algorithm. At this moment, our results may not be able to compare directly with other works in the area because our problem formulation and focus is slightly different; we are developing our algorithm to be capable of resolving general and more realistic model. At the same time, we may add more useful constraints and condition that is already embedded in the properties of the graph itself but we have not utilized.

6. CONCLUSIONS

We have proposed an evaluation function to be used in our tree search algorithm for solving network localization problem. Evaluation function is a function that is derived from a combination of graph properties. The results show that those properties, when apply individually, can reduce the algorithm average iterations. The number of connection to the realized graph, Nc, significantly reduces the number of iterations. Furthermore, the combination of the parameters D, Sd and Ss can significantly reduce the average iterations under noise-free and noisy environments by 46%. In the case of noisy environments, the threshold t can be adjusted in order for the algorithm to tolerate the noise at different levels. We also provide the outline for the complexity analysis.

ACKNOWLEDGEMENT

This research was partially supported by Faculty of Engineering, Thammasat University, Thailand. We thank the editors at ECTI-CIT who assisted us during the submission process. We would also like to thank all the reviewers for their valuable comments.

References

[1] P. Kaewprapha, J. Li and N. Puttarak, "Network Localization on Unit Disk Graphs," *2011 IEEE Global Telecommunications Conference - GLOBECOM 2011*, Houston, TX, USA, pp. 1-5, 2011.

[2] P. Kaewprapha, "Localization in Wireless Sensor Networks: Solvability Improvement Technique Using Priori Information from Sensing Data and Network Properties in Unit Disk Graph Model", *Advanced Materials Research*, Vols. 931-932, pp. 999-1003, 2014.

[3] P. Kaewprapha, and N. Puttarak, "Network Localization using Tree Search Algorithm: A heuristic search via graph properties", *2016 13th*

International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Chiang Mai, Thailand, pp.1-5, 2016.

[4] L. Cheng, C. Wu, Y. Zhang, H. Wu, M. Li, and C. Maple, "A Survey of Localization in Wireless Sensor Network," *International Journal of Distributed Sensor Networks*, January 2012.

[5] G. Mao, B. Fidan and B. D.O. Anderson, "Wireless sensor network localization techniques," *Computer networks*, Vol.51, Issue 10, pp. 2529-2553, 11 July 2007.

[6] A. Pal, "Localization algorithms in wireless sensor networks: Current approaches and future challenges," *Network Protocols and Algorithms*, Vol. 2, No. 1, pp.45-73, 2010.

[7] H. Breu and D. G. Kirkpatrick, "Unit disk graph recognition is NP-hard," *Computational Geometry*, Vol. 9, Issues 1, pp.3-24, Jan. 1998.

[8] D. Niculescu and B. Nath, "DV based positioning in ad hoc networks," *Telecommunication Systems*, pp. 267-280, 2003.

[9] A. T. Ihler, J. W. Fisher, R. L. Moses and A. S. Willsky, "Nonparametric belief propagation for self-localization of sensor networks," in *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 809-819, April 2005.

[10] S. Lee, H. Woo, and C Lee, "Wireless sensor network localization with connectivity-based refinement using mass spring and Kalman filtering," *EURASIP J Wireless Commun. and Networking*, pp. 1-11, 2012.

[11] J. Xiang and W. W. Tan, "An improved DV-hop algorithm based on iterative computation for wireless sensor network localization," *2013 IEEE International Workshop on Electromagnetics, Applications and Student Innovation Competition*, pp.171-174, 1-3 Aug. 2013.

[12] S. Ji, K. F. Sze, Z. Zhou, A. M. C. So and Y. Ye, "Beyond convex relaxation: A polynomial-time non-convex optimization approach to network localization", *2013 Proceedings IEEE INFOCOM*, Turin, pp. 2499-2507, 2013.

[13] G. Oliva, S. Panzieri, F. Pascucci and R. Setola, "Sensor Networks Localization: Extending Trilateration via Shadow Edges," in *IEEE Transactions on Automatic Control*, vol.60, no.10, pp.2752-2755, Oct. 2015.

[14] Nguyen C, Georgiou O, et al. Maximum likelihood based multihop localization in wireless sensor networks. In: Communications (ICC), 2015 IEEE International Conference on. IEEE; 2015. p. 6663-6668.

[15] J. Aspnes, T. Eren, D.K. Goldenberg, A.S. Morse, W. Whiteley, Y.R. Yang, B.D.O. Anderson and P.N. Belhumeur, "A Theory of Network Localization," in *IEEE Transactions on Mobile Computing*, vol.5, no.12, pp.1663-1678, Dec. 2006.

[16] S. V. Pemmaraju and I. A. Pirwani, "Good quality virtual realization of unit disk graphs," *Springer Berlin Heidelberg*, pp. 311-322, 2007.

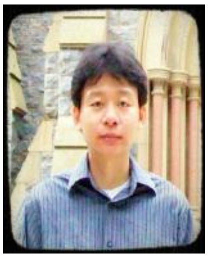
[17] M. Campbell, A. J. Hoane , F. Hsu, "Deep Blue," *Artifical Intelligence*, Vol. 134, Issues 1, pp. 57-83, 2002.

[18] Kaewprapha, P., Puttarak, N., Tansarn, T., "Multi-hop network localization in unit disk graph model under noisy measurement using tree-search algorithm with graph-properties-assist traversing selection," KKU-IENC, Proceeding of, Khon Kaen, Thailand, August 2016.

[19] G.H. Hardy, *Ramanujan. Twelve Lectures on Subjects Suggested by His Life and Work*, 3rd ed., Chelsea, New York, 1999, pp.67.

[20] Santaló, L. A., "Integral geometry and geometric probability. Encyclopedia of Mathematics and its Applications," Vol. 1. Addison-Wesley Publishing Co., Reading, Mass.-London-Amsterdam, 1976.

[21] R. B. Ellis, J. L. Martin, and C. H. Yan, "Random geometric graph diameter in the unit ball," *Algorithmica* , pp. 421-438, 2007.



Phisan Kaewprapha received his Ph.D. in Electrical Engineering, Lehigh University (USA), in 2012. He is a lecturer at Department of Electrical and Computer Engineering, Faculty of Engineering, Thammasat University Thailand. His area of interest includes computer and data communication networks, network localization, wireless networking, parallel and distributed computing, big data and machine learning.

Thaewa Tansarn received his bachelor of engineering degree in Electronics and Telecommunications Engineering from Rajamangala university of technology and currently a master degree student at Department of Electrical and Computer Engineering, Faculty of Engineering, Thammasat University, Thailand. He also work as a software engineer at Toyota Tsusho Electronics Thailand. His research interests includes network localization, microcontroller and embededed systems.

Nattakan Puttarak received her Ph.D in Electrical Engineering, Lehigh University (USA), in 2011, for her work in coding for data storage. She is currently a lecturer at the Telecommunications Engineering department, faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang (KMITL), Thailand. Her research interests fall in coding theory in the area of data storage, including both the mainstream systems of hard drives and the emerging technology of flash drives. This includes designing new error correction coding strategies and decoding algorithms to combat disk failure and recover lost data for small-scale disk arrays as well as large-scale data centers, analyzing their performances, and identifying best practices.

RCA on FPGAs Designed by the RTL Design Methodology and Wave-Pipelined Operation

Tomoaki Sato[1], Member, **Sorawat Chivapreecha**[2],
Phichet Moungnoul[3], and **Kohji Higuchi**[4], Non-members

ABSTRACT

Field-programmable gate arrays (FPGAs) are used in various systems that use reconfigurable function. Conventional FPGAs have been developed by a transistor-level description for minimizing routing delay. Although FPGAs developed by the register transfer level (RTL) design methodology provide various benefits to the designers of a system-on-a-chip (SoC), they have not been realized. Therefore, the authors have advanced their development. They should be shown to operate in a practical throughput. For this purpose, circuits on them need to be designed and evaluated. In this paper, a ripple-carry adder (RCA) is designed on them and the throughput of the RCA is evaluated. The throughput shows that it is applicable to network processors. In addition, a wave-pipelined operation without changing the RCA reveals that the problem of routing delay in the FPGAs developed by the RTL methodology is mitigated. The contributions of this paper are to clarify that a 4-bit adder can be implemented on the FPGAs and the throughput of it can be improved by wave-pipelined operations.

Keywords: FPGAs, RTL Design Methodology, RCA, Wave-pipeline, SoC

1. INTRODUCTION

A variety of equipment, such as network equipment [1] and firewalls [2] are running by using FPGAs because they use the reconfigurable features of the FPGAs. These features are to enable changes in circuit configurations on the FPGAs as in a software program when it is required to add or modify a function. Therefore, if the processing speed, power consumption or cost is not suitable in the processing of a central processing unit (CPU), the choice of FPGAs is useful.

In order to achieve the reconfigurable features in conventional FPGAs, a routing path is controlled by a transistor as a switch. This is different significantly from a SoC developed by using a standard cell library. Therefore, such a circuit configuration of the SoC cannot be changed. On the other hand, operations of FPGAs consume more power and are a lower frequency than that of the SoC.

To solve these problems, [3] and [4] achieved high-speed and low power operations in the architecture and transistor levels respectively. In [5], the problems of operating speed and power consumption of static random access memory (SRAM) used with FPGAs were described and solved. These studies are very profitable in conventional FPGAs developed by the transistor level.

By the way, the development of a SoC that builds in FPGAs is demanded. It is essential to embed the FPGA with the RTL level as a reason for large-scale SoC and shortening of a design period simultaneously. However, routing path delays of FPGAs designed by the RTL design methodology are larger than that of conventional FPGAs. This is a reason why a study on FPGAs in the RTL design methodology was not made.

To realize FPGAs designed by the RTL design methodology, the authors have developed [6-9]. The advantage is that the FPGAs themselves can be developed by using a HDL (hardware description language). It does not mean a circuit on the FPGAs. In addition, circuits on the FPGAs can be designed by the HDL as well as conventional FPGAs.

These FPGAs should be made clear that it is usable in a practical throughput. Thus, it is necessary to design and evaluate circuits on the FPGAs. In this paper, a 4-bit ripple-carry adder (RCA) is configured on them. The throughput of the 4-bit RCA is evaluated. In addition, wave-pipelined operations are made without changing a circuit configuration of the 4-bit RCA and indicates that contribute to easing the problem of routing delay.

This paper is organized as follows. Section 2 presents the outlines of CPUs and FPGAs in packet processing. Next, Section 3 explains the FPGAs which are designed by the RTL design methodology. In Section 4, the 4-bit RCA circuit is designed on the FPGAs and wave-pipelined operation of it is described in Section 5. Then, the 4-bit RCA is evalu-

Manuscript received on August 31, 2016 ; revised on December 23, 2016.
Final manuscript received on February 21, 2017.
[1] The author is with the Computing and Networking Center, Hirosaki University, Japan, E-mail: t-sato@hokusei.ac.jp
[2,3] The authors are with the Faculty of Engineering, KMITL, Thailand, E-mail: sorawat@telecom.kmitl.ac.th and phichet@telecom.kmitl.ac.th
[4] The author is with the Graduate School of Informatics and Engineering, UEC, Japan, E-mail: higuchi@ee.uec.ac.jp

Table 1: *CPU Time in the CPUs of [16].*

Frequency of the CPU	Minimum size frame (ns)	Standard size frame (ns)	Jumbo frame of 3000 KB (ns)	Maximum jumbo frame (ns)
500MHz	760.0	15280.0	30280.0	160080.0
800MHz	380.0	7640.0	15140.0	80040.0
1GHz	237.5	4775.0	9462.5	50025.0
1.2GHz	190.0	3820.0	7570.0	40020.0
1.5GHz	158.3	3183.3	6308.3	33350.0
2.0GHz	126.7	2546.7	5046.7	26680.0
3.0GHz	95.0	1910.0	3785.0	20010.0
4.0GHz	63.3	1273.3	2523.3	13340.0

ated in Section 6. In Section 7, the conclusions are made.

2. NEEDS AND PROBLEMS OF FPGAS FOR PACKET PROCESSING IN MOBILE DEVICES

Packet processing in a mobile device should be executed at high-speed and low-power consumption. The use of an FPGA is the better solution in these aspects. However, conventional FPGA devices have a problem that a CPU developed by an FPGA designer cannot be put it on as an ASIC. In this section, the outlines of CPUs and FPGAs in packet processing are explained. Subsequently, problems and solutions of packet processing on a mobile CPU are clarified based on our research results. Finally, the necessity of a CPU architecture that specializes in packet processing and an FPGA developed by the RTL design methodology is described.

High throughput of networks used in mobile devices has been advanced. In Wi-Fi LANs, the development of IEEE802.11ad is conducted. It will finally achieve the throughput of 6.8 Gbps [10-11]. Moreover, studies on 5G is advanced in mobile data communications. The purpose of the studies is to achieve the data communication of more than 10 Gbps [12]. Packet processing in mobile devices in such a high-speed data communication is essential to process using an ASIC or FPGA.

Super pipelining, parallel processing and processing that does not depend on the word size are easily achieved by using the FPGA. Therefore, the FPGA is very beneficial for packet processing. In [13], a packet classification engine had been achieved a very high throughput by super pipelining on the FPGA. Moreover, the use of an FPGA indicated that it is also beneficial for low power consumption in [14]. [15] used re-configure features in the FPGAs for packet processing.

However, detecting processing on unauthorized accesses needs not only the packet classification engine but also complex processing. Circuits on an ASIC or FPGA for complex processing make a problem that amount of hardware is huge. In addition, it cannot take advantage of software resources on unauthorized accesses. That is, a CPU for packet and detecting processing is required.

The authors have estimated performances of a CPU which is needed for packet processing [16]. According to the results, packet processing in 1 Gbps throughput and normal data communications needs the CPU of MIPS64 5K architecture with 1 GHz operations. In contrast, packet processing of repeating a packet frame of the shortest size and a packet frame of the standard size needs a clock frequency of 4GHz. Table 1 shows the CPU times required for packet transfer processing. If it is unable to process, it becomes a state in a denial of service (DoS) attack.

In [16], in order to protect a CPU with 1 GHz operations from DoS attack, out-of-order packets execution inside the CPU has been proposed. The function should be built inside the CPU as an ASIC circuit. In addition, the CPU needs to be developed as an ASIC for low-power operations because the processing speed and power consumption of the CPU influences the entire system. If the CPU has sufficient processing capacity for packet processing, it facilitates the change in the content of processing by software. In that case, re-configurable features such as an FPGA are not required.

When incorporated such a unique architecture on an FPGA, the CPU was only on the FPGA as a soft-core processor. Furthermore, circuits as an ASIC are often needed from the viewpoint of low-power and high-speed processing. Therefore, the FPGA the RTL design methodology as in this study facilitates the development of the CPU and circuits as an ASIC. In other words, ASIC-FPGA co-design is realized.

3. FPGAS DESIGNED BY THE RTL DESIGN METHODOLOGY

In this section, details of the FPGAs that the authors have developed are explained. Next, logic synthesis is executed to investigate delay times of the FPGAs. Develop environments for them are described.

The architecture of the FPGAs is shown in Fig. 1. The FPGAs are composed of parts of three types. The parts are a logic block (LB) of Fig. 2, a connection block (CB) of Fig. 3, and a switch block (SB) of Fig. 4. Parts greatly different from conventional FPGAS are the CB and the SB. These switches are not a transistor but a selector. This is the cause, the direction of the routing of the FPGAs is different from that of conventional FPGAs.

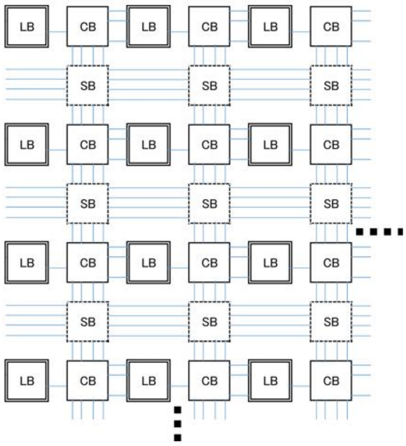


Fig.1: *Architecture of the FPGAs designed by the RTL design methodology.*

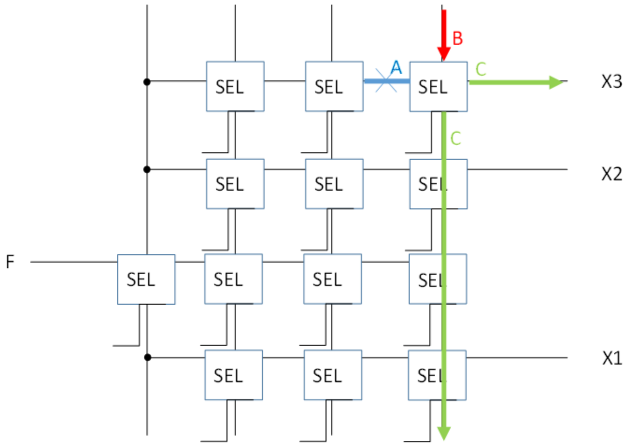
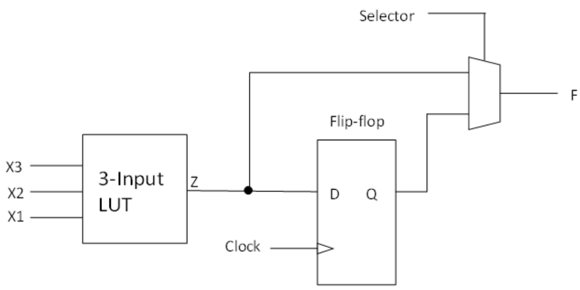
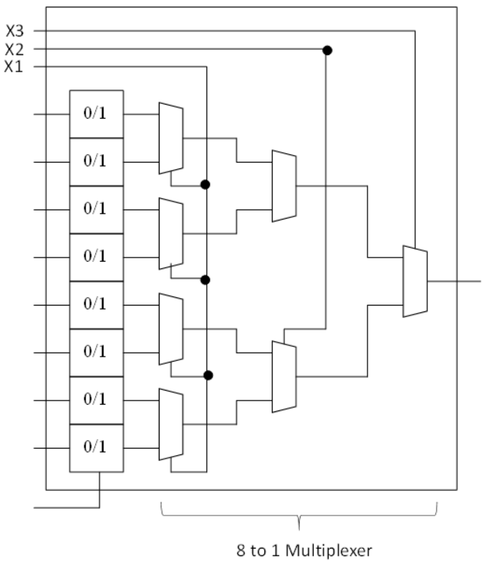


Fig.3: *Structure of the CB.*



(a)



(b)

Fig.2: *Structure of the LB. (a) Basic component (b) 3-input and 1-output LUT.*

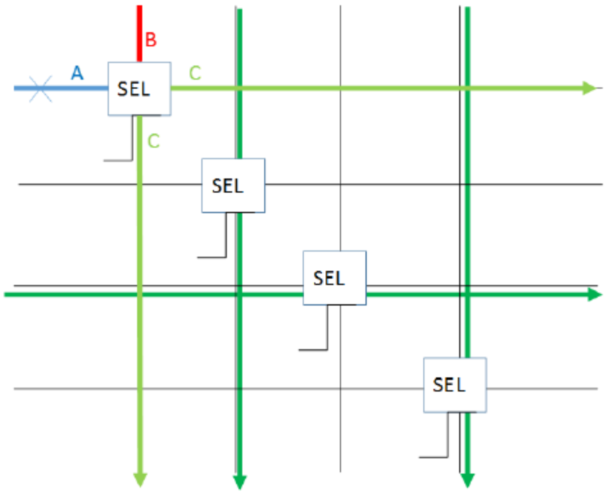


Fig.4: *Structure of the SB.*

Table 2: *Design environments.*

OS	Cent OS 5.9 x86
CPU	Intel Core 2 Duo E6600 (2.4GHz)
Memory	2 GBytes
Logic synthesis	Synopsys Design Compiler H-2013.03-SP2
Technology	Rohm 180 nm C-MOS
Standard cell library	The library provided by Rohm

As indicated in the selectors of Fig. 3 and Fig. 4, the selection result of A or B is outputted to C. In case of Fig. 3 and Fig. 4, the signal of B is selected and flows to C. The signal A cannot be used. Fig. 4 shows that 4 lines can cross at the same time. Therefore, it is confirmed that routings on the FPGA are possible [9].

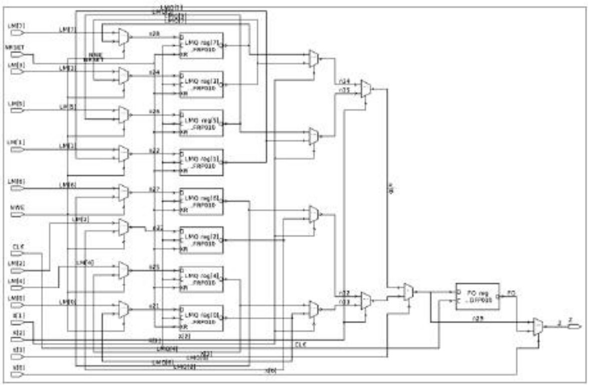


Fig.5: *Logic synthesis result of Fig. 2.*

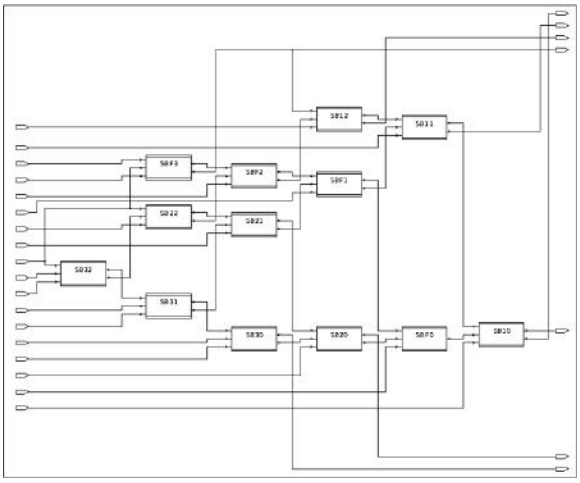


Fig.6: *Logic synthesis result of Fig. 3.*

The advantage of the FPGAs is that the FPGAs themselves can be developed by the RTL design methodology. It allows not only ASIC-FPGA co-design but also to arrange the FPGA architecture. If a large circuit is designed on the FPGAs, routing may not be possible with this FPGA configuration. In this case, increasing the number of wires solves the problem on routing. That is, it is easy to realize other arithmetic circuits.

For the evaluation carried out in Section 6, the authors run the logic synthesis of the parts of the FPGAs by using the design environments of Table 2. In this paper, the authors choose the standard cell library released by Rohm Inc. The CB of this paper is different from that of [7]. The number of selectors has been optimized to it. Logic synthesis results of Fig. 2, Fig.3 and Fig. 4 are shown in Fig. 5, Fig. 6 and Fig. 7, respectively.

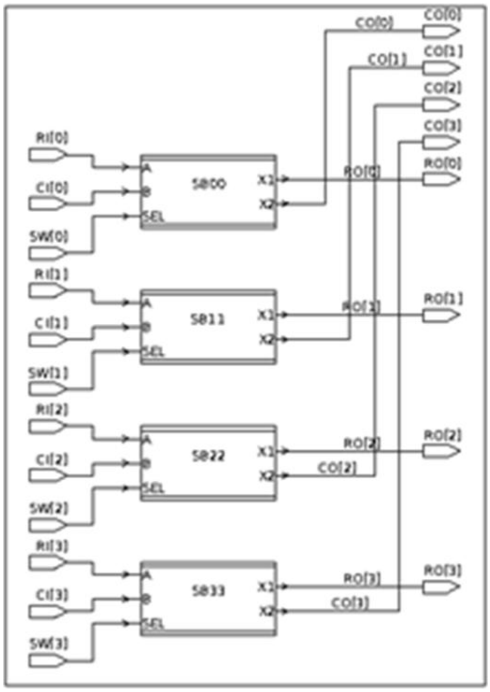


Fig.7: *Logic synthesis result of Fig. 4.*

CIN	A[0]	B[0]	S[0]
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	0
1	1	1	1

(a)

CIN	A[0]	B[0]	CIN[1]
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1

(b)

Fig.8: *Truth table for the full adder in the LUTs. (a) Sum (b) Carry.*

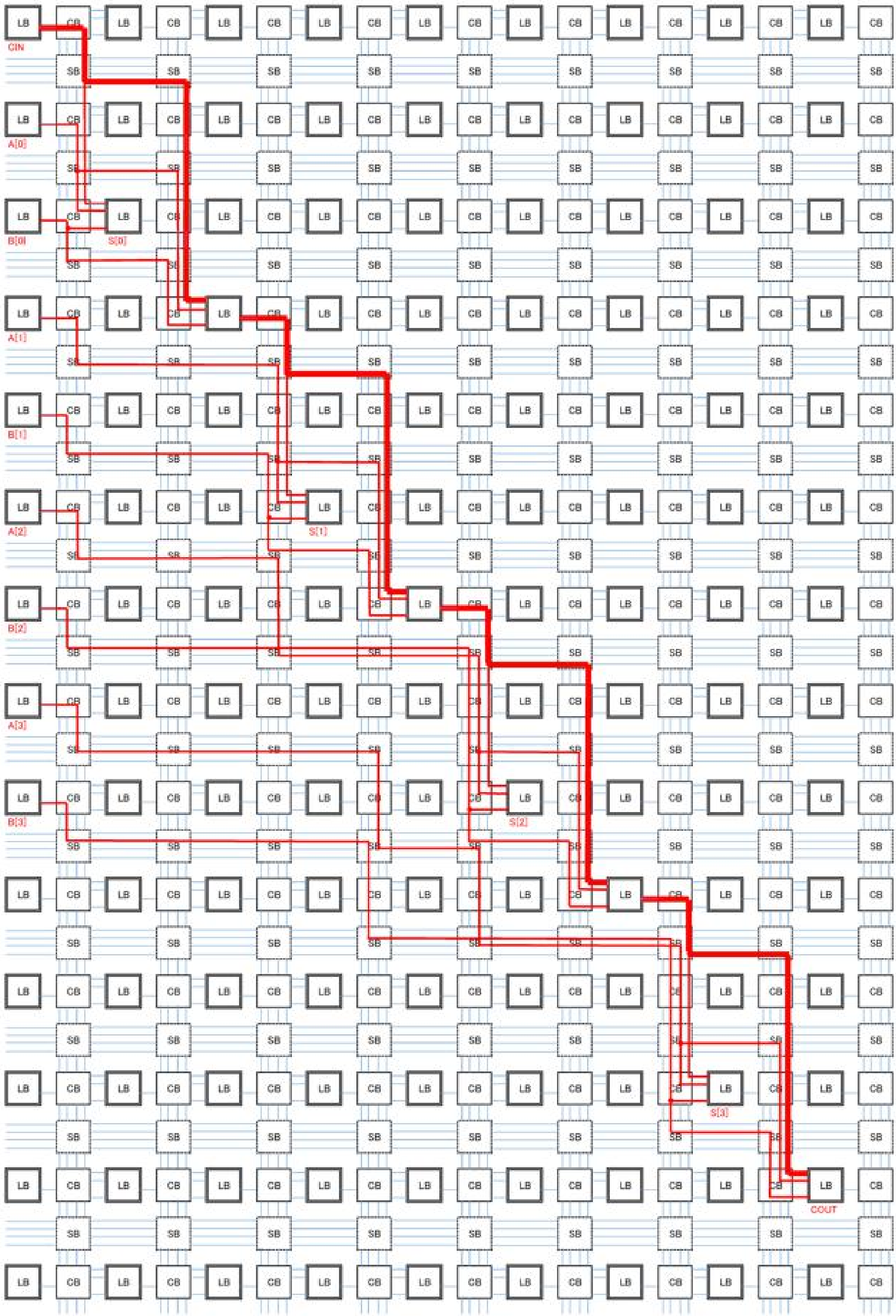


Fig.9: *4-bit RCA on the FPGAs.*

4. RCA ON THE FPGAS

The FPGAs are in order to clarify that they are working at a practical rate. A 4-bit RCA circuit is developed on the FPGAs in this section. The LUT of the FPGAs is 3-input and 1-output and can store a table shown in Fig. 8. Therefore, a full adder (FA) is designed with two LBs.

The 4-bit RCA circuit is shown in Fig. 9. Routing is made based on the explanation in Section 3. The operation frequency is calculated by using the results of the logical synthesis. For the calculation of the delay times, the authors make the software program. According to the results, the routing of the maximum delay time is from CIN to COUT shown in the heavy line of Fig. 9 and the maximum delay time is 24.12 ns.

5. WAVE-PIPELINED OPERATION

A wave-pipeline [17], [18] is a design method without using pipeline registers for pipeline operations. The method is superior in terms of power consumption because it does not use a register. In circuits on FPGAs, a design technique for high-speed operations is limited. In such a situation, the wave-pipeline is effective in the circuits on FPGAs [19].

Fig. 10 shows the overviews of pipelines. Conventional pipelines shown in Fig. 10 (a) need pipeline registers for pipelined operations. Only one set of signals can be operated in the circuit between the pipeline registers. On the other hand, pipeline registers are not used for wave-pipelines. Therefore, it is essential to make a collision-free interval so that the first signal does not collide with the next signal. Wave-pipelined operations are confirmed that two or more signals exist between registers.

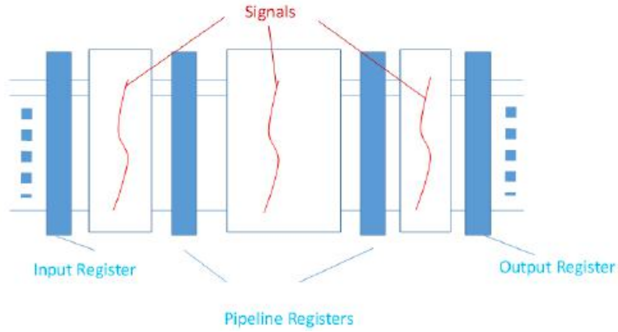
Wave-pipeline is also used in commercial processors. In our study, circuits constituting FPGAs for wave-pipelines have been studied [6-9, 21]. However, it is not achieved in the arithmetic circuit on the FPGA constructed by RTL proposed by us. However, they have not been achieved in an arithmetic circuit on the FPGAs proposed by us.

A clock cycle time for wave-pipelining, T_{CK} is calculated from the following equation [20]. Wave pipelined operations are achieved if this expression is satisfied.

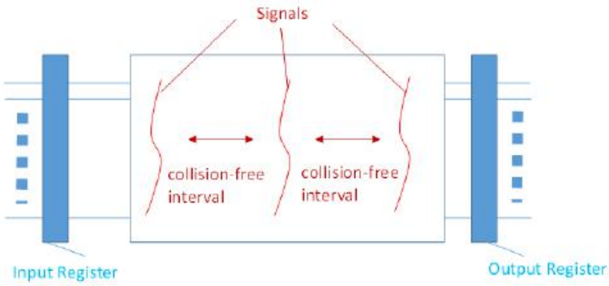
$$T_{CK} = (D_{MAX} - D_{MIN}) + T_{OV} \qquad (1)$$

In a circuit for wave-pipelined operations, D_{MAX} is the maximum delay time and D_{MIN} is the minimum delay time. T_{OV} is set as margins. The margins mean the influence of the condition of chip fabrication and the operating conditions from the temperature and voltage.

The novelty of the wave-pipeline in this paper is to allow wave-pipelined operations without changing the circuit configuration of Fig. 9. Here, the route



(a)



(b)

Fig.10: *Overviews of pipelines. (a) Conventional pipeline (b) Wave-pipeline.*

adjustments of only the outputs shown in the heavy line of Fig. 11 are executed. This is led to the design simplification of a wave-pipelined circuit. According to the Sec. 4, the maximum delay time is 24.12 ns. Also, the minimum delay route on a route cannot be adjusted is from B[3] to COUT. The minimum delay time is 10.07 ns. From these conditions, the routes of the outputs of S[0], S[1], S[2] and S[3] are derived from the following equation. D_{OUTPUT} is a delay time for the outputs.

$$10.07 \leq D_{OUTPUT} \leq 24.12 \qquad (2)$$

All the outputs of Fig. 9 satisfy Eq. (2).

The FPGAs enable ASIC-FPGA co design. Therefore, arithmetic circuits as an ASIC make to solve the problems of operation speed on the FPGAs. However, circuits as an ASIC cannot be changed and added after chip fabrication. Wave-pipelined operations on the FPGAs are needed for this reason.

6. EVALUATIONS

The FPGAs designed by the RTL design methodology is evaluated in this section using operation frequencies of Fig. 9 and Fig. 11, respectively. An operating frequency in normal operations of the RCA in Fig. 9 can be obtained from the maximum delay time. On the other hand, an operating frequency in wave-pipelined operations of the RCA in Fig. 11 is derived from Eq. (1). T_{OV}, the overhead time in the wave-pipelined operations is set at 2.0 ns. Actually,

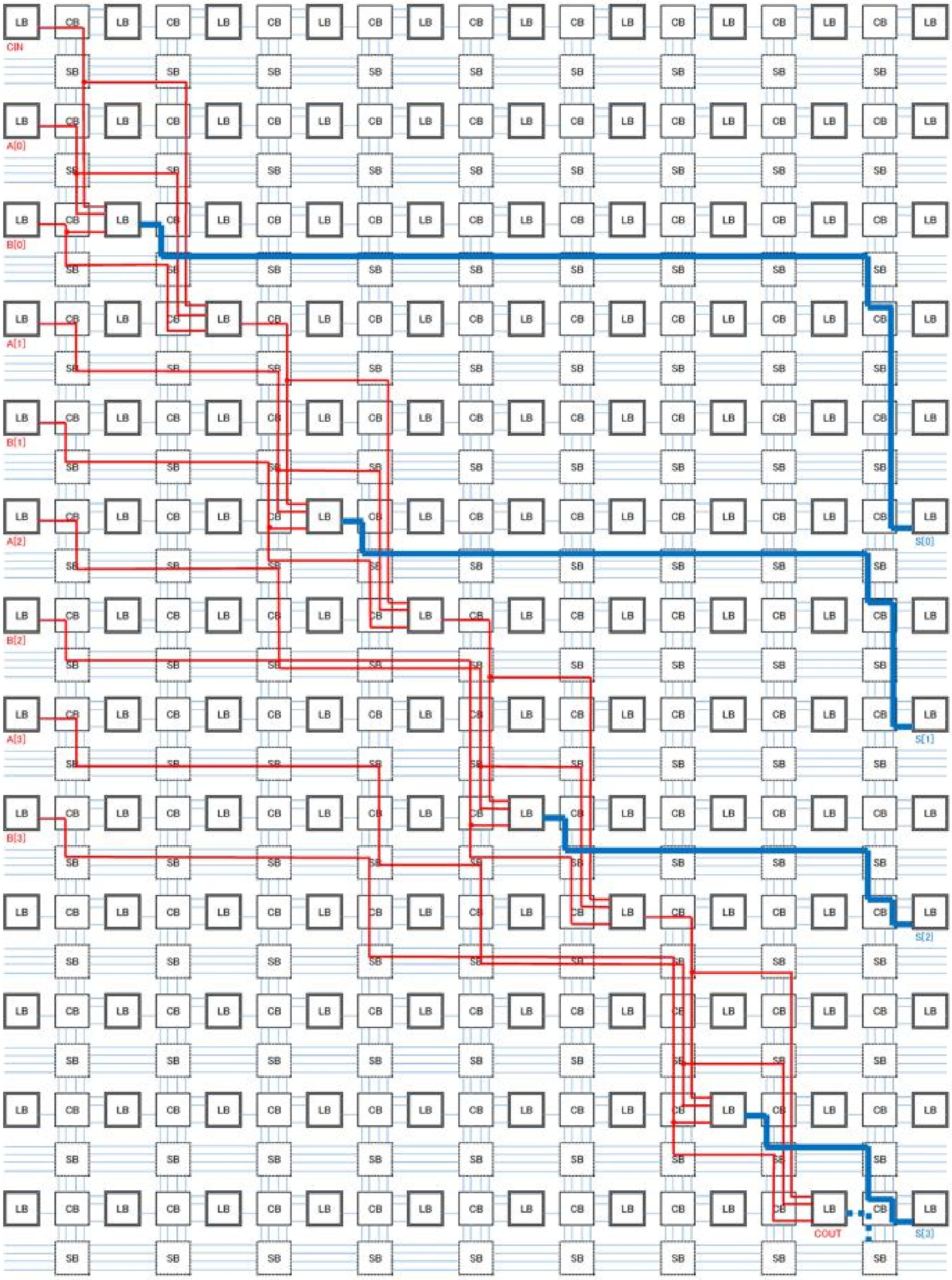


Fig.11: *4-bit RCA for wave-pipelined operations on the FPGAs.*

circuits on an ASIC fabricated in a 0.18 μm CMOS process operate 2.0 GHz [22]. Therefore, this value is very reasonable.

The clock cycle time of Fig. 11 in wave-pipelined operations, T_{RCA} is calculated from the following equation.

$$T_{RCA} = (24.12 - 10.07) + 2.0 \quad (3)$$

These results are shown in Fig. 12.

The operation frequency greatly depends on the process technology. MAX II of Altera's complex programmable logic device (CPLD) has been implemented in a 180 nm C-MOS technology [23]. The technology is same as the FPGAs. Operation frequencies of the Internal Oscillator of the CPLD are 13.33-22.22 MHz. That is, the operation frequency of the FPGA is more high-speed than that of the CPLD.

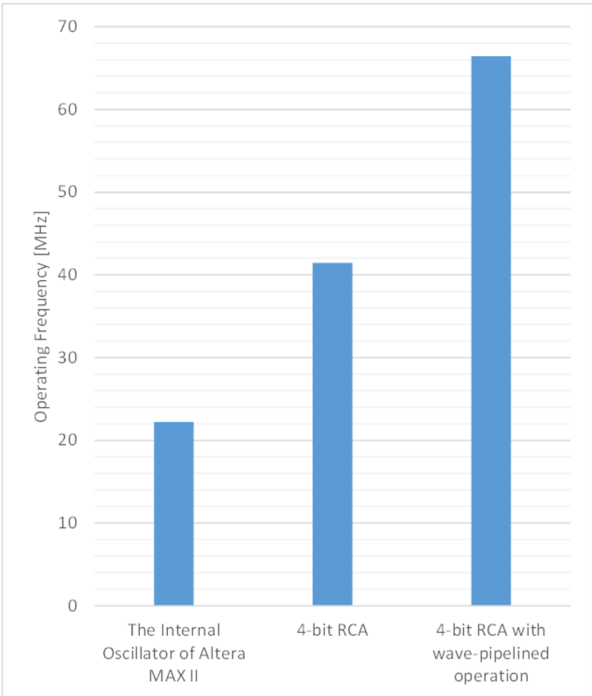


Fig.12: *Operating frequencies of the CPLD and 4-bit RCAs on the FPGAs.*

When packets processing of a computer network is executed on FPGAs, processing in the packet frame units is possible. Here, the operating frequency in the FPGAs is set to 60MHz. In the case of 1 Gbps, the word width of 17 bits or more enables the processing. Thus, it is clear that it is possible to sufficiently practical processing.

7. CONCLUSIONS

The FPGAs designed by the RTL design methodology have the following advantages:

- Easy integration of FPGA functions in a SoC is possible.
- Significant shortening of the design period of a SoC.
- Allows the selection of process rules.

The authors have developed the FPGAs in order to be able to capitalize on these advantages. In this paper, the 4-bit RCA on the FPGAs was designed and evaluated in order to demonstrate the practicality of the FPGAs.

As a reason why the FPGAs were not developed, there was a larger delay than conventional FPGAs. This problem was relaxed by wave-pipelined operations without changing the circuit configuration of RCAs. Wave-pipelined operations are very suitable for patterned circuits like the RCAs. That is, they are considered to be applicable in multiplication circuits.

Wave-pipelined circuits have the advantage of not increasing power consumption because they are not required pipeline registers. In addition the delay time of the entire circuit in wave-pipelined operations is the same as that in normal operations.

Therefore, the contributions of this paper are as follows:

- It was shown that the FPGAs can be put to practical use.
- A 4-bit adder can be implemented on the FPGAs.
- The problem of routing delay can be solved by improving the throughput by easy wave-pipelined operations.

Future works are fabricating the FPGAs chips using the 180 nm CMOS standard cell library and evaluations by the measurement of the chip.

ACKNOWLEDGEMENT

This work has been supported in part by VLSI Design and Education Center (VDEC), the University of Tokyo in collaboration with Synopsys, Inc., Rohm Corporation and Toppan Printing Corporation and KAKENHI Grant Numbers 25330149.

References

[1] Y. R. Qu and V. K. Prasanna, "High-Performance and Dynamically Updatable Packet Classification Engine on FPGA," *IEEE Trans. Parallel Distrib. Sys.*, vol. 27, no. 1, pp. 197-209. 2016.

[2] T. Sato, S. Imaruoka and M. Fukasse, "Verifying Firewall Circuits by Wave-Pipelined Operations," in *Proc. IEEE TENCON 2009*, pp. WED3.P.14.1-WED3.P.14.6, 2009.

[3] S. Redif and S. Kasap, "Novel Reconfigurable Hardware Architecture for Polynomial Matrix Multiplications," *IEEE Trans. VLSI*, vol. 23, no. 3, pp. 254-265, 2015.

[4] H. Marzouqi, M. Al-Qutayri, K. Salah, D. Schinianakis, and T. Stouraitis, "A High-Speed FPGA Implementation of an RSD-Based ECC

Processor," *IEEE Trans. VLSI*, vol. 24, no. 1, pp. 151-164, 2016.

[5] K. Huang, R. Zhao, W. He and Y. Lian, "High-Density and High-Reliability Nonvolatile Field-Programmable Gate Array With Stacked 1D2R RRAM Array," *IEEE Trans. VLSI*, vol. 24, no. 1, pp. 139-150, 2016.

[6] T. Sato, S. Chivapreecha and P. Moungnou, "A Crossbar Switch Circuit Design for Reconfigurable Wave-Pipelined Circuits," in *Proc. WMSCI 2014*, vol. II, pp. 200-205, 2014.

[7] T. Sato, S. Chivapreecha and P. Moungnou, "Wiring Control by RTL Design for Reconfigurable Wave-Pipelined Circuits," in *Proc. APSIPA ASC 2014*, pp. WP1-3-1-WP1-3-6, 2014.

[8] T. Sato, S. Chivapreecha and P. Moungnou, "Fine-Tuning of Wave-Pipelines on FPGAs Developed by the RTL Design," in *Proc. ECTI-CON 2015*, pp. 1230.1-1230.6, 2015.

[9] T. Sato, S. Chivapreecha and P. Moungnou, "The Potential of Routes Configured with the Switch Matrix by RTL," *Applied Mechanics and Materials J.*, vol. 781, pp. 189-192, 2015.

[10] W. Hong, K.-H. Baek and A. Goudelev, "Multilayer Antenna Package for IEEE 802.11ad Employing Ultralow-Cost FR4," *IEEE Trans. Antennas and Propagation*, vol. 60, no. 12, pp. 5932-5938, 2012.

[11] H. Sawada; S. Takahashi and S. Kato, "Disconnection Probability Improvement by Using Artificial Multi Reflectors for Millimeter-Wave Indoor Wireless Communications," *IEEE Trans. Antennas and Propagation*, vol. 61, no. 4, pp. 1868-1875, 2013.

[12] A. H. Fazlollahi and J. Chen, "Copper Makes 5G Wireless Access to Indoor Possible," in *Proc. 2015 IEEE Global Communications Conference (GLOBECOM)*, pp. 1-5, 2015.

[13] Y. R. Qu and V. K. Prasanna, "High-Performance and Dynamically Updatable Packet Classification Engine on FPGA," *IEEE Trans. Parallel and Distributed Systems*, vol. 27, no. 1, pp. 197-209, 2016.

[14] A. Kennedy and X. Wang, "Ultra-High Throughput Low-Power Packet Classification," *IEEE Trans. Very Large Scale Integration (VLSI) Systems*, vol. 22, no. 2, pp. 286-299, 2016.

[15] G. Brebner and W. Jiang, "High-Speed Packet Processing using Reconfigurable Computing," *IEEE Micro*, vol. 34, no. 1, pp. 8-18, 2014.

[16] T. Sato, P. Moungnoul, S. Chivapreecha and K. Higuchi, "Performance Estimates of an Embedded CPU for High-Speed Packet Processing," in *Proc. ECTI-CON 2014*, pp.1298.1-1298-5, 2014.

[17] L. Cotton, "Maximum Rate Pipelining Systems," in *Proc. AFIPS Spring Joint Computer Conference*, pp. 581-586, 1969.

[18] F. Klass and M. J. Flynn, "Comparative Studies of Pipelined Circuits," *Stanford University Technical Report*, no. CSL-TR-93-579, 1993.

[19] I. B. Eduardo, L. Sergio and M. M. Juan, "Some Experiments About Wave Pipelining on FPGA's," *IEEE Trans. Very Large Scale Integration (VLSI) Systems*, vol. 6, no. 2, pp. 232-237, 1998.

[20] W. P. Burleson, M. Ciesielski, F. Klass, and W. Liu, "Wave-Pipelining: A Tutorial and Research Survey," *IEEE Trans. Very Large Scale Integration (VLSI) Systems*, vol. 6, no. 3, pp. 464-474, 1998.

[21] T. Sato, P. Moungnoul, S. Chivapreecha and K. Higuchi, "A Connection Block Implemented in the RTL Design for Delay Time Equalization of Wave-Pipelining, *J. Systemics, Cybernetics and Informatics*, vol. 14, no. 1, pp. 49-54, 2016.

[22] C. L. Jin, X. P. Yu and W.-Q. Sui, "1-2 GHz 2 mW injection-locked ring oscillator based phase shifter in 0.18 m CMOS technology," *Electronics Letters*, vol. 52, no. 22, pp. 1858 - 1860, 2016.

[23] Altera Inc., "Using the Internal Oscillator in Altera MAX Series," https://www.altera.com/en_US/pdfs/literature /an/an496.pdf, 2014.

Tomoaki Sato received the B.S. and M.S. degrees from Hirosaki University, Japan, in 1996 and 1998 respectively, and the Ph.D. degree from Tohoku University, Japan, in 2001. From 2001 to 2005, he was an Assistant Professor of Sapporo Gakuin University, Japan. Since 2005, he has been an Associate Professor of Hirosaki University. Between 2012 and 2015, he was a Visiting Associate Professor of the Open University of Japan. From April 2017, he will become a Professor of Hokusei Gakuen University, Sapporo, Japan. His research interests include VLSI Design, Computer Hardware, Computer Architecture and Computer and Network Security. He is also a member of ECTI, IEEE, IEICE, IPSJ, and SSI.

Sorawat Chivapreecha received his B.Eng. degree in telecommunication engineering from the Suranaree University of Technology (SUT), Nakhon Ratchasima, Thailand, in 1998, M.Eng. degree and D.Eng. degree in electrical engineering from the King Mongkut's Institute of Technology Ladkrabang (KMITL), Bangkok, Thailand, in 2001 and 2008, respectively. Since 2002, he has been a faculty member of the Department of Telecommunication Engineering, Faculty of Engineering, KMITL, where he is currently an Assistant Professor. In 2011, he was awarded the distinguished lecturer from Faculty of Engineering, KMITL. His research interests include digital filter design and implementation, VLSI for digital signal processing, information science and satellite engineering. He is also a member of IEEE and IEICE.

Phichet Moungnoul received the B.Ind.Tech. degree in telecommunication technology, M.Eng and D.Eng degree in Electrical Engineering from King Mongkut's Institute of Technology Ladkrabang (KMITL), Bangkok, THAILAND, in 1992 ,1997 and 2001, respectively. Since 1998, he has been a member of the Department of Telecommunication at Faculty of Engineering, KMITL, where he is currently an assistant professor of telecommunication. His research interests wireless and mobile communication and Broadband Communication.

Kohji Higuchi received his Ph.D. degree from Hokkaido University, Sapporo, Japan in 1981. In 1980 he joined the University of Electro-Communications, Tokyo, Japan, as a Research Associate, where he became an Assistant Professor in 1982 and currently an Associate Professor in the Dept. of Mechanical Engineering and Intelligent Systems, Electronic Control System Course. His interests include Power Electronics, Control Engineering and Digital Signal Processing.

The Improvement of The Efficiency for The Quality Assurance Information System Based on The Six-Sigma Principle

Thitirath Cheowsuwan[1], **Somporn Saipanya**[2], and **Supan Tongphet**[3], Non-members

ABSTRACT

This paper presents the development of the efficiency of a quality assurance information system based on the six-sigma principle for schools in northern Thailand. The aim of this research is to solve all problems occurred in the previous system such as the awkward user interface, no quality standard support, users' confusions in workflow process, the inaccurate and unreliable information system process etc. The six-sigma foundation is operated by five processes, which are the define phase, the measure phase, the analyze phase, the improve phase, and the control phase (DMAIC). According to the studies, the measurement of the users' satisfaction and expectation is divided into three parts: standard-based service, system reliability, and processing speed. As the results, the quality measurements from both technical and non-technical are ranged on 2-sigma level. These measurements confirm that the proposed system provides the improved services better than the previous system and it can solve all problems as mentioned above effectively. Moreover, it can raise the efficiency of the service system based on the quality assurance standard for the mission of the schools in the northern region.

Keywords: Six-Sigma Principle, DMAIC, Quality Assurance Information System, Sigma Level

1. INTRODUCTION

National Education Act B.E.2542 provides and considers quality assurance in schools, and it is one of the processes of educational administration, which was managed and developed continuously to a standard quality in education for supporting the external quality assurance [1]. However, many schools in the northern region still have the problems with the information criteria management and the quality assurance indicators, which cannot completely integrate the routine works with the quality assurance in education works. Due to the large amounts of data and many of the people in several segments, it is so difficult to manage all information because most of them have not been arranged in electronic documents. Consequently, it increases the teacher workloads to gather each information in the schools during the period of quality assurance in education causing the deviation of the information and it cannot reflect the real quality of the schools. Therefore, those problems will be solved by using the quality assurance information system, which is one of the databank development projects and business intelligence for quality assurance system. It uses for reducing the complicated tasks in information management for the people who are responsible for quality assurance in the schools of the northern part of Thailand [2]. The databank system can collect user's information from both the daily tasks and other sources. All information will be used for supporting the manager's decision in terms of decision support system [2]. The proposed software in this research, which is used for solving the quality assurance in educational issues, starts from the business process model development. It identifies a level of working functions connecting with the analysis system and leads to develop in the information system by explaining as a symbolic model, which is called business process modelling notation (BPMN) [3]. The system model is created by a sequence of unified modelling language (UML) [4]. Then, the information system is employed for testing the system efficiency by following the law of the Six Sigma standard in a way of quality in dimensions. It consists of a standard pattern, an appropriate management, and a responsibility of the organized missions. These allow both the customer and the manufacturer to get the benefits of the investment both resources and productive values [5]. The Six Sigma has five steps, which are the define phase, the measure phase, the analyze phase, the improve phase, and the control phase (DMAIC). Its quality development is the way either to reduce the flow or to rebuild the level of quality to achieve the level of Six Sigma by using the quality process and statistics method development. This research purposes all information based on the efficiency of the quality assurance information system used the Six Sigma to survey the satisfaction of the sample in schools. In addition, it is used for studying the reasons for user dissatisfaction on the services to solve its quality problems and seeking the way to

---

Manuscript received on September 9, 2016 ; revised on February 28, 2017.

Final manuscript received on March 6, 2017.

[1,2,3] The authors are with School of Information and Communication Technology, University of Phayao, Phayao 56000, Thailand, E-mail: thitirath.ch@up.ac.th, tanktai@hotmail.com and tp_supan@hotmail.com

control on working process more efficiently.

2. RESEARCH METHODOLOGY

The Six Sigma is the idea to improve the quality of the organization. It can reduce errors that occur in the process by using the principles of statistics. Many businesses have adopted the Six Sigma technique to improve the quality of their organization, for instance, the case study narrowly focuses on reduction/elimination of two imperative responses in spray painting process producing shock absorbers, namely peel off and blisters using the Six Sigma Define-Measure-Analyze-Improve-Control (DMAIC) approach that highly impacts quality at customer end [6]. The present article presented a Lean Six Sigma (LSS) project management process improvement model and a case study test developed in a real enterprise environment which has a formal and established project management system PMI based. The LSS proposed approach is a DMAIC cycle-based proposal [7].The five stages of six sigma improvement model DMAIC (define, measure, analyse, improve, control) is the core Six sigma management system based on JMP/SAP system. In this system, operators can acquire batch data of production in different stages through the top of the SAP system as required, and then make connection between JMP and SAP system, finally put forward an efficient improvement plan after analysing on production data by using JMP [8].

As mentioned above, the solution process of the Six Sigma standard has 5 phases known as DMAIC. Both the define phase and the measure phase will be improvably focused in this research.

1) The studying of the Define phase consists of the Quality Assurance Information Systems (QAIS), system results, and service problems. Firstly, the information from all studying is formed as a Macro Process Mapping. Secondly, both analysis and mapping of the process (or SIPOC analysis) are used for seeking the user's need and expectation (or SIPOC and Requirement Analysis) to create a Micro Process Mapping, which is shown the minor process of the system [9]. Thirdly, all information from those processes is used to build the need of the system users. Finally, the summarization is employed to make a project indicator for this research (or Metric) leading to solve all problems in the Business Metric, which is in both user level (or Project Metric) and topic level (or Theme). As the results, the proposed technique can synthesise the important factors affecting to both service quality and user's satisfaction. Furthermore, the questionnaire is randomly corrected from all users by using non-probability sampling [10]. Due to users' sample is clear, the purposive sampling is employed in the 26-schools sample divided into 4 categories. Each category requires for 2 samples per school from executives, teachers, quality assurance staffs, and students (totally 208 samples).

2) Measure phase is to measure an important aspect of the project and to analyze the quality process by using the tree diagram to find and to choose the indicators [9]. First, the criteria weighting technique is used for creating an indicator account. Second, a check sheet is employed for collecting information in the QAIS testing for 240 days [11]. Third, from the information in a check sheet, it is manipulated to seek all flaws and mistakes in the services. Fourth, the analysed information from the processes based on baseline sigma is used for finding the system mistakes, which is randomly compared with one million samples (or Defect per Million Opportunity). The founded mistakes are analysed based on
DPMO = Defect Counted ÷ (Unit Counted × Defect Opportunity) × 106 solutions

Fifth, the quality level (or Sigma Level) is calculated from the number of both the receiving problems and the unsolvable problems. In the case of counted information, it can be calculated from the bad proportion, $\bar{P}$[12], as

$$\bar{P} = \frac{\Sigma np}{\Sigma p}, \tag{1}$$

where Σnp is total number of defect products overall, and Σp is total number of determine products overall.

The Z benchmark technique is used for calculating the level of quality (or Sigma level) to estimate the counting performance searching all information at the Sigma level; it is named as the long terms indicator, Pp. Finally, the counting information is compared with the divided scale (or the standard scale), Z, by calculating the bad proportion (lookup the defect per opportunity table). The Pp indicator is used for evaluating the efficiency of the process as

$$Pp = \frac{1}{3} Z_{Bench}. \tag{2}$$

3) Analyze phase is the analysis of the data collected from the research tasks such as frequency, variance, the use of the Pareto diagram, system processes, and the hypotheses of the possible causes shown in the tree diagram. Finally, it also proofs the hypotheses and concludes the root cause statement summary.

4) Improve phase is the process of searching the alternative ways to rectify system. The main purpose of this section is to prevent all defects and to satisfy all users in the system. Also, it is utilized for preparing to the real situation by using the proposed process, analysed for the Sigma level to show the quality level of the modified process.

5) Control phase is the process of bringing the improved and tested scheme to use in the real practice by specifying an indicator and a control point. Furthermore, many aspects are created in this section such as writing user's manual, monitoring all processes, both tracking and fixing all problems, and im-

proving system continually.

3. RESEARCH RESULTS

The results in the Define phase show that the QAIS in schools of the northern region in Thailand is designed to compatible with online systems. So, all users are able to access the proposed system anywhere and anytime through .NET technology, Windows Server, SQL Server, HTML, and JavaScript. The main function of the system is composed of nine sections: (1) The management of the schools. (2) The management of criteria and the indicator of the quality assurance in education. (3) The record of the overall operations and the indicator of the quality assurance. (4) The information management and the quality assurance indicator criteria evidence. (5) The quality assurance in education information analysis. (6) The report evaluation management. (7) The committee evaluation management. (8) To create activities for self-evaluation. (9) The store of Educational Quality Assurance (EQA) information. Moreover, the proposed system can report the indicators, which are based on the Office for National Education Standards and Quality Assessment standard, for all users such as executives, instructors, administrative teachers, students, and administrators. Some reports of the QAIS in schools of the northern part of Thailand can be shown in figure 1.

There is a procedure of searching for measuring the users' needs and expectations in the system such as (D1) To specify the users' problems by answering 5 questions (What? Where? When? Why? How?) (D2) To bring all answers from (D1) to draw a Macro Process Mapping, SIPOC model, and Micro Process Mapping (D3) To study basic information of system by studying sources, developed systems, functions of procedural system, results, connections, and securities (D4) During the system testing, as shown in figure 2, the studying of the secondary information is analyzed for searching users' issues such as problems, complaints, comments, and satisfies.

The solutions of the Six Sigma method can be explained by receiving all users' complaints during the system testing for 244 days. From the collection of information from the sample group of the trial system, we found 60 complaints as following: 18 complaints about system speed, 12 complaints about system reliability, 11 complaints about accessibility and usage, 8 complaints about system accuracy, 6 complaints about system security, and 5 complaints about system administrator service. In the testing works of the system found that 60 complaints can be solved within 3 hours.

In the specification of the Metric is used for solving the practical problems: (1) Business Metric is to satisfy all users with the QAIS (2) In Project Metric, we found 2 users' problems: First, the technical defection is appeared by the users' confusions such as the workflow system, data processing, and creating reports. Second, the non-technical defection is raised from administrators' error. (3) Theme (or research title) is the EQA system's quality improvement indicated by Key Process Output Variable (KPOV). It is used to rate the EQA system's satisfaction and to specify the way for evaluating the new system by searching for all problems from sample group for 244 days (March 1 - October 30, 2015). The employee satisfaction index of the purposed system is corrected from the survey with 10 questions: (1) Does the information system cover all functions of the users' work based on quality assurance standard completely? (2) Is the information system the friendly user interface design? (3) Does the imported data section update all information automatically? And is it useful for reducing the waste time from users' work? (4) Do the information system process and display all information speedily, accurately, and reliably? (5) Does the information system store, categorise, and search all information smoothly? (6) Does the information system backup and recover all information effectively? (7) Is all information from the information system up-to-date and compatible with other systems? (8) Is the information system useful for the practical work? (9) Does the information filling control in the information system work securely and correctly? (10) Does service maintenance of administrators instantly respond and resolve all problems effectively?

As all answers from those questions, we found that the Pareto chart can display all non-technical defections in service maintenance of administrators perfectly, e.g., instant responsibilities, fast resolutions. Its cumulative sum is 100%. It also assists to decide to fix the defections in service maintenance of administrators firstly. Then, all technical defections are being solved, e.g., information filling control, security process, backup data section, processing errors, data display, and instant responsibilities respectively as shown in figure 2.

The Measure phase has been used for searching and analyzing the capacity of the administrators' maintenance, the instant responsibilities, and urgent resolutions consequently. By the use of the Tree diagram, it can be used to reflect the procedure of resolution system service and the impact factors on both defections and indicators. The influential factors for receiving all problems can be shown as (1) Communication. The indicator is a time (per minute) (2) Helpfulness. The indicator is the decrease of the users' complaints by the random survey. (3) Courtesy, the indicator is also the decrease of the users' complaints by the random survey as used in (2). All influential factors shown in Tree diagram can be represented in figure 3. Moreover, the Criteria Weighting technique is employed for selecting all indicators by scoring the possibility of the data correction, e.g., simple, resistance, preparation, and advantages. The most scores

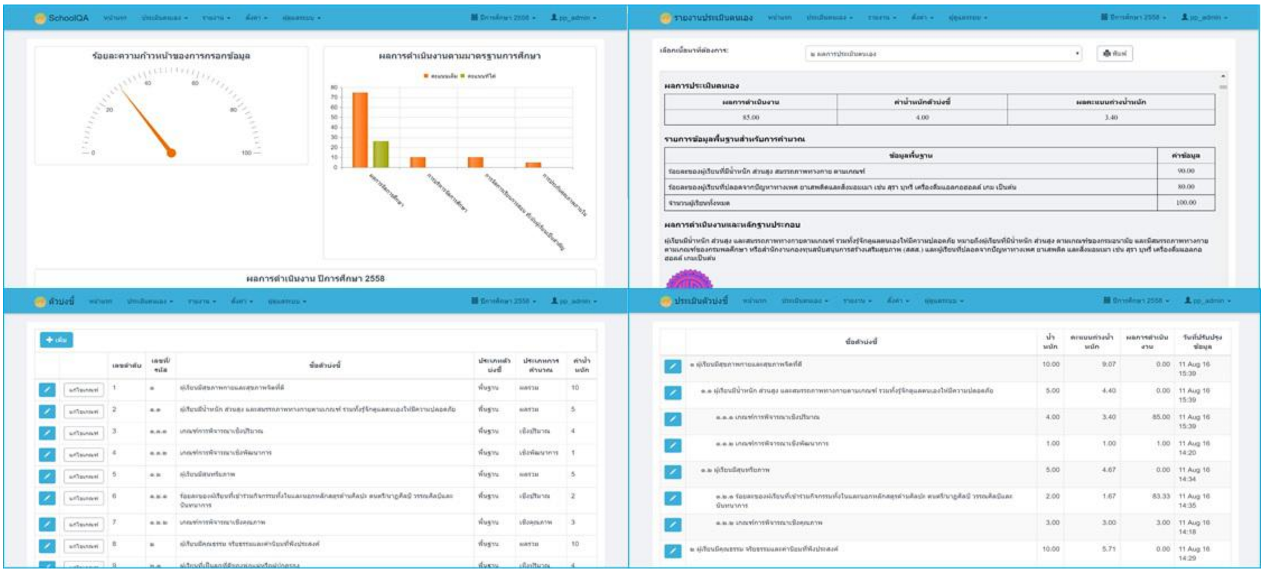


Fig.1: *The QAIS screens.*

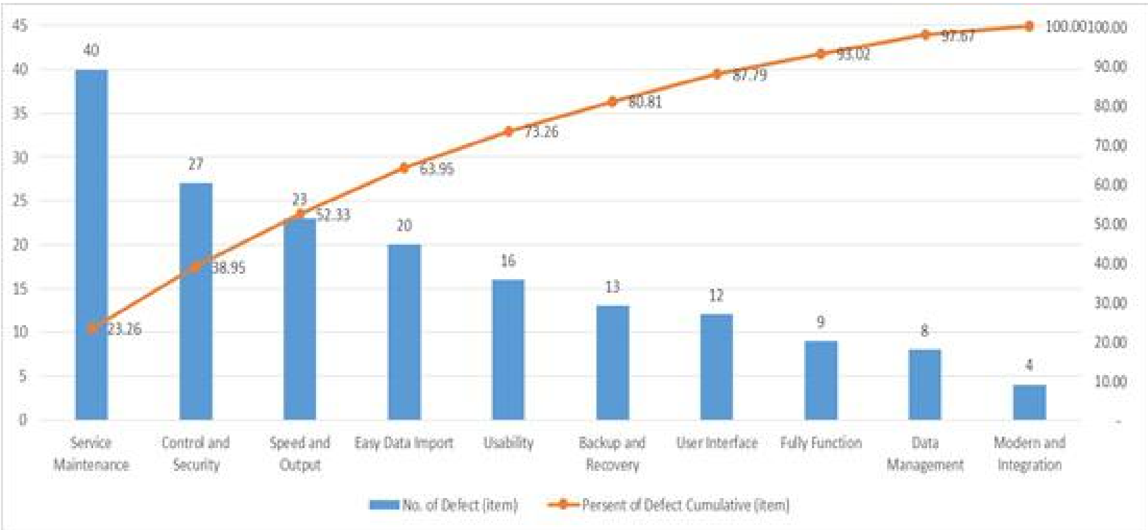


Fig.2: *The defections occurred during the QAIS experiments.*

of the indicators are chosen for data correction by the Check Sheet technique. Then, the service information is corrected by the Tree diagram, e.g., technical and non-technical issues.

From the data correction of the information system's usage for 60 times, we found that 19 deficiencies can be analyzed as the sigma rate standard compared with the number of deficiency and its possibility in 1 million samples (Defect per Million Opportunity). 19 deficiencies are found. According to the formula in [12], Unit counted is the amount of piece of work (Using the number of testing) for 60 pieces of works (or times). Finally, the defect opportunity of the chance of deficiencies is 7 chances per one piece of work (or times).

$$\text{DPMO} = \text{Defect Counted} \div (\text{Unit Counted} \times \text{Defect Opportunity}) \times 106, \tag{3}$$

e.g.,

$$\text{DPMO} = 19 \div (60 \times 7) \times 106,$$

$$\text{DPMO} = 45,238.10.$$

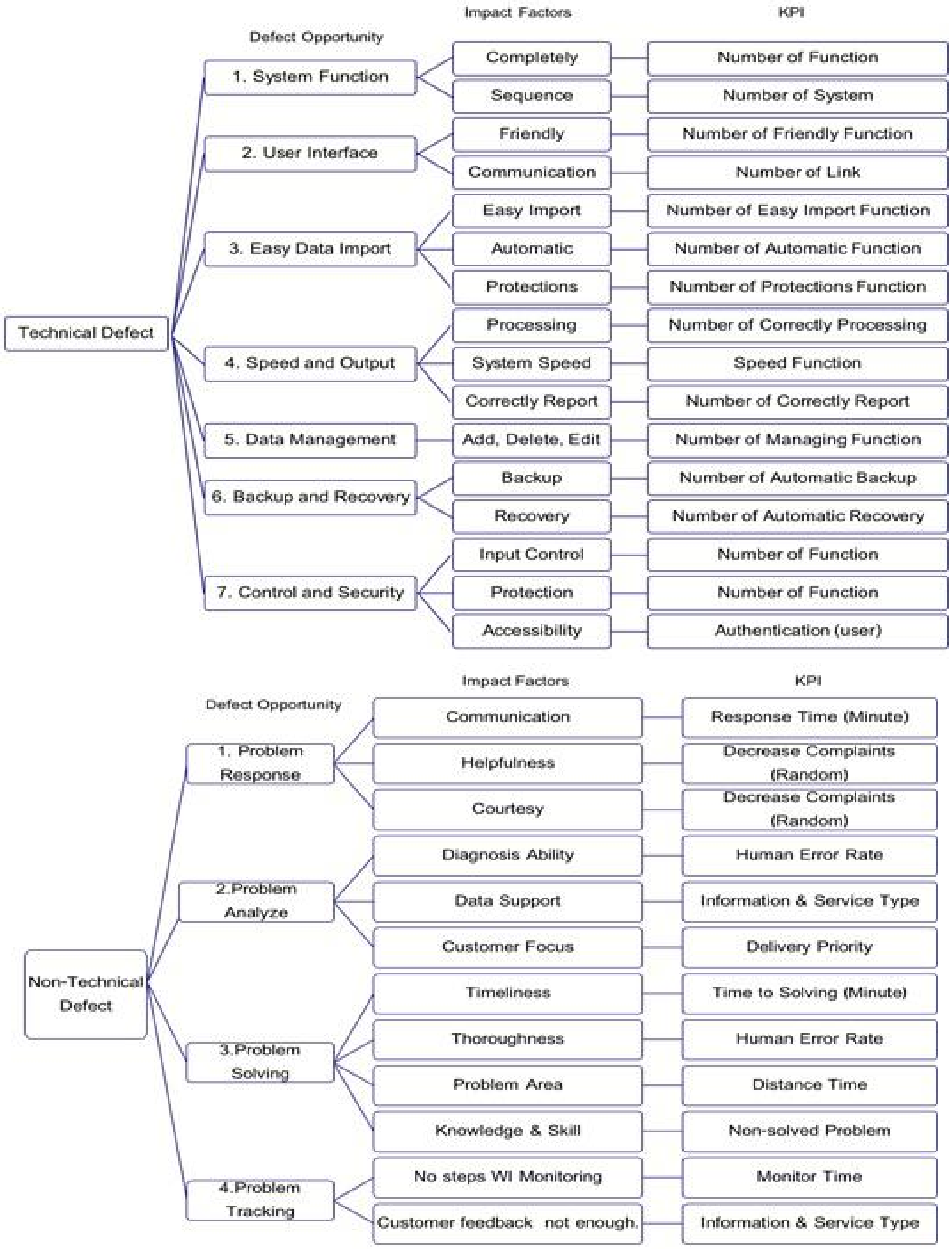


Fig.3: *The tree diagram of the efficiency of both technique and non-technique ways.*

From the lookup sigma conversion table, the short term is 3.20 sigma and the long term is 1.70 sigma. By the use of DPM0 = 45,238.10, the sigma rate is ranged at 2-sigma level by lookup the table. In other words, the deficiency per million is 45,500.124. The Sigma level is used for calculating the quality level by analysing technical information for 60 times. We found 19 deficiencies for this case. In the case of counting data, the $\bar{P}$ is calculated by the ratio of waste as:

$$\bar{P} = \frac{The\ amount\ of\ overall\ deficiency\ products\ (\Sigma np)}{The\ amount\ of\ overall\ examine\ (\Sigma p)} \tag{4}$$

e.g.,

$$\bar{P} = 19 \div 60 = 0.316.$$

The overall deficiency chance in one work is 0.316 The overall deficiency chance is calculated from $\bar{P} \div 2$. So, the deficiency chance is $0.316 \div 2 = 0.158$ chance.

The data correction of the information system for searching all non-technical deficiencies is administrators' information system services for 45 times. We found that 8 deficiencies can be analysed as the sigma rate standard compared with the number of deficiency and its possibility in 1 million samples (Defect per Million Opportunity). 8 deficiencies are found. Unit counted is the amount of piece of work (Using the number of testing) for 45 pieces of works (or times). Finally, the defect opportunity of the chance of deficiencies is 4 chances per one piece of work (or times).

$$\text{DPMO} = \text{Defect Counted} \div (\text{Unit Counted} \times \text{Defect Opportunity}) \times 106, \tag{5}$$

e.g.,

$$\text{DPMO} = 8 \div (45 \times 4) \times 106,$$

$$\text{DPMO} = 44,444.444.$$

From the lookup sigma conversion table [13], short term is 3.10 sigma and long term is 2- sigma. In other words, the deficiency per million is 45,500.124. The Sigma level is used to calculate the quality level by analysing technical information for 45 times. We found 19 deficiencies for this case. In the case of counting data, the $\bar{P}$ is calculated by the ratio of waste as:

$$\bar{P} = \frac{The\ amount\ of\ overall\ deficiency\ products\ (\Sigma np)}{The\ amount\ of\ overall\ examine\ (\Sigma p)}, \tag{6}$$

e.g.,

$$\bar{P} = 19 \div 60 = 0.316,$$

the overall deficiency chance in one work is 0.177, and the overall deficiency chance is calculated from $\bar{P} \div 2$. Therefore, the deficiency chance is $0.177 \div 2 = 0.088$.

Figure 4 shows that both the lower specification limit (LSL) and upper specifications limit (USL) are in range at the normal period. According to the sigma conversion table, its quality is in range at sigma level 2.

Figure 5 shows the performance assessment of the counting process from all analysed data. It is used as the index of the long-term process, Pp. The counting data is utilized for comparing in the scale of standard normal distribution, Z, by calculating from waste rate of $\bar{P}$. Also, the index Pp can be used for assessing the ability of capacity process by

$$Pp = \frac{1}{3} Z_{Bench}. \tag{7}$$

e.g.,

$$Pp = \frac{1}{3}(2.00),$$

$$= 0.666.$$

According to the ability process table and waste rate, we found that the ability of process of the long term is arranged in the low rate. (The index of Cp is lower than 0.67).

In the analyze phase, from the Adapted Waterfall system development cycle [14], the technical problems is solved by using system analysis and development for a new system to resolve all problems in the previous system. For the non-technical problems, we analyse the system's gap, processing inefficiency, and the ways to improve the system by using the flow analysis respectively. As the results, the technical problems from administrators are found such as the unclear workflow to solve the problems, the bottleneck of the administrators' works, and their time limitation. Those problems will seriously cause the disruption in the system, and the system has to restart everything again. This leads to waste the response time to all users in the system. A solution to prevent these negative issues is the use of the flow analysis. In addition, it also assists all administrators to respond all problems quickly, to store all information from users effectively, and to plan the solutions of the future problems accurately. The FAQs are corrected from all users to tackle the non-technical problems. By this technique, the users' satisfaction is increased with their convenience. Then, both the Improve phase and Control phase are operated to solve all old problems. They will focus on both the improvement of the problem analysis and the problem's solutions by creating the procedure manual, workflow instructions, and the user's manual.

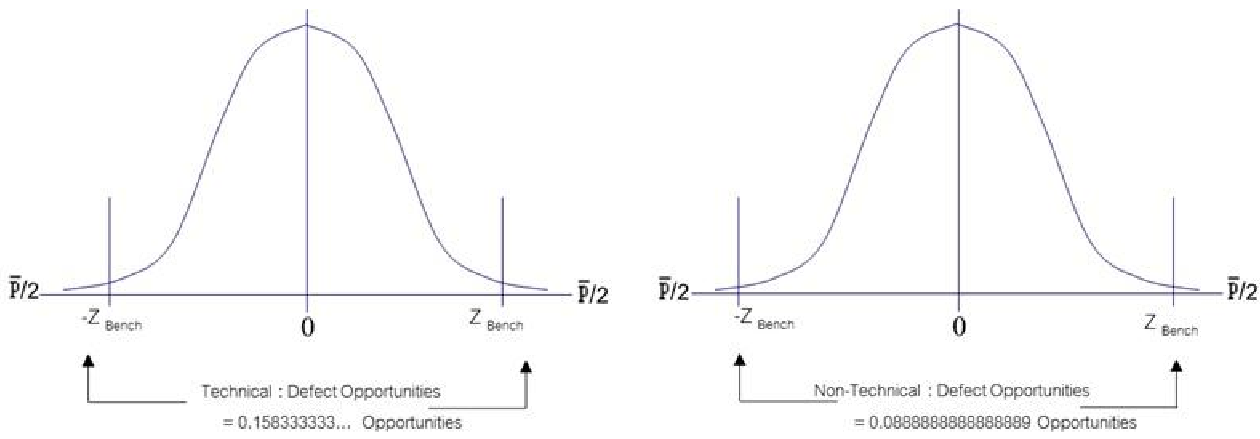


Fig.4: *The deficiency probabilities in both technique and non-technique ways.*

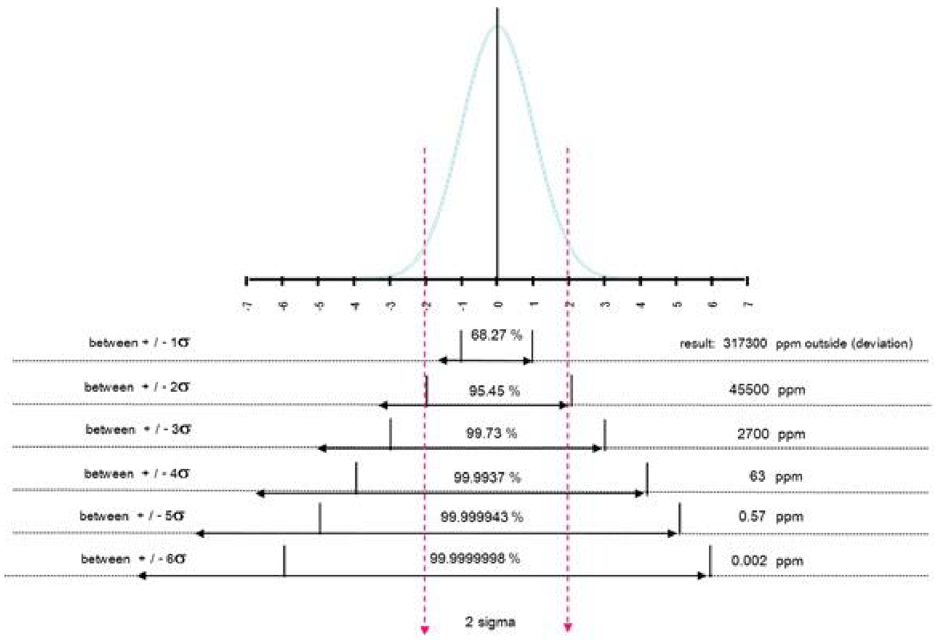


Fig.5: *The levels of processing quality.*

4. CONCLUSION

This paper presents the improvement for the quality assurance information system based on the six-sigma principle. By the use of the data correction from sample groups of the schools in the northern region in Thailand, the proposed technique gains the users' satisfaction by studying the service efficiency of the QAIS based on Sigma standard. Moreover, it also studies the causes of the users' dissatisfaction in the system in order to improve the system process effectively and to enhance the users' satisfaction including finding the alternative ways to control all working flows continually. As the results, both the technical process and the non-technical process are in the same range at the Sigma level, which is rated in the Sigma level 2. In other words, the deficiency per million is 45,500.124. Therefore, the total probabilities of deficiency for technical processes in one task are 0.316. On the other hands, the total probabilities of deficiency for the non-technical process in one task are 0.088.

To improve the technical problems, this research provides both the system analysis and the developed system for solving the old system based on Adapted Waterfall method. In addition, for the non-technical problems, this paper employs the gap analysis, the inefficiency in each process, the novel processes controlled and tracked with the developed processes. Finally, the information system is practically used in the real situations by defining both indicators and quality control points. By the use of the proposed QAIS, the system is continually tracked all problems by the improvement of the system effectively.

ACKNOWLEDGEMENT

I have furthermore to thank all participants, executives, teachers, software developers and system administrators who share their experiences and information in the pre and post surveys along with my colleagues and friends who directly and indirectly help me completing this research.

References

[1] Office of the National Education Commission. (2002). National Education Act B.E. 2542 (1999) Ministerial Regulations Rules Governing the System of Quality Assurance Methods. (2553, April 23). Government Gazette.

[2] N. Warren, M. T. Neto, J. Campbell and S. Misner. *Business Intelligence in Microsoft SharePoint 2010*. California : O'Reilly Media, Inc., 2011.

[3] Business Process Management Initiative, Business Process Modeling Notation (BPMN) Version 1.2 (Online). Available From: http://www.omg.org/spec/BPMN/1.2/(2004, January 3).

[4] O.H. Booch, J. Rumbaugh and I. Jaboson. *The Unified Modeling Language User Guide*. Reading, MA: Addison Wesley, 1999.

[5] M. Harry and R. Schroeder. *Six Sigma: The Breakthrough Management Strategy Revolutionizing the World's Top Corporations*. USA:Random House, Inc., 2005.

[6] K.Srinivasana, S.Muthub, N.K.Prasadc and G.Satheeshd. *Reduction of paint line defects in shock absorber through Six Sigma DMAIC phases*. Procedia Engineering 97 (2014) 1755 - 1764 (Online). Available From: http://www.sciencedirect.com

[7] Alexandra Teneraa,b , Luis Carneiro Pinto. A Lean Six Sigma (LSS) project management improvement model. Procedia - Social and Behavioral Sciences 119 (2014) 912 - 920 (Online). Available From: http://www.sciencedirect.com

[8] Siyu Chena,b, Shuhai Fan a,b, Jiawei Xionga,b,Wenqian Zhanga. The Design of JMP/SAP Based Six Sigma Management System and its Application in SMED. Procedia Engineering 174 (2017) 416 - 424 (Online). Available From: http://www.sciencedirect.com

[9] Sittisak Prukpitikul. (2003). Quality jump with Six Sigma. Bangkok: Technology Promotion Association Thailand-Japan.

[10] Boonchom Srisa-ard. (2002). Preliminary Research Revised Edition. Bangkok:Wattana Panich.

[11] Kitisak Ploypanichcharoen. (2005). Statistical Problem Solving (SPS). Bangkok: Technology Promotion Association Thailand-Japan.

[12] Kitisak Ploypanichcharoen. (2001). Process Capability Analysis. Bangkok: Technology Promotion Association Thailand-Japan.

[13] F. Voehl, H. J. Harrington, C. Mignosa and R. Charron. *The Lean Six Sigma Black Belt Handbook: Tools and Methods for Process Acceleration*. NW:Taylor & Francis Group, 2014.

[14] Kitti Pakdeewattanakul, Panida Panichkul. (2003). System Analysis and Design. Bangkok: KTP Comp & Consult.

Thitirath Cheowsuwan received the BA (General Management) degree from the Valaya Alongkorn Rajabhat University, Ayutthaya, Thailand, in 1994, both the MBA (Business Administration) and the M.Sc. (Information Technology) degrees from the Naresuan University, Phitsanulok, Thailand, in 1999 and 2003 respectively, and the PhD. (Computer Science) degree from Devi Ahilya Vishwavidyalaya, Indore, Madhya Pradesh, India, in 2008. Since 2011, she has been with the School of Information and Communication Technology, University of Phayao, Payao, Thailand, where she currently holds a faculty post of Associate Professor, and also currently as the Dean of the School of Information and Communication Technology, University of Phayao.

From 2001 to 2009, she was the director of the Center of Training and Computer Network Control, and then, from 2009 to 2010, she was the head of the Business Computer, Faculty of Management and Information Science, both in Naresuan University, Thailand, respectively. Her research interests include system analysis and design, object oriented analysis and design, enterprise architecture, business process analysis, and database management and development.

Somporn Saipanya received the B.S. (Management Information System) degree from Suranaree University of Technology, Nakhon Ratchasima, Thailand, in 2001. And the M.Sc. (Information Technology) degree from King Mongkut's University of Technology Thonburi, Bangkok in 2004. Her research interests include process capability, business process analysis, statistical science, mathematics and applied mathematics, and database design and development. Currently, she is a Lecturer in Bachelor of Business Administration in Business Computer, School of Information and Communication Technology, University of Phayao, Phayao, Thailand.

Supan Tongphet received the BA. (Business Computer) degree from Naresuan University, Phayao Campus, Phayao, Thailand, in 2005 and the M.Sc. (Applied Statistics and information technology) from National Institute of Development Administration, Bangkok in 2012. Her research interests business process analysis. Currently, she is a Lecturer in Bachelor of Business Administration in Business Computer, School of Information and Communication Technology, University of Phayao, Phayao, Thailand.

Iterative Based Time Domain Equalization Method for DFTS-OFDM under Highly Mobile Environments

Pongsathorn Reangsuntea[1], **Arttapol Reangsuntea**[2], **Pisit Boonsrimuang**[3], **Kazuo Mori**[4], and **Hideo Kobayashi**[5], Non-members

ABSTRACT

Under highly mobile environments, signal quality of discrete Fourier transform spreading-orthogonal frequency division multiplexing (DFTS-OFDM) with a frequency domain equalization method would be degraded relatively due to the occurrence of inter-channel interference (ICI). To solve this problem, this paper proposes an iterative based time domain equalization (TDE) method with a time domain channel impulse response (CIR) estimation method for DFTS-OFDM signal. The salient features of proposed method are to employ a time domain training sequence (TS) in the estimation of CIR instead of using the conventional pilot subcarriers and to employ the TDE method with a maximum likelihood (ML) estimation method instead of using the conventional one-tap minimum mean square error frequency domain equalization (One-Tap MMSE-FDE) method. This paper also proposes a low-complexity iterative based TDE method by using a good property of symmetric banded CIR transfer matrix for solving the simultaneous equations instead of using a direct calculation of inverse matrix. This paper presents various simulation results under highly mobile environments to demonstrate the effectiveness of proposed iterative based TDE with the CIR estimation method for the TS inserted DFTS-OFDM signal as comparing with the conventional One-Tap MMSE-FDE method.

Keywords: DFTS-OFDM, Time Domain Equalization (TDE), Channel Impulse Response (CIR) Estimation, Time Domain Training Sequence (TS), Maximum Likelihood (ML) Estimation, PCGS Algorithm, Banded Matrix

Manuscript received on August 31, 2016 ; revised on February 9, 2017.

Final manuscript received on March 26, 2017.

[1,4,5] The authors are with Electrical and Electronic Engineering, Graduate School of Engineering, Mie University, Tsu-shi, 514-8507 Japan., E-mail: pongsathorn@com.elec.mieu.ac.jp, kmori@elec.mie-u.ac.jp and koba@elec.mie-u.ac.jp

[2,3] The authors are with Telecommunication Engineering, Faculty of Engineering, King Mongkuts Institute of Technology Ladkrabang, 10520 Thailand., E-mail: arttapol.ice@gmail.com and kbpisit@kmitl.ac.th

1. INTRODUCTION

Discrete Fourier transform spreading-orthogonal frequency division multiplexing (DFTS-OFDM) has been received a lot of attentions as an alternative technique to OFDM from its lower peak to averaged power ratio (PAPR) and robustness to multipath fading by using a simple one-tap minimum mean square error frequency domain equalization (One-Tap MMSE-FDE) method [1-4]. From these advantages, the DFTS-OFDM with One-Tap MMSE-FDE method has been adopted as the standard transmission technique for the uplink from the user terminal to the base station in the 4th generation mobile communication system (LTE: Long Term Evolution) [5-6].

When the DFTS-OFDM is employed under highly mobile environments such as high speed vehicles or high speed trains, the transmitted DFTS-OFDM signal experiences severe inter-channel interference (ICI) due to the Doppler frequency spreads. Under these conditions, the time domain channel impulse response (CIR) would be no more constant even during one DFTS-OFDM symbol period. Accordingly the channel frequency response (CFR) which be used in the One-Tap MMSE-FDE method at the receiver is also changed relatively during one DFTS-OFDM symbol period. From this fact, it is impossible to mitigate the ICI by using the One-Tap MMSE-FDE method under highly mobile environments which leads the fatal degradation of bit error rate (BER) performance [7].

Up to today, many FDE methods of using the CFR transfer matrix were proposed for the OFDM signal which can mitigate ICI under highly mobile environments [8-10]. The authors proposed an iterative based FDE method for the OFDM signal which can solve the inverse of CFR matrix iteratively with a smaller computation complexity than that of using the inverse matrix calculation [10]. By using the iterative based FDE method, the order of computation complexity $O(N^3)$ required in the inverse matrix calculation can be reduced to $O(N_{Aver} \cdot N^2)$. Where N is the number of FFT points and N_{Aver} is the average number of required iterations. However, the order of complexity for the iterative based FDE method is still higher and this method is conducted in the frequency domain which requires higher computation complexity with the order of $O(N^2 \cdot \log_2 N)$ in the construction of CFR transfer matrix. To solve the problem on the FDE method, the authors proposed a time domain equalization (TDE) method for the DFTS-OFDM signal [11]. Although this method can mitigate the ICI precisely with lower complexity in the construction of time domain CIR transfer matrix, it is required to calculate the inverse matrix at every data symbol which leads extremely higher computation complexity and it is unsuitable in the practical implementation of DFTS-OFDM receiver.

To solve the above problems, this paper proposes an iterative based TDE method for the time domain training sequence (TS) inserted DFTS-OFDM signal under highly mobile environments [12]. The salient features of proposed TDE method are to employ the CIR transfer matrix in the mitigation of ICI and to employ the preconditioned conjugate gradient squared (PCGS) algorithm [13] in the calculation of inverse matrix instead of using the direct inverse matrix calculation. By using the proposed TDE method for the TS inserted DFTS-OFDM signal, the order of complexities required in the construction of time domain CIR transfer matrix and in the calculation of inverse matrix are $O(S^2 \cdot N)$ and $O(2S \cdot N \cdot N_{Aver})$, respectively which are much lower complexity than $O(N^2 \cdot N_{Aver})$ and $O(N^2 \cdot \log_2 N)$ for the conventional iterative based FDE method for the OFDM signal [10]. Where S is the length of time domain TS signal which corresponds to the guard interval (GI) of OFDM signal. The proposed iterative based TDE method can achieve much better BER performance than that for the conventional One-Tap MMSE-FDE method and almost the same BER performance as that for the TDE method of using the direct inverse matrix calculation with much lower computation complexity [11] under highly mobile environments.

The remainder of this paper is organized as follows. Section 2 introduces a problem of conventional One-Tap MMSE-FDE method for the DFTS-OFDM signal under highly mobile environments. Section 3 proposes a low-complexity iterative based TDE with a time domain CIR estimation method for the DFTS-OFDM signal. Section 4 presents various computer simulation results to verify the effectiveness of proposed iterative based TDE method as comparing with the conventional One-Tap MMSE-FDE method, and Section 5 draws some conclusions.

2. PROBLEM OF CONVENTIONAL ONE-TAP MMSE-FDE FOR DFTS-OFDM SIGNAL

When assuming the quasi-static or lower time-varying fading channels, the channel impulse response (CIR) can be considered as a constant during one DFTS-OFDM symbol period. Fig. 1 shows a schematic diagram for the relationships between the

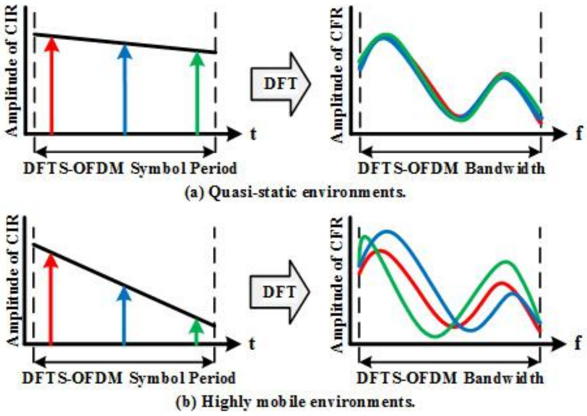


Fig.1: *Relationships between CIR and CFR in quasi-static and highly time-varying fading channels.*

CIR in the time domain and the channel frequency response (CFR) in the frequency domain where the CFR can be obtained by performing the DFT to the CIR at a certain sampling time during one DFTS-OFDM symbol period. Fig. 1(a) shows the relationships between the CIR and CFR in the quasi-static channels. From the figure, it can be seen that the CIRs at any sampling times are almost the constant. Accordingly the CFRs converted from the CIRs at any sampling times during one symbol period are also almost the constant as shown in Fig. 1(a). From this fact, the received DFTS-OFDM signal caused by lower time-varying fading distortion can be equalized precisely by using the One-Tap MMSE-FDE method with the CFR converted from the CIR at any sampling time during one symbol period. However the CIR would be no more constant even during one DFTS-OFDM symbol period in highly time-varying fading channels [11]. Fig. 1(b) shows the relationships between the CIRs and CFRs in highly time-varying fading channels. The CIRs are changing during one symbol period and the CFRs converted from the CIRs at the different sampling times are also changing as shown in Fig. 1(b). From this fact, it is difficult to mitigate the time-varying fading distortions by the One-Tap MMSE-FDE method of using the fixed CFR during one symbol period which leads the fatal degradation of BER performance.

3. PROPOSAL OF ITERATIVE BASED TIME DOMAIN EQUALIZATION METHOD

To solve the above problem on the conventional One-Tap MMSE-FDE method under highly mobile environments, this paper proposes an iterative based TDE method for the TS inserted DFTS-OFDM signal by using the estimated CIR at every sampling time.

3.1 TS inserted DFTS-OFDM system

Fig. 2(a) shows the frame format for the TS inserted DFTS-OFDM signal which be used in the CIR estimation at every sampling time. The authors proposed the CIR estimation method at every sampling time for the TS inserted OFDM signal by using the maximum likelihood (ML) estimation method [11]. This paper employs this CIR estimation method for the TS inserted DFTS-OFDM signal. The time domain training sequences TS1 and TS2 with the length of S samples are added at the both ends of every data symbol which be used in the estimation of CIR at every sampling time and also used as the role of GI to remove the inter-symbol interference (ISI). The employment of time domain TS signal in the estimation of CIR under higher mobile environments can achieve higher estimation accuracy than that for the conventional pilot subcarriers in the frequency domain. This is the reason that the time duration of TS signal is much shorter than the duration of DFTS-OFDM symbol and the fluctuation of CIR due to the Doppler spread can be considered as the constant during the period of TS signal. As for the pilot subcarriers method, the pilot subcarriers are inserted into the data subcarriers with a certain interval in the frequency domain in which the assumption of constant CIR over one DFTS-OFDM symbol time period is no more satisfied [11].

Fig. 3 shows a structure of transceiver for the TS inserted DFTS-OFDM system with the proposed iterative based TDE method. At the transmitter, the data information is encoded by a forward error correction (FEC) code [14] and the encoded data is modulated in the time domain. Then, the M modulated data is converted to M data subcarriers by M-points discrete Fourier transform (DFT) which is given by,

$$X_D(m,k) = \sum_{n=0}^{M-1} x_D(m,n) \cdot e^{-j\frac{2\pi kn}{M}} \tag{1}$$

where $x_D(m,n)$ is the time domain data signal at the n-th sampling time of the m-th symbol and $X_D(m,k)$ is the frequency domain data signal at the k-th subcarriers. Then, the M data subcarriers of $X_D(m,k)$ are mapped into within the allocated frequency bandwidth with N subcarriers which is called the subcarrier mapping. In the subcarrier mapping, M data subcarriers are mapped into the certain frequency band continuously from the data subcarrier number N_{Z1} to N_{Z2} ($N_{Z2} - N_{Z1} = M$) within N subcarriers in which the zero padding (null subcarrier) with the length of $(N-M)/2$ subcarriers are added at the both ends of M data subcarriers. Here, it should be noted that the DFT and inverse DFT (IDFT) processing are required both at the transmitter and receiver in the proposed DFTS-OFDM system. Although the required processing loads for the DFT and IDFT are higher than those for the fast Fourier transform

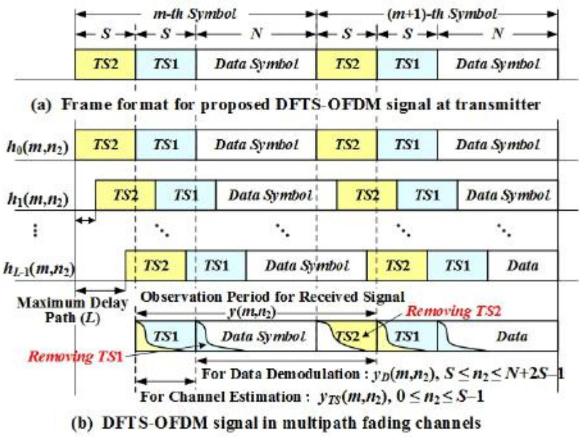


Fig.2: *Frame format of DFTS-OFDM at transmitter (a) and receiver (b) in multipath fading channels.*

(FFT) and inverse FFT (IFFT), the DFTS-OFDM technique is already employed in the uplink of LTE systems as the standard transmission technique [5-6]. From this fact, the proposed system of using DFTS-OFDM technique is also possible to implement in the practical systems.

The frequency domain signal over N subcarriers after subcarrier mapping is given by,

$$X(m,k_1) = \begin{cases} 0\ (zero\ padding), & 0 \le k_1 \le N_{Z1}-1 \\ X_D(m,k_1-N_{Z1}), & N_{Z1} \le k_1 \le N_{Z2} \\ 0\ (zero\ padding), & N_{Z2}+1 \le k_1 \le N-1 \end{cases} \tag{2}$$

where N_{Z1} and N_{Z2} in (2) are the starting and end data subcarrier numbers for M data subcarriers in the frequency domain. The starting and end data subcarrier numbers are given by $N_{Z1} = (N-M)/2$ and $N_{Z2} = M + N_{Z1} - 1$, respectively. Here, it should be noted that the zero padding is employed to avoid the aliasing occurred at the output of D/A converter. By using the zero padding, the simple analogue bandpass filter located after the D/A converter can be used to reject the aliasing which leads the easy implementation of transmitter. After subcarrier mapping, $X(m,k_1)$ including the zero padding is converted into the time domain signal as similar to the conventional OFDM signal by N-point IFFT which is given by,

$$x(m,n_1) = \frac{1}{N} \sum_{k_1=N_{Z1}}^{N_{Z2}} X(m,k_1) \cdot e^{j\frac{2\pi n_1 k_1}{N}}, \quad 0 \le n_1 \le N-1 \tag{3}$$

where $x(m,n_1)$ is the transmitted time domain DFTS-OFDM signal at the n_1-th sampling time of m-the symbol. Here, it should be noted that since the computation complexity both for the FFT and IFFT are decided by N even including the zero padding.

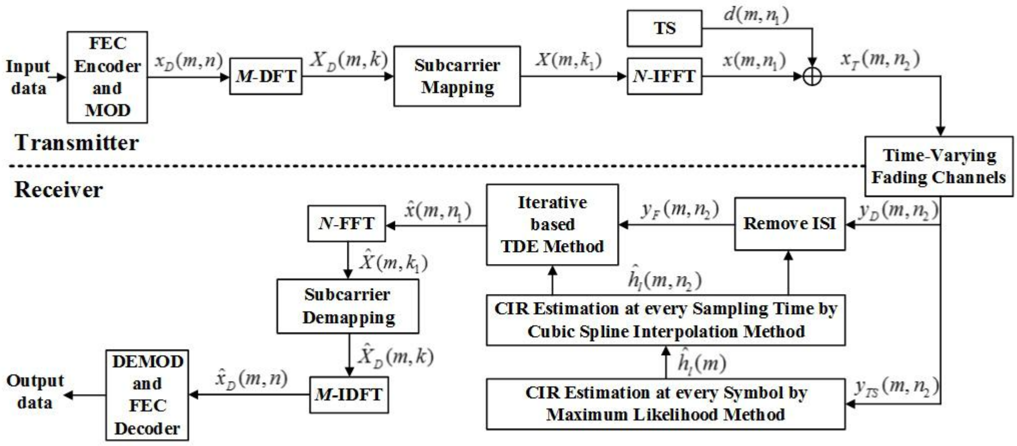


Fig.3: *Structure of transceiver for DFTS-OFDM signal with proposed iterative based TDE method.*

From this fact, the number of zero padding would not affect the computation complexity. In the proposed method, the data pattern of TS signal is generated by using some part of time domain DFTS-OFDM signal so as to achieve higher CIR estimation accuracy. The optimum data pattern of TS signal which can achieve higher CIR estimation accuracy in the multipath fading channels is selected from the computer simulation results by changing the randomly generated data patterns of DFTS-OFDM symbols.

For simplicity, this paper assumes that the data patterns both for TS1 and TS2 are assumed to be the same as $d(m, n_1)$ at the m-th symbol. The length of S should be taken longer than the length of delay paths (L) as the same as the role of GI to remove the ISI. After adding the TS1 and TS2 signals at the both ends of data symbol as shown in Fig. 2(a), the transmitted time domain signal including TS1 and TS2 signals can be expressed by,

$$x_T(m,n_2) = \begin{cases} d(m,n_2), & 0 \le n_2 \le S-1 \\ x(m,n_2-S), & S \le n_2 \le N+S-1 \\ d_2(m,n_2-N-S), & N+S \le n_2 \le N+2S-1 \end{cases} \tag{4}$$

where $x_T(m, n_2)$ is the transmitted time domain signal including TS1 and TS2 signals with the length of $N + 2S$ sampling time, d_1 and d_2 are the time domain TS1 and TS2 with the length of S sampling time ($0 \le n_2 \le S-1$) of which data patterns are known at the receiver.

As shown in Fig. 2(b), the received DFTS-OFDM signal $y(m, n_2)$ can be divided into two parts which consist of the observation period for the CIR estimation $y_{TS}(m, n_2)$ with the length of S sampling time and for the data demodulation period $y_D(m, n_2)$ with the length of $N + S$ sampling time, respectively.

3.2 CIR Estimation method for TS inserted DFTS-OFDM signal [11]

Consider for the observation period for CIR estimation, the received time domain TS1 signal $y_{TS}(m, n_2)$ during the observation period for CIR estimation from 0 to $S-1$ as shown in Fig. 2(b) can be expressed by,

$$y_{TS}(m,n_2) = \sum_{l=0}^{L-1} h_1(m,n_2) \cdot d_1(m,n_2-l) + z(m,n_2), \quad 0 \le n_2 \le S-1 \tag{5}$$

where $d_1(m, n_2)$ is the TS1 signal given in (4), $hl(m, n_2)$ is the complex amplitude of CIR for the l-th delay path at the n_2-th sampling time of m-th symbol and $z(m, n_2)$ is the additive white Gaussian noise (AWGN). Here it is assumed that $h_l(m, n_2)$ is the constant during the short period of TS1 even under highly mobile environments. Under the above assumption, the expected received TS1 signal passed through the multipath fading channels can be expressed by,

$$\hat{y}_{TS}(m,n_2) = \sum_{l=0}^{S-1} \hat{h}_1(m) \cdot d(m,m_2-l) \quad 0 \le n_2 \le S-1 \tag{6}$$

The unknown parameters of complex amplitude of $\hat{h}l(m)$ at the m-th symbol can be estimated by using the maximum likelihood (ML) estimation method under the constraint with minimizing the difference between the actual received time domain TS signal in (5) and the expected received time domain signal in (6). Here the estimated CIR $\hat{h}_l(m)$ corresponds to the CIR at the middle sampling time of TS1. Here, it should be noted that the fluctuation of CIR even under higher mobile environments can be considered as the constant because of employing the short time period of TS signal. From this fact, the CIR estimated by using the TS signal under higher mobile environments can achieve higher estimation accuracy than that for using pilot subcarriers assigned over one DFTS-OFDM symbol in the frequency domain as mentioned in section 3.1. Then, the time domain CIR $\hat{h}_l(m, n_2)$ at every sampling time can be estimated by applying the cubic spline interpolation method for the CIR estimated at every symbol over one frame. By using the estimated CIR at every sampling time over one data symbol duration, the received data symbol can be equalized in the time domain so as to mitigate the Doppler spread. In other words, the proposed TDE method can mitigate the Doppler spread at every time sampling basis in the time domain which is completely different from the conventional frequency domain equalization by using the estimated CFR for each subcarrier [11].

3.3 Proposed Time Domain Equalization Method

In the proposed time domain equalization (TDE) method, the estimated CIRs at every sampling time are employed in the construction of CIR transfer matrix which enables the compensation of Doppler spread at every sampling time basis. This feature of proposed TDE method is completely different from the conventional One-Tap MMSE-FDE method in which the CFR converted from the fixed CIR during one symbol period is employed which leads the fatal degradation of BER performance. By assuming the actual channel impulse response $h_l(m, n_2)$ at every sampling time, the received time domain data signal $y_D(m, n_2)$ during the observation period for the data demodulation from S to $N + 2S - 1$ as shown in Fig. 2(b) can be expressed by,

$$y_D(m, n_2) = \sum_{l=0}^{L-1} h_l(m, n_2) \cdot x_T(m, n_2 - l) + z(m, n_2), \quad S \leq n_2 \leq N + 2S - 1 \tag{7}$$

where $x_T(m, n_2 - l)$ corresponds to the transmitted time domain signal given in (4) and the following relationships are employed in the derivation of (7).

$$n_2 - l \leq S - 1, \quad x_T(m, n_2 - l) = d(m, n_2 - l) \tag{8}$$

$$n_2 - l \leq N + S, \quad x_T(m, n_2 - l) = d(m, n_2 - N - S - l) \tag{9}$$

The received data signal in (7) includes the ISI which are added at the start and end of data symbol from the TS1 and TS2 signal, respectively as shown in Fig. 2(b). Since the data pattern of TS signal are known at the receiver, the ISI caused from the TS signals in the multipath fading channel can be removed by using the estimated CIRs at every sampling time and the known data pattern of TS signal which is given by,

$$y_F(m, n_2) = \begin{cases} y_D(m, n_2) - \sum_{l=n_2-S+1}^{S-1} \hat{h}_l(m, n_2) \cdot d(m, n_2 - l), \\ \quad (S \leq n_2 \leq 2S - 2) \\ y_D(m, n_2), (2S - 1 \leq n_2 \leq N + S - 1) \\ y_D(m, n_2), \sum_{l=0}^{n_2-N-S} \hat{h}_l(m, n_2) \cdot d(m, n_2 - N - S - l), \\ \quad (N + S \leq n_2 \leq N + 2S - 2) \end{cases} \tag{10}$$

where $y_F(m, n_2)$ is the received time domain signal after removing the ISI from the actual received DFTS-OFDM signal $y_D(m, n_2)$ in (7). When the transmitted time domain data signal $x(m, n_1)$ given in (3) is assumed as the unknown parameters, the expected time domain received data signal $\hat{y}_E(m, n_2)$ without the ISI which corresponds to (10) can be given by,

$$\hat{y}_E(m, n_2) = \begin{cases} \sum_{l=0}^{n_2-S} \hat{h}_l(m, n_2) \cdot \hat{x}(m, n_2 - S - l), \\ \quad (S \leq n_2 \leq 2S - 2) \\ \sum_{l=0}^{S-1} \hat{h}_l(m, n_2) \cdot \hat{x}(m, n_2 - S - l), \\ \quad (2S - 1 \leq n_2 \leq N + S - 1) \\ \sum_{l=n_2-N-S+1}^{S-1} \hat{h}_l(m, n_2) \cdot \hat{x}(m, n_2 - S - l), \\ \quad (N + S \leq n_2 \leq N + 2S - 2) \end{cases} \tag{11}$$

The unknown parameter of time domain data signal $\hat{x}(m, n_1)$ can be estimated by solving the following maximum likelihood (ML) equation under the constraint with minimizing the difference between the actual received data signal $y_F(m, n_2)$ in (10) and the expected received data signal $\hat{y}_E(m, n_2)$ in (11) which can be expressed by,

$$\Upsilon = \arg\min_{\hat{x}(m,s)} \left[\sum_{n_2=S}^{N+2S-2} |y_F(m, n_2) - \hat{y}_E(m, n_2)|^2 \right], 0 \leq n_2 \leq N + 2S - 2 \tag{12}$$

The ML equation in (12) can be solved by taking the partial differentiation for unknown parameters of $\hat{x}^*(m, n_3)$ which can be expressed by,

$$\frac{\partial \Upsilon}{\partial \hat{x}^*(m,n_3)} = \frac{\partial\left(\sum_{n_2=n_3}^{N+2S-2} |y_F(m,n_2) - \hat{y}_E(m,n_2)|^2\right)}{\partial \hat{x}^*(m,n_3)} = 0$$
$$, 0 \le N_3 \le N-1 \tag{13}$$

where $(\cdot)^*$ represents the conjugate complex. By using (13), the ML equation (12) can be expressed by the following simultaneous equations with N unknown parameters of $\hat{x}(m,n_1)$.

$$[b(m,n_3)]_{N\times 1} = [A_m(n_3,n_1)]_{N\times N} \cdot [\hat{x}(m,n_1)]_{N\times 1} \tag{14}$$

where $b(m,n_3)$ and $A_m(n_3,n_1)$ can be expressed by,

$$\begin{aligned} b(m,n_3) &= \sum_{n_2=S}^{N+2S-2} y_F(m,n_2) \frac{\partial \hat{y}_E^*(m,n_2)}{\partial \hat{x}^*(m,n_3)} \\ &= \underbrace{\left[\hat{h}_l^H(m,n_3)\right]}_{N\times(N+S-1)} \cdot \underbrace{[y_F(m,n_1)]}_{(N+S-1)\times 1},\ 0 \le n_3 \le N-1 \end{aligned} \tag{15}$$

$$\begin{aligned} A_m(n_3,n_1) &= \sum_{n_2=S}^{N+2S-2} \hat{y}_E(m,n_2) \frac{\partial \hat{y}_E^*(m,n_2)}{\partial \hat{x}^*(m,n_3)} \\ &= \underbrace{\left[\hat{h}_l^H(m,n_3)\right]}_{N\times(N+S-1)} \cdot \underbrace{\left[\hat{h}_l(m,n_1)\right]}_{(N+S-1)\times 1},\ 0 \le n_1 \le N-1 \end{aligned} \tag{16}$$

where $(\cdot)^H$ represents the Hermitian transpose operation and the following relationships are used in the derivation of (15) and (16).

$$n_3 = n_2 - S - l \ \text{ in } \ \partial \hat{y}_E^*(m,n_2)/\partial \hat{x}^*(m,n_3) \tag{17}$$

$$n_1 = n_2 - S - l \ \text{ in } \ \hat{y}_E^*(m,n_2) \tag{18}$$

Here the CIR matrix $[\hat{h}_l(m,n_1)]$ in (16) is given by,

$$[\hat{h}_l(m,n_1)] = \begin{bmatrix} \hat{h}_0(m,S) & 0 & \cdots & 0 \\ \hat{h}_1(m,S+1) & \hat{h}_0(m,S+1) & \ddots & \vdots \\ \vdots & \hat{h}_1(m,S+2) & \ddots & 0 \\ \hat{h}_{S-1}(m,2S-1) & \vdots & \ddots & \hat{h}_0(m,N+S-1) \\ 0 & \hat{h}_{S-1}(m,2S-1) & \vdots & \hat{h}_1(m,N+S) \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \hat{h}_{S-1}(m,N+2S-2) \end{bmatrix} \tag{19}$$

From (19), the order of complexity required in the construction of CIR transfer matrix $A_m(n_3,n_1)$ in (16) which is obtained by the multiplication of $\hat{h}_l(m,n_1)$ and $\hat{h}_l^H(m,n_1)$, can be given by $O(S^2 \cdot N)$. The proposed TDE method of using the CIR transfer matrix can achieve lower complexity than that for the conventional FDE method [10] which employs the full elements of CFR transfer matrix with the complexity of $O(N^2 \cdot \log_2 N)$. The CIR transfer matrix $[A_m(n_3,n_1)]$ in (16) which is obtained after the partial differentiation can be represented by,

$$A_{i,j} = \begin{bmatrix} A_{0,0} & A_{0,1} & \cdots & A_{0,S-1} & 0 & \cdots & 0 \\ A_{0,1}^H & A_{1,01} & & & \ddots & \ddots & \vdots \\ \vdots & & \ddots & & & \ddots & 0 \\ A_{0,S-1}^H & & & \ddots & & & A_{N-S,N-1} \\ 0 & \ddots & & & \ddots & & \vdots \\ \vdots & \ddots & \ddots & & & A_{N-2,N-2} & A_{N-2,N-1} \\ 0 & \cdots & 0 & A_{N-S,N-1}^H & \cdots & A_{N-2,N-1}^H & A_{N-1,N-1} \end{bmatrix} \tag{20}$$

In (20), $A_m(i,j)$ in (16) is represented by $A_{i,j}$ and the index of m-th symbol is omitted for brevity. From (20), it can be seen that the matrix is the banded matrix with the upper and lower bandwidth (S-1) whose non-zero entries are confined to a diagonal bands. Also the lower band matrix with the index of (j,i) below the diagonal terms is the Hermitian transpose of upper band matrix with the index of (i,j). From these facts, the CIR transfer matrix $[A_m(n_3,n_1)]$ in (16) is the symmetric banded matrix with the block size of $(N \times N)$. This paper employs the good property of symmetric banded matrix in the proposed iterative based TDE method for further reduction of computation complexity.

From (14), the unknown parameters $\hat{x}(m,n_1)$ can be simply solved by using the inverse matrix of $[A_m(n_3,n_1)]$ which is given by,

$$[\hat{x}(m,n_1)]_{N\times 1} = [A_m(n_3,n_1)]^{-1}_{N\times N} \cdot [b(m,n_3)]_{N\times 1} \quad (21)$$

where $[\cdot]^{-1}$ represents the inverse matrix. In the demodulation of time domain data information for the DFTS-OFDM signal, the estimated $\hat{x}(m,n_1)$ in (21) is converted to the frequency domain signal $\hat{X}(m,k)$ as given in (2) by N-points FFT. After subcarrier demapping to $\hat{X}(m,k)$, M data subcarriers $\hat{X}_D(m,k)$ as given in (2) can be obtained in the frequency domain. Then $\hat{X}_D(m,k)$ is converted to the time domain data $\hat{x}_D(m,n)$ by M-points IDFT as given in (1). The time domain information data can be obtained after demodulation and FEC decoding for $\hat{x}_D(m,n)$ which are all the opposite processing at the transmitter side as shown in Fig. 3.

In (21), the order of computation complexity for the calculation of inverse matrix with size of $N\times N$ is $O(N^3)$ which is required at every data symbol demodulation. To reduce the computation complexity in solving the simultaneous equations for the proposed TDE method, the next section proposes an iterative based TDE method which employs the precondition conjugate gradient squared (PCGS) algorithm for the symmetric banded CIR transfer matrix as given in (20).

3.4 Proposed Iterative based TDE Method

The conjugate gradient squared (CGS) algorithm [13] is well known as one of the iterative methods which can solve the linear simultaneous equations for N unknown parameters with much smaller computation complexity as compared with that of using the inverse matrix calculation. Let consider the simultaneous equations $\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{b}$, where $\boldsymbol{A}$ corresponds to $A_m(n_3,n_1)$ in (14) and its matrix is the symmetric banded matrix with the size of $N\times N$. The exact CGS solution can be obtained after at most N steps. Hence, stopping the iteration after $N_{iter}(<N)$ steps would yield an approximate solution for the problem. In the equalization of every data symbol, the CGS algorithm minimizes iteratively the cost function in a reduced-rank Krylov subspace. When the spectral condition number of the matrix $\boldsymbol{A}$ is too high, a preconditioned matrix $\boldsymbol{D}$ is employed which is called the precondition CGS (PCGS) algorithm. The PCGS algorithm solves the simultaneous equations by,

$$\boldsymbol{D}^{-1}\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{D}^{-1}\boldsymbol{b} \quad (22)$$

where the inversion of matrix $\boldsymbol{D}$ should be a computationally efficient operation. In the rest of our analysis, assuming the simplicity that the matrix $D(m)$ is constructed by the diagonal of matrix $A_m(n_3,n_1)$ and the initial solution of $\hat{x}(m,n_1)$ can be given by,

$$[\hat{x}(m,n_1)]^{(0)}_{N\times 1} = [D(m)]^{-1}_{N\times N} \cdot [b(m,n_3)]_{N\times 1} \quad (23)$$

In PCGS algorithm [13], the residual vector $\boldsymbol{r^{(i)}}$ can be regarded as the product of $\boldsymbol{r^{(0)}}$ and an i-th degree polynomial in matrix $\boldsymbol{A}$ which is expressed by,

$$\boldsymbol{r^{(i)}} = \boldsymbol{P_i(A)r^{(0)}} \quad (24)$$

where $\boldsymbol{P_i(A)}$ is the polynomial of $\boldsymbol{A}$ at the i-th degree and $\boldsymbol{r^{(0)}}$ is obtained from the initial solution given in (23) as $\boldsymbol{r^{(0)}} = \boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}^{(0)}$. The iteration coefficients can be recovered from the i-th vectors $\boldsymbol{r^{(i)}}$ and it turns out to be easy to find the corresponding approximations for $[\hat{x}(m,n_1)]$. From (24), the multiplication of $\boldsymbol{P_i(A)r^{(0)}}$ is required at every symbol. Since the CIR transfer matrix $\boldsymbol{A}$ given in (20) is the symmetric banded matrix, the order of complexity for this multiplication requires only $O(2S\cdot N)$ which is lower than the conventional iterative based FDE method of using the full elements of CFR transfer matrix [10] which requires the order of complexity $O(N^2\cdot\log_2 N)$.

The repetition of PCGS algorithm is stopped when the following normalized mean square error (NMSE) between the $(i-1)$-th and i-th solutions of $[\hat{x}(m,n_1)]$ is smaller than the predetermined threshold level (TOL) [10].

$$NMSE = \frac{\sum_{n_1=0}^{N-1}\left|[\hat{x}(m,n_1)]^{(i)} - [\hat{x}(m,n_1)]^{(i+1)}\right|^2}{\sum_{n_1=0}^{N-1}\left|[\hat{x}(m,n_1)]^{(i)}\right|^2} \quad (25)$$

In the proposed iterative based TDE with the PCGS algorithm, the following procedures are repeated up to either the value of $NMSE$ in (25) becomes less than the predetermined threshold level (TOL) or the number of iterations reaches to the predetermined maximum number (N_{max}).

Step 1: The maximum iteration number is set to N_{max} ($N_{max} = 5$), the threshold level is set to TOL and the initial solution of $[\hat{x}(m,n_1)]^{(0)}$ is given by (23).

Step 2: Calculate the $i-th$ solution of $[\hat{x}(m,n_1)]^{(i)}$ by the PCGS algorithm and calculate the $NMSE$ by (25).

Step 3: Compare the $NMSE$ obtained at the $i-th$ iteration with the predetermined threshold level of TOL. If the $NMSE$ is less than TOL, the $[\hat{x}(m,n_1)]^{(i)}$ is output as the estimated data information. If not, repeat the same procedures. If the number of iterations reaches to predetermined N_{max}, $[\hat{x}(m,n_1)]^{(N_{max})}$ is output as the estimated data information.

The order of computation complexity for the proposed PCGS algorithm can be evaluated by $O(2S \cdot N \cdot N_{Aver})$ which is lower complexity than that for $O(N^2 \cdot N_{Aver})$ when the PCGS algorithm is applied to the conventional FDE method [10] with the full elements of transfer CFR matrix. The order of complexity ratio between TDE with the inverse matrix calculation $O(N^3)$ and proposed TDE with the iterative method is evaluated by the following equation.

$$R_c = \frac{2S \cdot N \cdot N_{Aver}}{N^3} = \frac{2S \cdot (N_{Aver})}{N^2} \tag{26}$$

where N_{Aver} is the average number of required iterations which satisfies the predetermined threshold level of TOL. The average number of required iterations which may depend on the predetermined threshold level TOL, mobile environments and operating carrier to noise power ratio (C/N), which are evaluated by computer simulations in the next section.

4. PERFORMANCE EVALUATIONS

In this section, various computer simulations are conducted to evaluate the performance of proposed iterative based TDE method as comparing with the conventional One-Tap MMSE-FDE method in highly time-varying fading channels. The simulation parameters to be used in the following evaluations are listed in Table 1. The total numbers of subcarriers (FFT/IFFT points) is N=128, the numbers of data subcarriers (DFT/IDFT points) is M=96, and the numbers of null subcarriers both ends of data subcarriers (zero padding) is $N - M = 32$. The communications channel is modelled by the Rician multipath fading channel which is usually experienced by the user on the higher speed vehicles or trains [7]. In the following evaluations, the normalized Doppler frequency $R_D = f_{dmax}/\Delta f$ (%) is employed as the measure of mobile conditions where f_{dmax} is the maximum Doppler frequency and Δf is the subcarrier spacing of DFTS-OFDM signal.

The estimation accuracy for the time domain CIR at every sampling time is evaluated by the normalized mean square error (NMSE) which can be expressed by,

$$\Psi_{NMSE} = \frac{\sum_{l=0}^{L-1} \sum_{m=0}^{N_s-1} \sum_{n_2=0}^{N+2S-1} \left| \hat{h}_l(m, m_2) - h_l(m, m_2) \right|^2}{\sum_{l=0}^{L-1} \sum_{m=0}^{N_s-1} \sum_{n_2=0}^{N+2S-1} \left| h_l(m, n_2) \right|^2} \tag{27}$$

where N_S is the number of data symbols per one DFTS-OFDM frame.

Fig. 4 shows the time domain CIR estimation accuracy at every sampling time which is evaluated by the normalized mean square error (NMSE) given in (27) for the proposed CIR estimation method of using the time domain TS when changing the normalized Doppler frequency R_D and operation C/N. From the figure, it can be observed that the proposed CIR estimation method of using the time domain TS signal can keep higher estimation accuracy even at lower operation C/N when R_D is up to 20% which corresponds to the vehicle speed at 381km/hr.

Table 1: *Simulation parameters.*

Parameters	Values
No. of FFT/IFFT points (N)	128
No. of DFT/IDFT points (M)	96
No. od data subcarriers (M)	96
No. of zero padding ($N - M$)	32
Length of GI for One-Tap MMSE-FDE	16
Length of training sequence (S)	16
Modulation for training sequence (TS)	16QAM
Modulation for data information	16QAM
No. of symbols per one frame (N_S)	33
Allocated frequency bandwidth	1MHz
Radio frequency	5.9GHz
Forward Error Correction (FEC) Codec [14]	
Encoding	Convolution
FEC rate	1/2
Constraint length	7
Decoding	Viterbi with hard decision
Interleaver	Matrix size with one frame
Rician multipath fading channel model	
Rice factor (K)	6dB
Delay profile for Rayleigh fading	Exponential
Decay constant	-1dB
No. of delay paths(L)	14
No. of scattered rays per one delay path	20

Fig. 5 shows BER performances for the proposed TDE method of using the direct inverse matrix calculation and the proposed iterative method with the PCGS algorithm when changing the threshold level (TOL) of $NMSE$ at the normalized Doppler frequency R_D=15% which correspond to the vehicle speed at 286 km/hrs. Here it should be noted that all the BER performances in Fig. 5 are evaluated by performing 5 iterations with regardless of TOL. From Fig. 5, it can be seen that the BER performance of proposed iterative method can achieve almost the same performance as that for the direct inverse matrix calculation when the TOL is smaller than 2×10^{-2}. From the figure, it can also be observed that the proposed iterative method can achieve almost the same BER performance as that for the direct inverse matrix calculation after performing 5 iterations even at $TOL = 10^{-3}$ which would require more iterations to satisfy the $TOL = 10^{-3}$ than $TOL = 2 \times 10^{-2}$. From the above results, this paper employs $N_{max} = 5$ and the threshold level $TOL = 2 \times 10^{-2}$ in the following evaluations.

Fig. 6 shows the average number of iterations N_{Aver} when N_{max} and TOL are fixed by 5 and 2×10^{-2} based on the results in Fig. 5 and changing the normalized Doppler frequency (R_D). From the figure, it can be observed that the average number of

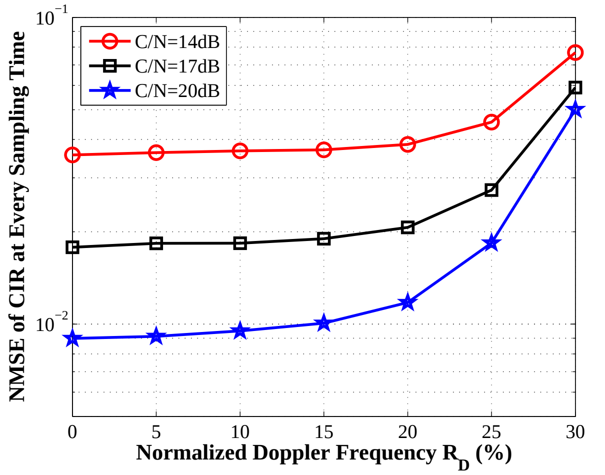


Fig.4: *CIR estimation accuracy when changing R_D and operation C/N.*

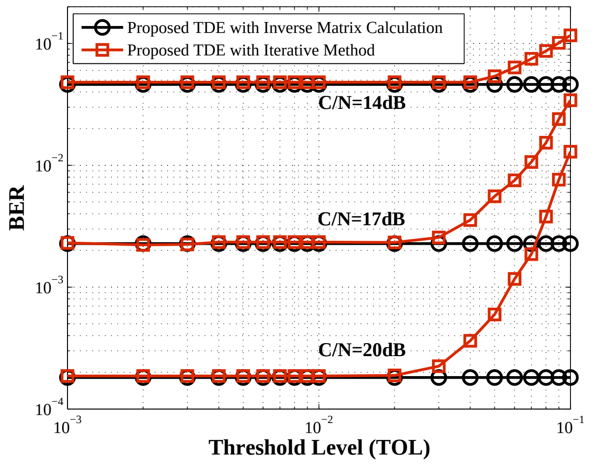


Fig.5: *BER performance for proposed iterative based TDE method when changing TOL and C/N at R_D=15%.*

iterations N_{Aver} becomes larger as increasing the normalized Doppler frequency R_D and as decreasing the operation C/N. From Fig. 6, it can also be observed that the average number of iterations N_{Aver} to satisfy $TOL = 2 \times 10^{-2}$ is always less than 3.5 even at lower C/N and higher R_D=15%. From Figs. 5 and 6, it can be concluded that the proposed iterative method can achieve almost the same BER performance as that for the inverse matrix calculation with keeping lower N_{Aver} than N_{max}=5 which enables the reduction of computation complexity.

Table 2 shows the order of complexity ratio R_C defined in (26). The orders of computation complexities for the proposed iterative based TDE method and the inverse matrix calculation method can be evaluated by $O(2S \cdot N \cdot N_{Aver})$ and $O(N^3)$, respectively. From these facts, the orders of complexities for both methods are highly depending on the size of parameters N. From the table, it can be observed that the average number of iterations for the proposed iterative method is decreasing as increasing the operation C/N. This is the reason that the larger number of iterations is required to satisfy the predetermined small TOL at lower C/N. However, the average number of iterations NAver is always less than 3.5 for all operation C/N and R_D. From these results in table 2, it can be concluded that the proposed iterative based TDE method can reduce the computation complexity by about 150 times as compared with the inverse matrix calculation method. Here it should be noted that the proposed iterative based TDE method can achieve much lower computation complexity especially when employing larger N.

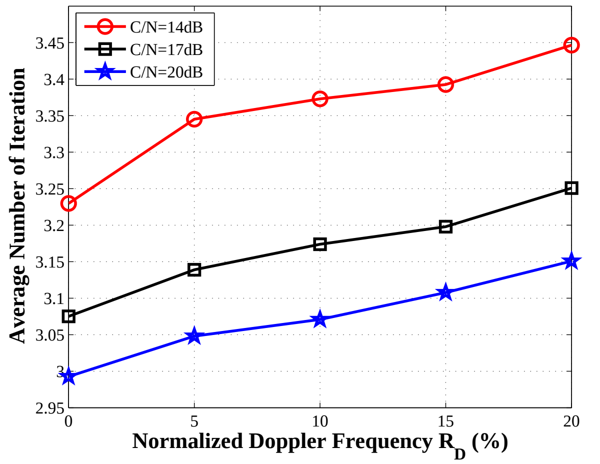


Fig.6: *Average number of required iterations for proposed iterative based TDE method when changing R_D and C/N at TOL= 0.02.*

Table 2: *Ratio of computation complexity for proposed iterative based TDE method.*

C/N	Proposed iterative based TDE method ($N = 128, S = 16, TOL = 0.02$)			
	$R_D(= f_{dmax}/\Delta f) = 5\%$		$R_D(= f_{dmax}/\Delta f) = 15\%$	
	N_{Aver}	Complexity ratio R_C in (26)	N_{Aver}	Complexity ratio R_C in (26)
14dB	3.34	0.0065	3.39	0.0066
17dB	3.13	0.0061	3.19	0.0062
20dB	3.04	0.0059	3.11	0.0061

Fig. 7 shows the BER performances when changing the normalized Doppler frequency R_D at C/N=20dB for the conventional One-Tap MMSE-FDE and the proposed TDE with both the inverse matrix calculation and proposed iterative methods. From the figure, it can be observed that the proposed iterative based TDE method shows much better BER performance than the conventional One-Tap MMSE-FDE method especially when the normalized Doppler frequency R_D is higher. The proposed iterative based TDE method shows almost the same BER performance as that for the inverse matrix calculation method with much lower complexity even when the normalized Doppler frequency R_D is up to 15%.

Fig. 8 shows the BER performances when changing C/N at R_D=15% for the conventional One-Tap MMSE-FDE method, proposed TDE with both the iterative method and direct inverse matrix method. From the figure, it can be observed that the proposed TDE with the iterative method can achieve much better BER performance than the conventional One-Tap MMSE-FDE method and achieve almost the same BER performance as the proposed TDE with the inverse matrix calculation method. From the results in Table 2 and Fig. 8, it can be concluded that the proposed TDE with the iterative method can reduce the computation complexity by 150 times with keeping the same BER performance as that for the inverse matrix calculation method.

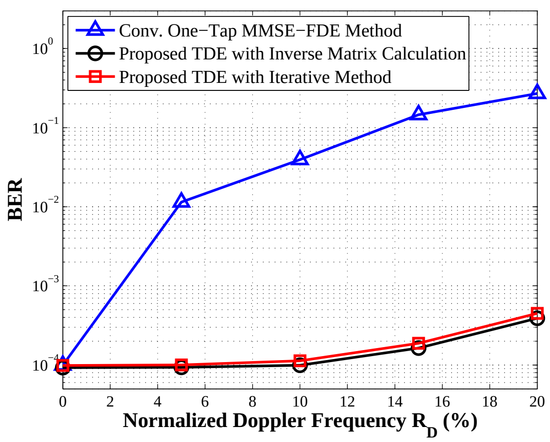


Fig.7: *BER performance for proposed iterative based TDE method when changing R_D at C/N=20dB.*

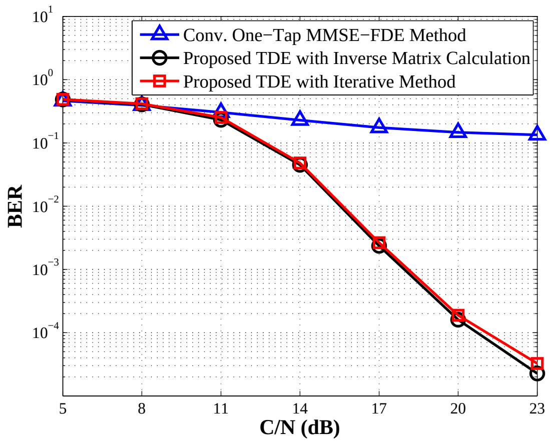


Fig.8: *BER performance for proposed iterative based TDE method when changing C/N at R_D =15%.*

From Figs. 7 and 8, it can be seen that the BER performances of proposed method are not exactly equal to the performance of conventional inverse matrix calculation. This is the reason that the proposed iterative method is one of approximation methods for the calculation of inverse matrix which can reduce the computation complexity with keeping almost the same BER performance as that for the inverse matrix calculation.

5. CONCLUSIONS

This paper proposed the low-complexity iterative based TDE method with the time domain CIR estimation method for the TS inserted DFTS-OFDM signal under highly mobile environments. The proposed method employs the partial differentiation in solving the ML equation so as to be the symmetric banded CIR transfer matrix which allows the employment of iterative method for solving the simultaneous equations iteratively with much lower computation complexity than that for the inverse matrix calculation. From the computer simulation results, it can be concluded that the proposed iterative based TDE method can achieve much better BER performance than the conventional One-Tap MMSE-FDE method under highly mobile environments. It can be also concluded that the proposed TDE with the iterative method can achieve much lower computation complexity than that for the inverse matrix calculation method with keeping almost the same BER performance.

ACKNOWLEDGEMENT

The authors would like to thank to the Japanese Government (Monbukagakusho: MEXT) Scholarships who supported some part of this research.

References

[1] D. Falconer, S. L. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson, “Frequency domain equalization for single-carrier broadband wireless systems,” *IEEE Commun. Mag.*, vol. 40, pp. 58-66, Apr. 2002.

[2] N. Benvenuto and S. Tomasin, “On the comparison between OFDM and single carrier modulation with a DFE using a frequency-domain feed forward filter,” *IEEE Trans. Commun. Mag.*, vol. 50, pp. 947-955, Jun. 2002.

[3] G. M. Hyung, L. Junsung, and J. G. David, “Single carrier FDMA for uplink wireless transmission,” *IEEE Vehicular Technology. Mag.*, vol. 1, no. 3, pp. 30-38, Sep. 2006.

[4] D. Y. Seol, U. K. Kwon, and G. H. Im, “Performance of single carrier transmission with cooperative diversity over fast fading channels,” *IEEE Trans. Commun.*, vol. 57, no. 9, pp. 2799-2807, Sep. 2009.

[5] A. Ghosh, R. Ratasuk, B. Mondal, N. Mangalvedhe, and T. Thomas, "LTE-advanced: next-generation wireless broadband technology," *IEEE Wireless Commun.*, vol. 17, no. 3, pp. 10-22, Jun. 2010.

[6] 3GPP TS 36.211, Evolved Universal Terrestrial Radio Access (EUTRA); *Physical Channels and Modulation*, 3GPP Standard, Rev. 10.0.0, 2011.

[7] L. Yang, G. Ren, B. Yang, and Z. Qiu, "Fast Time-Varying Channel Estimation Technique for LTE Uplink in HST Environment," *IEEE Trans. Vehicular Technology*, vol. 61, no. 9, Nov. 2012.

[8] G. Li, H. Yang, L. Cai and L. Gui, "A Low-complexity Equalization Technique for OFDM System in Time-Variant Multipath Channels," in *Proc. IEEE on vehicular technology conference (VTC 2003-Fall)*, Vol. 4, pp. 2466-2470, Oct. 2003.

[9] C. Ma, S. Liu and C. Huang, "Low-Complexity ICI Suppression Methods Utilizing Cyclic Prefix for OFDM Systems in High-Mobility Fading Channels," *IEEE Trans. Vehicular Technology*, vol. 63, no. 2, pp. 718-730, Feb. 2014.

[10] P. Reangsuntea, M. Hourai, P. Boonsrimuang, K. Mori, and H. Kobayashi, "Iterative based ML Demodulation Method for OFDM Signal under Higher Mobile Environments," in *Proc. IEEE on vehicular technology conference (VTC-Spring)*, pp. 1-6, May 2015.

[11] P. Reangsuntea, P. Boonsrimuang, K. Mori, and H. Kobayashi, "Time domain equalization method for DFTS-OFDM signal without GI under high mobile environments," in *Proc. of 12nd International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, pp. 1-6, June 2015.

[12] A. Reangsuntea, P. Reangsuntea, P. Boonsrimuang, K. Mori, and H. Kobayashi, "Iterative based Time Domain Equalization Method for DFTS-OFDM under Highly Mobile Environments," in *Proc. of 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommuni- cations and Information Technology (ECTI-CON)*, June 2016.

[13] C. Vulk, "Iterative Solution methods," *Delft Institute of Applied Mathematics*, Netherlands, 2012.

[14] Andrew J. Viterbi, "Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm," *IEEE Transactions on Information Theory*, Volume IT-13, pp. 260-269, Apr. 1967.

Pongsathorn Reangsuntea received the B.Eng and M.Eng degrees in Telecommunications Engineering from King Mongkuts Institute of Technology Ladkrabang (KMITL), Thailand in 2010, and 2012, respectively. He is currently working toward the Ph.D. degree at Mie University, Japan. His research interests include high-speed mobile communications, and wireless LAN systems. He received IEEE VTS Japan 2015 Young Researcher's Encouragement Award in VTC2015-Spring, Glasgow, UK, May 2015.

Arttapol Reangsuntea received the B.Eng degree in Instrumentations Engineering from King Mongkuts Institute of Technology Ladkrabang (KMITL), Thailand in 2012. He is currently pursuing his M.Eng degree in Telecommunications Engineering at the same place. His research interests include next generation of mobile communications, and wireless LAN systems.

Pisit Boonsrimuang received the B.Eng, M.Eng, and Doctor degrees in Telecommunications Engineering, in 1997, 2000, and 2007, respectively. He is currently an associate professor at the King Mongkut's Institute of Technology Ladkrabang (KMITL), Thailand. His research interests include transmission techniques for future multimedia wireless LAN systems and next generation of mobile communication systems. He received the Student Award of Young Research's Encouragement Award from IEICE Tokai branch, and Doctor Thesis Award for information technology from Nation Research Council of Thailand (NRCT) in 2005 and 2008, respectively.

Kazuo Mori received the B.E. degree in computer engineering from Nagoya Institute of Technology, Japan, in 1986 and received the Ph.D. degree in information electronics engineering from Nagoya University, Japan in 2000. In 1986, He joined the Hyper-media Research Center, SANYO Electric Co., Ltd. and was engaged in research and development on telecommunication systems. From 1995 to 2000, he was a research engineer at YRP Mobile Telecommunications Key Technology Research Laboratories Co., Ltd., where he was engaged in research on mobile communication systems. Since August 2000 he has been with Mie University, Japan, where he is currently a Professor in the division of Electrical and Electronic Engineering, Graduate School of Engineering. During 2005–2006, he was a Visiting Research Fellow at King's College London, UK. His research interests include mobile communication systems, wireless sensor networks, radio resource management, and teletraffic evaluation. He received the Excellent Paper Award from IEICE, Japan in 2002. Dr. Mori is a member of IEEE.

Hideo Kobayashi received the B.E., M.E., and Ph.D in 1975, 1977 and 1989, respectively from Tohoku University in Japan. He joined KDD in 1977, and engaged in the research and development of digital fixed satellite and mobile satellite communications systems. From 1988 to 1990, he was with International Maritime Satellite Organization (INMARSAT) at its London Headquarters as the Technical Assignee and involved in the design and development of Inmarsat-M communications system. From 1997 to 1998, he was with the NEC ICO project office as the technical consultant at its London office and involved in the development of the ICO personal satellite communications system. Since 1998 he has been the Professor of Graduate School of Engineering, Mie University in Japan. His current research interests include mobile communications and wireless LAN systems. Dr. Kobayashi is a member of IEEE.

Quantification of Valve Stiction using Particle Swarm Optimisation with Linear Decrease Inertia Weight

Pongsurachat Aksornsri[1] and **Sarawan Wongsa**[2], Non-members

ABSTRACT

Valve stiction is one of the most common problems on industrial process control loops. The detection and quantification of valve stiction in control loops is therefore important to ensure the high quality of the products and maintain the reliable performance of control loops. This paper presents an algorithm for quantifying valve stiction in control loop based on linear decrease inertia weight particle swarm optimisation to obtain more accurate estimates of stiction parameters. The amount of stiction present in the valve is estimated by identifying parameters of Kano model which is a two-parameter data-driven stiction modelling based on the parallelogram of MV-PV phase plot. Simulation results have demonstrated the efficacy of this algorithm in valve stiction quantification and also its robustness to oscillations due to inappropriate controller tuning and external disturbances. Results are confirmed by application to real process industrial data.

Keywords: Valve Stiction Quantification, Particle Swarm Optimisation, Linear Decrease Inertia Weight, Kano Model

1. INTRODUCTION

Valve stiction problem has been known to be one of the main causes of performance deterioration in control loops. Stiction induces oscillations in process variables and cannot be eliminated by controller detuning. Therefore, detecting and quantifying this valve problem is essential to identify the sticky valves so that they can be isolated and repaired before degrading the performance of the control loop and affecting the product quality [1, 2].

In the literature, there are a significant number of methodologies to detect and quantify the valve stiction [3], including the nonlinear analysis [1], model-based methods [4], and the shape analysis [5, 6]. The methods belonging to the nonlinear analysis utilises the fact that the presence of stiction in a control valve introduces nonlinearity in the control loop and often produces non-Gaussian time series. Therefore, the very first step of detecting valve stiction is to analyse whether there is nonlinearity and non-Gaussian data in the control loop. The most-used techniques for detection of valve or process nonlinearity are higher-order statistical methods. In Choudhury et al. [7], a non-Gaussianity index (NGI) and a nonliearity index (NLI) have been defined using the bicoherence of the signal to quantify the size of the non-Gaussianity and nonlinearity in control loops. However, in order to detecting valve stiction the nonlinearity detection technique has to be combined with other methods, typically with ellipse fitting [1].

The shape-based methods have been shown to be reliable for different types of control loops including flow control and level control loops [5, 6]. The shape-based methods use only routine operation data for detecting stiction. The shape is considered from the relationship between controller output (OP) and manipulated variable (MV) in the two-dimensional space. In practice, flow rate is used as MV instead of the valve position if MV is not available. When stiction occurs, it produces a special shape in the phase plot of OP and MV. Based on the obtained shape, a pre-determined stiction metric is calculated to quantify the valve stiction. In Kano et al. [5], stiction is detected by detecting a parallelogram shown in Fig. 1. When stiction occurs, MV stays constant while OP decreases or increases. The degree of stiction can be evaluated by taking the length where MV stays constant into account.

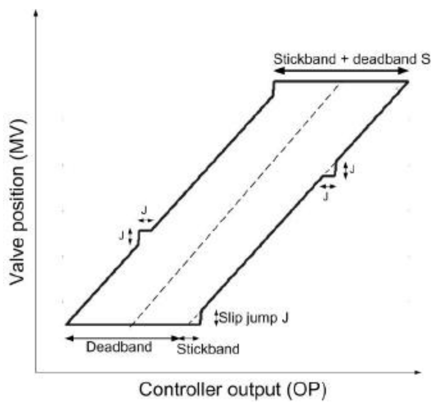


Fig.1: *Typical MV-OP characteristic of a sticky valve.*

Manuscript received on August 15, 2016 ; revised on March 11, 2017.

Final manuscript received on March 25, 2017.

[1,2] The authors are with Control Systems and Instrumentation Engineering, King Mongkut's University of Technology Thonburi, Bangkok, Thailand, Email: jummimo@gmail.com and sarawan.won@kmutt.ac.th

The main difficulty of using stiction metrics proposed in [5] in practice is that there are several predefined parameters that need to be selected, such as the thresholds for the difference between the maximum and the minimum of OP and MV when stiction occurs. However, there is no systematic way to select appropriate threshold values presented in their work. Moreover, the proposed metrics are unscaled. This makes it hard to gauge how strong the stiction is after the stiction size is quantified in the form of these metrics. In order to address this problem, it might be better to diagnose the valve stiction by directly identifying the stiction parameters which are more physically meaningful.

In this work, we propose a stiction quantification method that incorporates the stiction parameter estimation by particle swarm optimation (PSO) with the two-parameter data-driven stiction modelling. We apply the concept of linear decrease inertia weight in PSO to obtain more accurate estimates. As the identified parameters are the direct measure of the amount of valve stiction, they intuitively provide the interpretation of how strong the stiction is. We show through simulation how the proposed stiction quantification can be carried out and that the algorithm is not affected by controller tuning.

2. STICTION MODELLING & QUANTIFICATION

Consider a typical valve-controlled loop in Fig. 2, where a valve is included between the process and the controller. The controller compares the process variable (PV), for example the level, flow rate, and pressure, to the desired process condition or the set point (SP) and sends the controller output (OP) signal to correctively adjust the manipulated variable (MV) which is the valve travel position in this case (units of % opening), such that the PV well tracks the SP.

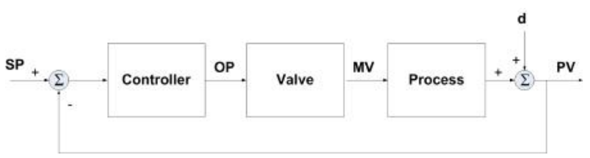


Fig.2: *Typical scheme of feedback valve-controlled loop. (SP: set point; OP: controller output; MV: manipulated variable/valve position; PV: process variable.*

A normal valve has the linear relationship between MV and OP. However, in the presence of stiction the relationship is changed to a nonlinear shape as shown in Fig. 3. Stiction describes the situation where the valve's stem is sticking when small changes are attempted. It occurs when the static friction exceeds the dynamic friction inside the valve [5]. As a consequence, the valve position cannot be changed until OP overcomes the static friction. In presence of stiction, the movement of the position becomes unsmooth and jumpy.

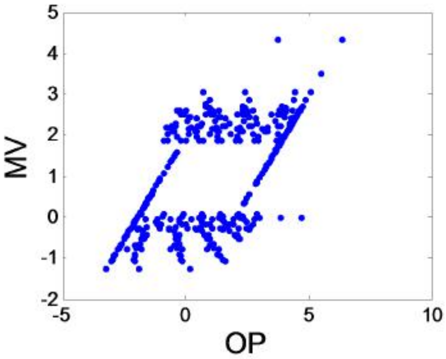


Fig.3: *Typical input-output behavior of a sticky valve.*

2.1 Stiction Modelling

In the literature, there are 2 major categories of stiction modelling: the physically-based modelling and the data-driven modelling. The physically-based models attempt to describe the friction phenomenon that causes the valve stiction by using the balance of forces and Newton's second law of motion. Examples are the model equations that describe the sticking or sliding body [8] and dynamic friction model proposed in [9]. For more details, see [2] and references therein. A detailed physically-based stiction model has a major problem to be applied in a real plant because it requires the knowledge of several critical parameters that are difficult to estimate.

Instead of relying on the first principles to model the stiction, data-driven stiction models use extensive collection of data to establish the detection and quantification of the valve stiction. The data-driven modelling requires only a few parameters to identify and has become more favorable in the literature in the recent years. Some of the data-driven models need only one parameter to be identify, such as the stiction model proposed by Stenman [4]. The problem with the one-parameter data-driven model however is that it cannot capture some input-output behaviours, such as the phase plot, of the real sticky valve during the fast stoke test [1].

The two parameter data-driven model has been proposed to more accurately capture the dynamics of the valve stiction. The Choudhury model [10] and the Kano model [5] require two parameters: the J and S parameters which represent the size of the stem slip and the combination of the stickband and deadband, respectively, while the He model [6] describes the valve stiction through the parameters f_S (static friction band) and f_D (kinetic friction band).

Garcia [11] has performed tests according to ISA standards on several different valve stiction models, including the Choudhury, the Kano, and the He models, under different input signals and with valve with different friction coefficients. He demonstrated that among these three models, the Kano model was the only model that could represent the expected stick-slip phenomenon of the sticky valve. The Kano model also has other advantages, including its ability to cope with both the stochastic and deterministic inputs and to change the degree of stiction according to the direction of the valve movement [5]. Therefore in this work, the Kano model is used to describe and simulate the stiction nonlinearity.

The Kano model describes the valve stiction using the parallelogram of the MV-OP plot, shown in Fig. 3. The deadband and stickband represent the behaviour of the valve when it is not moving through the controller output keeps changing. The magnitude of the deadband and stickband is estimated as the S parameter. The slip jump J presents the abrupt release due to the conversion of potential energy stored in the actuator to kinetic energy.

The magnitudes of S and J determine the characteristics of oscillation caused by valve stiction. The deadband provokes oscillations and reduces performance of the control loops. The magnitudes of J are also crucial to determine the amplitudes and frequencies of limit cycles [1]. Stiction is measured in percent of the controller output necessary to move the valve stem. Generally, 1% of stiction is considered enough to cause performance problems [3].

The algorithm of Kano model can be presented in a flowchart shown in Fig. 4 [5]. The input u and output y of this model are the OP and MV, respectively. The variable $stp = \{0, 1\}$ represents the moving ($stp = 1$) or resting states ($stp = 0$) of the stem, u_s represents the controller output at the moment the valve state changes from moving to resting, and $d = \pm 1$ indicates the direction of frictional force. When the valve stops or changes its direction while its state is moving, us is updated and changes its state to the resting state. Then, in the resting state the valve will change its state to moving if i) the valve changes its direction and overcomes the maximum static friction and/or ii) the valve moves in the similar direction and overcomes friction. After the valve changes its state to moving, the valve position is updated via the following equation:

$$y(t) = u(t) - \frac{d(S-J)}{2} \qquad (1)$$

2.2 Stiction Detection Method based on Parallelogram

The stiction detection algorithm, proposed by Kano et al. [5], uses the parallelogram in Fig. 1 to consider the sections where the valve does not change even though the controller output changes.

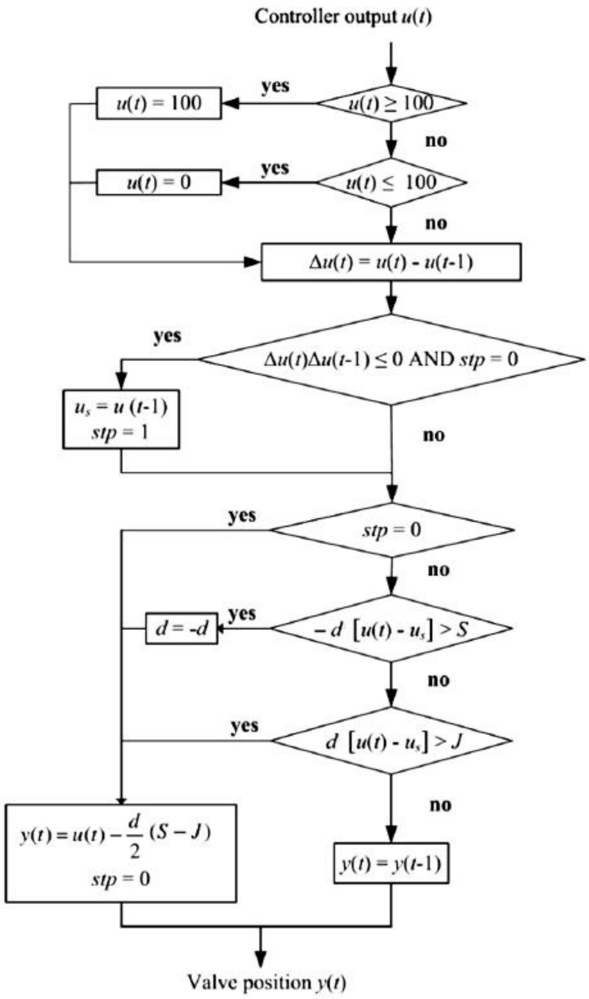


Fig.4: *Flowchart for the Kano model [5].*

The longer such sections are, the stronger the stiction is. The possibility of stiction is estimated as a ratio between the total length of intervals when stiction occurs to total length of all intervals. The stiction size is quantified by calculating the mean of the difference between the maximum and the minimum of the controller output (defined as $\tilde{u}$) when stiction occurs.

Even though this algorithm has been illustrated its success in stiction detection and quantification via both simulation studies and real chemical processes, the main problems are that there is no information on how to optimally tune several parameters, including the thresholds ε, ε_u and ε_y, and the resulting indicator σ, which quantifies the degree of stiction, has no direct connection to the parallelogram. Therefore, it is hard to gauge how strong the stiction is once the parameter σ is obtained.

To solve these limitations we propose a particle swarm optimation-based technique to directly estimate the stiction parameters J and S which have a more direct interpretation to the valve stiction.

3. PROPOSED METHOD

Particle swarm optimisation (PSO) is a stochastic population-based optimisation method, motivated by social interaction behaviour of bird flocking. The advantages of PSO are its requirements of a few parameters to adjust and easiness in implementation. Moreover, it does not require linearity in the parameters that makes it suitable to the identification nonlinearities properties [12].

In the basic PSO algorithm, m particles are placed in the n-dimensional search space. The current position of each particle represents the potential solution of the problem evaluated by a pre-defined fitness function. In every iteration, each particle determines its movement according to the history of its own current and best locations, and the best location obtained so far by any particle in the population (global best PSO). The equations for updating velocity and position of each particle are:

$$v_{id} = v_{id} + c_1 r_1 (p_{id} - x_{id}) + c_2 r_2 (p_{gd} - x_{id}) \quad (2)$$

$$x_{id} = x_{id} + v_{id} \quad (3)$$

where

$d = 1, 2, \ldots, n$ represents the dimension.

$i = 1, 2, \ldots, m$ represents the particle index, where m is the size of the swarm.

g is the index of the best particle in the swarm. c_1 and c_2 are constants, called cognitive and social scaling parameters, respectively.

r_1 and r_2 are random numbers drawn from a uniform distribution from 0 to 1.

$\boldsymbol{x_i}=(x_{i1}, x_{i2}, \ldots, x_{in})^T$ represents the position vector of the i particle.

$\boldsymbol{v_i}=(v_{i1}, v_{i2}, \ldots, v_{in})^T$ represents the velocity vector of the i particle.

$\boldsymbol{p_i}=(p_{i1}, p_{i2}, \ldots, p_{in})^T$ represents the best previous position of the i particle.

The concept of the inertia weight was introduced in 1998 by Shi and Eberhart [16] and the proposed velocity update equation is:

$$v_{id} = w v_{id} + c_1 r_1 (p_{id} - x_{id}) + c_2 r_2 (p_{gd} - x_{id}) \quad (4)$$

where w is the inertia weight.
The inertia weight is added to balance global exploration and local exploitation of the searching process. A large w facilitates a global search while a small one facilitates a local search. There have been a number of strategies proposed for adjusting the value of the inertia weight during a course of run, such as adaptive inertia weight strategy [13], chaotic inertia weight [14], and linearly decreasing strategy [15, 16]. According to the results of the comparative study of 15 different inertia weight strategies over five optimisation problems presented in [17], the linear decreasing inertia weight produced the best performance in terms of the minimum error in comparison to other methods.

In the linear decreasing inertia weight, the value of w is linearly decreased from an initial value (w_{max}) to a final value (w_{min}) according to the following equation:

$$w(iter) = \frac{iter_{max} - iter}{iter_{max}}(w_{max} - w_{min}) + w_{min} \quad (5)$$

where $iter$ is the current iteration of the algorithm and $iter_{max}$ is the maximum number of iterations allowed.

In this work, we use the PSO with the linear decreasing inertia weight to estimate the stiction parameters, J and S, of the control valve. The proposed framework is illustrated in Fig. 5.

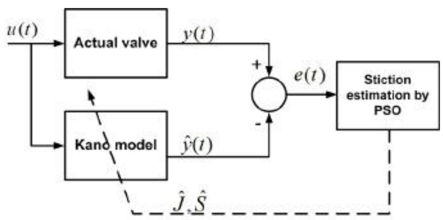


Fig.5: *PSO based parameter estimation procedure.*

The algorithms in Fig. 5 is summarised as follows:

i) Set up all PSO parameters and initialise the inertia weight to w_{max}.

ii) Initialise a population array of m particles with random positions and velocities on two dimensions, representing the parameters J and S of the stiction, in the search space.

iii) The fitness function for each particle in the initial population is evaluated. We select the mean squared error (MSE) as the fitness function for determining how well the estimates fit the system. The MSE_i of estimation for particle i is defined as

$$MSE_i = \frac{\boldsymbol{e}_i^T \boldsymbol{e}_i}{N} \quad (6)$$

where $\boldsymbol{e}_i \in R^N = \boldsymbol{y_i} - \boldsymbol{\hat{y}_i}$ is error vector of the i particle, $\boldsymbol{\hat{y}_i} \in R^N = \{\hat{y}_i(1), \hat{y}_i(2), \ldots, \hat{y}_i(N)\}$ is the vector of estimated valve positions obtained from the Kano model, $\boldsymbol{y}_i \in R^N = \{y_i(1), y_i(2), \ldots, y_i(N)\}$ and is the vector of measured valve positions, and N is the number of input-output datapoints used in the identification. p_{id} is set to each initial searching point. The initial best evaluated value among p_{id} values is set to p_g.

iv) Update the velocity and position of each particle using (4) and (3).

v) Search with new position and the fitness functions are calculated. If the fitness function of each particle is better than the previous p_{id} ,the value is set to p_{id}. If the best p_{id} is better than p_{gd}, the value is set to p_{gd}. All p_{gd} are stored as the current estimates of J and S.

vi) Update the inertia weight according to (5).

vii) Repeat from step iv) until the maximum number of iteration is exceeded.

4. SIMULATION EXAMPLES

The following experiments demonstrate the optimal configuration of the proposed stiction quantification method and its performance. The flow control system proposed in [5] is used to represent the industrial control loop. Fig. 6 shows the block diagram of the system.

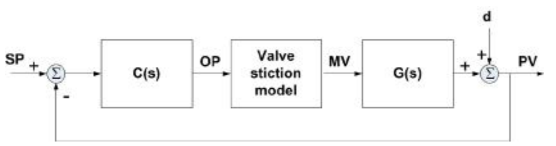


Fig.6: *Block diagram of the flow control system.*

The process transfer function of flow $G(s)$ is given by

$$G(s) = \frac{1}{0.2s + 1} \tag{7}$$

The controller $C(s)$ is the proportional-integral (PI) controller which is implemented in the following form:

$$C(s) = K_p \left(1 + \frac{1}{\tau_i s}\right) \tag{8}$$

The proportional gain (K_p) and reset time (τ_i) are set to 0.5 and 0.3 min, respectively. The sampling interval for flow control is 0.5 min.

Table 1: *Stiction Parameters used to Evaluate the Proposed Method.*

Case	J(%)	S(%)
No stiction	0	0
Weak stiction	0.3	1
Strong stiction	1	5

Table 1 summarises all three cases investigated in this study, including the normal case where there is no stiction, the weak stiction where $J = 0.3\%$ and $S = 1\%$ and the strong stiction where $J = 1\%$ and $S = 5\%$. For each pair of parameter $\{J, S\}$, we generate the input-output datapoints by simulation to represent the actual data. These data are then used in our stiction quatification procedure to determine the estimates of $\{J, S\}$ in the performance evaluation process.

It should be noted that generally there is no particular rule of thumb to strictly pinpoint how much stiction would represent each state of valve stiction. In practice, we could consider a valve to be *normal* even when it exhibits some behaviour of stiction if such magnitude of stiction does not greatly deteriorate the loop performance. However in order to be able to simulate three different states of the valve, in this simulation we selected the magnitudes of J and S similar to those in [5] to represent the normal, weak stiction, and strong stiction states. For each stiction case, a band-limited white noise is forced into the control loop as the setpoint to simulate the control loop and 100 sampling points of MV-OP data are used in our algorithm to quantify the stiction parameters J and S.

Table 2: *Parameters of LDIW-PSO Stiction Estimation.*

Parameter	Value
Population size	9, 25, 49
Number of iterations	40
Initial and final inertia weight	{0.9, 0.4}
c_1 and c_2	2

The parameters of the LDIW-PSO (linear decrease inertia weight-particle swarm optimisation) are listed in Table 2. The initial and final values of the inertia weight are set to 0.9 and 0.4, respectively, according to the suggestion in [16]. The coefficients c_1 and c_2 are set to the recommended values of 2. Note that the estimated J and S are constrained to be positive, meaning that when negative estimates are obtained, they are set to zero.

In the first experiment, we vary the number of particles or the population size to 9, 25, and 49 to investigate the estimation performance and use it as a guideline to select the appropriate population size for this problem. For each siction case, 50 trials are carried out and the average MSE of parameter estimation are calculated. Note that this MSE is different from the one we used for calculating the fitness function of PSO in (6) and the MSE of the stiction estimation is defined as:

$$MSE_{stiction} = \frac{1}{50}\sum_{i=1}^{50}(\theta_i - \hat{\theta}_i)^2 \tag{9}$$

where θ_i and $\hat{\theta}_i$ represent the actual values and the estimates of the parameters $\{J, S\}$ at the trial number i, respectively.

From the results of the average MSEs shown in Figs. 7 - 9, it can be clearly seen that the population size of 25 is sufficient for this problem because its MSE values are greatly reduced when compared with the MSE of the population size of 9 and by increasing it to be 49, the performances are not significantly improved, or for the normal case are even degraded. So the population size of 25 is used for further analyses.

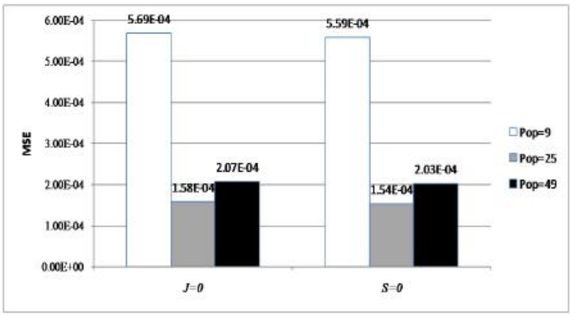


Fig.7: *MSE in each population size for normal case.*

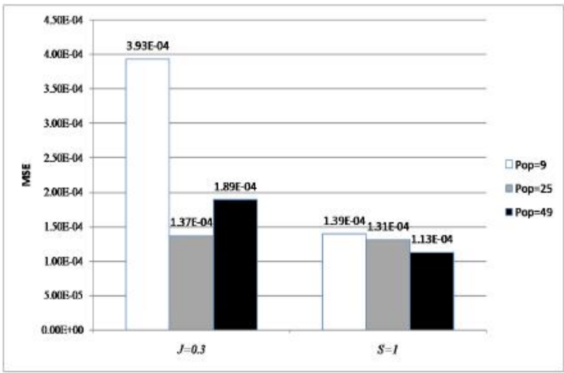


Fig.8: *MSE in each population size for weak stiction case. (a) J estimates (b) S estimates.*

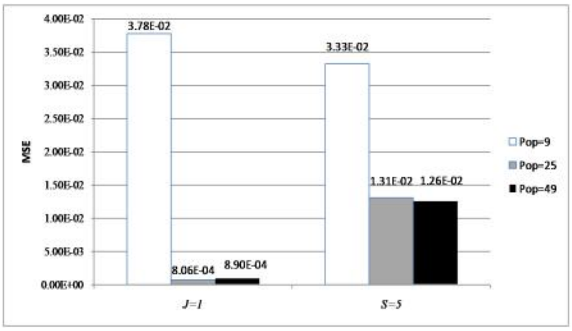


Fig.9: *MSE in each population size for strong stiction case. (a) J estimates (b) S estimates.*

In the next experiment, the LDIW-PSO is compared with the conventional PSO which uses the constant inertia weight. Two constant inertia weights, w = 0.9 and 0.4, corresponding to the initial and final values of the inertia weight of LDIW-PSO are selected for conventional PSO and the results are compared based on the MSE of the stiction estimation in (9). All PSO variants use the population of 25. Table 3 summarises the results in terms of the MSE from 50 trials. The best result for each test case is shown in bold. In 5 out of 6 stiction parameter values, LDIW-PSO performs as the best approach. Even in the case of $S = 5$ where the LDIW-PSO does not perform the best, but its obtained MSE is just slightly smaller than the best one. So it may be concluded that overall the LDIW-PSO outperforms the constant inertia weight scheme.

Table 3: *MSE for Each Stiction Case Versus the PSO Variants.*

Case	MSE		
	Fixed w=0.4	*Fixed w=0.9*	*LDIW-PSO*
J=0	1.99E-01	5.91E-04	**1.58E-04**
S=0	1.94E-01	5.32E-04	**1.54E-04**
J=0.3	5.25E-01	1.57E-02	**1.37E-04**
S=1	1.94E-01	5.32E-04	**1.54E-04**
J=1	9.84E-04	6.70E-03	**8.06E-04**
S=5	**1.23E-02**	**1.23E-02**	1.31E-02

In the last experiment, we investigate how the proposed technique performs in detecting oscillation other than stiction. Sources of oscillation commonly present in industrial processes include incorrect controller tuning and periodic external disturbances.

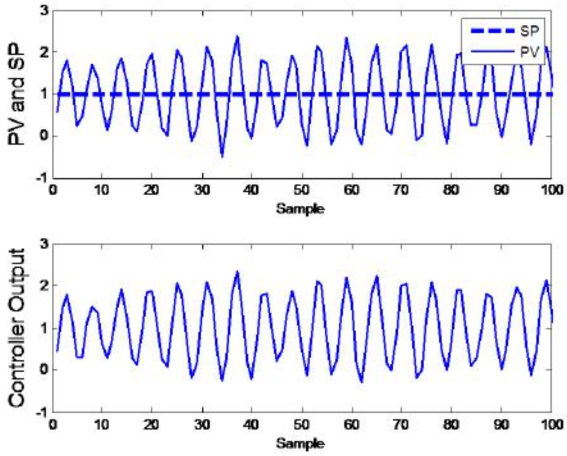


Fig.10: *Oscillatory control loop simulated with aggressive tuning without stiction for flow control loop.*

A very common scenario of inappropriate controller tuning is when an aggressively tuned controller is applied. In such scenario, the stiction quantification algorithm should be able to confirm that the cause of the oscillation is not valve stiction. We simulate this scenario by deliberately apply an aggressive tuning specification to the flow control loop by setting the integral time to 0.26 min and keeping the proportional gain to the value used in the previous simulations. The measurement noise with variance of 0.01 (SNR around 10) is injected to the measured flow to simulate corrupted data. This setting causes the loop to be oscillating as shown in Fig. 10 even when the valve is free from stiction.

To simulate the loop oscillation due to an external disturbance, a sinusoidal signal with amplitude of 0.1 and period of 20 min is added to the measured flow. Even when the stiction is absent, the injected disturbance causes a regular oscillation to the process variable as illustrated in Fig. 11.

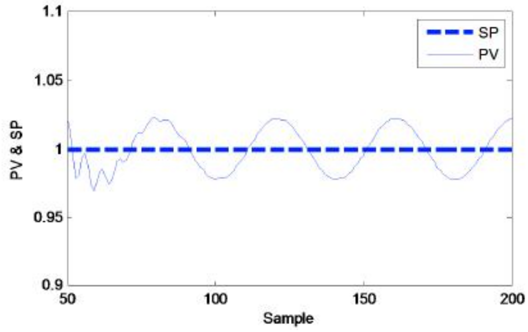


Fig.11: *Oscillatory control loop simulated with external sinusoidal disturbance without stiction for flow control loop.*

To evaluate the robustness of the proposed methodology, the LDIW-PSO is applied to different sources of oscillation for the cases of normal valve, weak stiction, and strong stiction. The values of J and S for all simulated stiction levels are similar to the ones that we used in the previous experiments.

The average estimates of the stiction parameters over 50 runs are summarised in Table 4. It can be seen that for all oscillation cases, the method successfully estimates valve stiction with high accurate quantification. Moreover, in the case of normal valve the proposed method is robust to non-stiction oscillations as the obtained values are zero for both J and S, indicating that it can quantify the stiction size correctly.

Table 4: *Average Estimates and Standard Deviations(in parentheses) of Different Stiction Levels for Different Sources of Loop Oscillation.*

Case	Actual J	Actual S	Est.J	Est.S
Normal + aggressive tuning	0	0	0(0)	0(0)
Weak stiction + aggressive tuning	0.3	1	0.3105 (0.0237)	1.0105 (0.0237)
Strong stiction + aggressive tuning	1	5	1.0045 (0.0097)	5.0045 (0.0097)
Normal + sinusoidal disturbance	0	0	0(0)	0(0)
Weak stiction + sinusoidal disturbance	0.3	1	0.2998 (0.0042)	0.9998 (0.0042)
Stroing stiction + sinusoidal	1	5	0.9676 (0.0470)	4.9675 (0.0470)

Fig. 12 shows the MV-OP plot of the prediction from the model for normal valve with aggressive tuning as an example. From the figure, we can clearly observe a linear dependence between the estimated manipulated variable and the controller output and there is no sign of any stiction pattern as it should be.

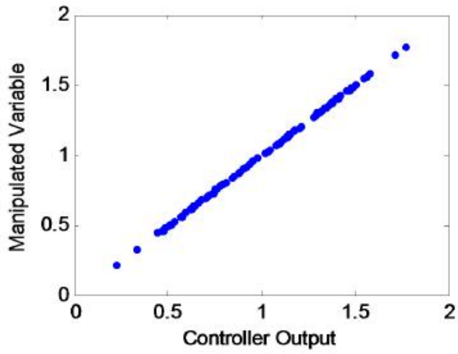


Fig.12: *MV-OP plot from the estimated stiction model of the normal valve.*

5. APPLICATION TO INDUSTRIAL DATA

The objective of this section is to evaluate the performance of the proposed method when applied to real process industrial valves. A set of data, collected from five valves and provided by a large-scale petrochemical plant in Thailand, were analysed. Valves EX1 - EX4 in Table 5 were selected from normally-operated control loops. Bear in mind that the normal valves considered in this study are not completely free from stiction. Instead, we accept a valve as normal as long as it gives a satisfactory control loop performance. For example, the plot of MV-OP of valve EX1 which is considered as normal clearly shows a non-zero deadband plus side slip. However, as this valve could still perform its control duty with excellent performance it was regarded as normal valve in this study.

Table 5: *Industrial Control Valves in This Study.*

Valve	Size	Valve Type	Control Loop
EX1	1”	Globe	Tank level
EX2	1.5”	Globe	Tank level
EX3	4”	Globe	Gas flow
EX4	10”	Butterfly	Level of a distillation column
EX5	14”	Ball	Gas flow

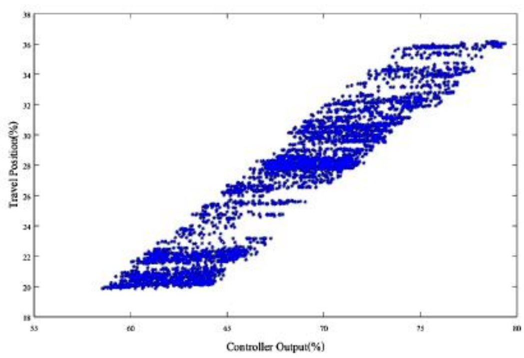
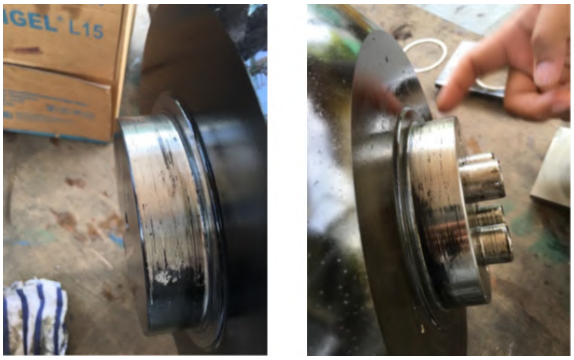


Fig.13: *MV-OP plot of valve EX1.*

For valve EX5, it was taken from the control loop which exhibited severe oscillations. We suspected that the oscillations in this loop were due to the sticky behaviour of this valve. To confirm the condition of valve EX5, it was completely disassembled and closely examined during the maintenance period. Apparently, as seen in Fig. 14, the clear damage on the surfaces of upper and lower trunnions of the valve was visually evident. Therefore, it can be confirmed that the loop oscillations were indeed a result of stiction in valve EX5.



(a) (b)

Fig.14: *Damage on the surfaces of (a) upper trunnion (b) lower trunnion of the ball valve EX5.*

For each loop, the controller output (OP) and the valve stem travel position (MV) were collected with the sampling interval of 1 second and scaled between 0 to 100%. To illustrate the ability of the proposed method to perform the online valve stiction analysis, we divided the time-series data into sub-windows with window size of 1 hour, corresponding to 3600 data points in each window. 96 samples were extracted from the normal valves (EX1-EX4) and 29 samples from the sticky valve. 68 samples of normal data and 20 samples of abnormal data were randomly selected as training data, while the remaining data were used for testing. The slip jump (J) and deadband plus slip band (S) of all samples are estimated using the proposed method. To perform the online stiction detection, the estimated J and S are compared to a set of predefined thresholds. If any of the estimated J and S is greater than its threshold, the sample is classified as faulty and the valve is stuck. Once the valve is defined as a sticky valve, its stiction level can readily be classified into weak or strong stictions by considering the magnitudes of the estimated J and S. Here, we used two sets of thresholds in order to indicate the stiction levels of the valve as described below:

$$\text{Stiction index}(i)=\begin{cases}\text{normal,if } \hat{J}_i<J_1 \ \& \ \hat{S}_i<S_1\\ \text{weak stiction,if } J_1\leq\hat{J}_i\leq J_2 \text{ or } S_1\leq\hat{S}_i\leq S_2\\ \text{strong stiction,if } \hat{J}_i>J_2 \text{ or } \hat{S}_i>S_2\end{cases} \tag{10}$$

where Stiction index (i) is the verdict of valve stiction for data in window i, J_i and S_i are the estimated J and S of the data in window i, J_1 and $J_2 (J_2 > J_1)$ are the thresholds of slip jump for detecting weak and strong stiction respectively, and S_1 and $S_2 (S_2 > S_1)$ are the thresholds of deadband plus slip band for detecting weak and strong stiction respectively.

Obviously, the thresholds influence the robustness of this stiction detection and classification scheme. Unfortunately, from the experience of the authors, there is no single value of threshold that can guarantee to successfully detect stiction of all control loops. In this work, we obtained the thresholds empirically from the training dataset. Specifically, J_1 and S_1 were obtained from the maximum values of the estimated J and S of the normal samples in the training dataset, while J_2 and S_2 were obtained from the minimum values of the estimated J and S of the severely sticky valve in this example to represent the lower limits of strong stiction.

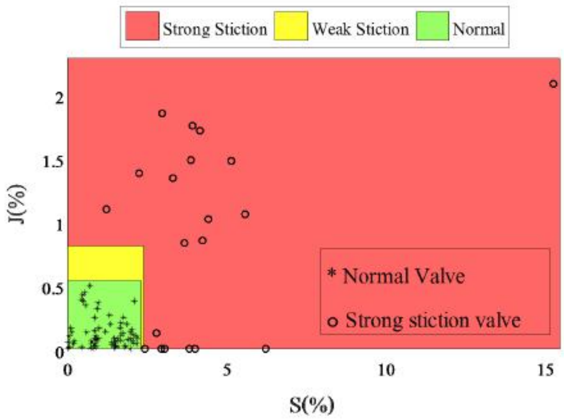


Fig.15: *Stiction condition zones and estimated J and S of the training data.*

With $\{J_1, S_1\} = \{0.53, 2.3\}$ and $\{J_2, S_2\} = \{0.8, 2.4\}$, we could plot the decision boundaries for normal, weak stiction and strong sticion as shown in Fig. 15 for the training data. To provide a convenient and quick analysis of valve stiction level, the $J - S$ plot area is divided as a green-yellow-red condition zone. The colours green, yellow, and red specify the zones for normal, weak stiction, and strong stiction. When plotting the estimates of test dataset on the $J - S$ chart in Fig. 16, we see that the stiction conditions of all samples are correctly predicted.

6. CONCLUSIONS

In this paper, a method for quantification of valve stiction has been proposed. The slip-jump J and the deadband plus stickband S could be simultaneously estimated. Stiction parameters were identified accurately through Kano model through MV-OP plot

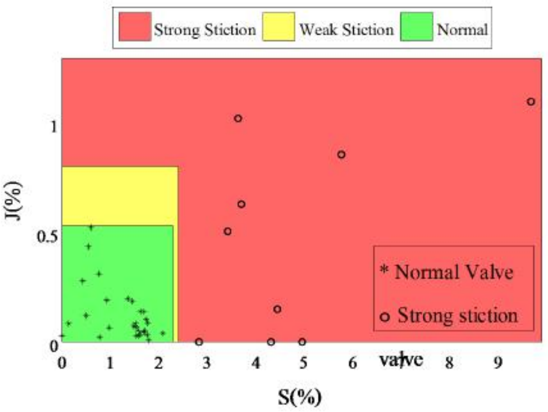


Fig.16: *Stiction condition zones and estimated J and S of the test data.*

with the use of particle swarm optimisation. The linear decrease inertia weight was suggested to be used during the update process of PSO to improve the identification convergence and accuracy. The experimental results clearly show the superiority of the LDIW-PSO over the fixed inertia weight PSO. The comparisons are made in terms of solution accuracy. In addition, the robustness of the proposed method to incorrect controller tuning and external disturbances was demonstrated. The simulations have shown that the estimated parameters obtained from this method can clearly indicate if the oscillation is caused by valve stiction or other problem. Industrial examples have demonstrated the implementation of this technique in an online fashion and confirmed the effectiveness of the proposed stiction quantification method in practice.

References

[1] M.A.A. S. Choudhury, M. Jain and S. L. Shah, "Stiction - Definition, Modelling, Detection and Quantification," *Journal of Process Control*, Vol. 18, No. 3-4, pp. 232-243, March 2008.

[2] M. Jelali, *Control Performance Management in Industrial Automation*, 1^{st} ed., UK: Springer-Verlag London, 2013, pp. 266.

[3] M. Jelali and B. Huang, *Detection and Diagnosis of Stiction in Control Loops: State of the Art and Advanced Methods*, 1^{st} ed., UK:Spinger-Verlag London, 2010.

[4] A. Stenman, F. Gustafsson, and K. Forsman, "A Segmentation-based Method for Detection of Stiction in Control Valves," *International Journal of Adaptive Control and Signal Processing*, Vol. 17, No. 7-9, pp. 625-634, September - November 2003.

[5] M. Kano, H. Maruta, H. Kugemoto, and K. Shimizu, "Practical Model and Detection Algorithm for Valve Stiction," *Proceedings of 7th IFAC DYCOPS*, Cambridge, USA, 2004.

[6] Q. P. He, J. Wang, M. Pottmann, and S. J. Qin, "A Curve Fitting Method for Detecting Valve Stiction in Oscillating Control Loop," *Industrial & Engineering Chemistry Research*, Vol. 46, No. 13, pp. 4549-4560, 2007.

[7] M.A.A.S. Choudhury, S.L. Shah, N.F. Thornhill, "Diagnosis of Poor Control Loop Performance using Higher Order Statistics," *Automatica*, Vol. 40, No. 10, pp.1719-1728, 2004.

[8] D. Karnopp, "Computer Simulation of Stick-Slip Friction in Mechanical Dynamic Systems," *Journal of Dynamic Systems, Measurement, and Control*, Vol. 107, No. 1, pp. 100-103, 1985.

[9] C. C. de Wit, H. Olsson, K. J. Åström, and P. Lischinsky, "A New Model for Control of Systems with Friction," *IEEE Transactions of Automatic Control*, Vol. 40, No 3, pp. 419-425, 1995.

[10] M. A. A. S. Choudhury, N. F. Thornhill, and S. L. Shah, "A Data- Driven Model for Valve Stiction," *Proceedings of 5th ADCHEM*, Hong Kong, China, 2004, pp. 261-266.

[11] C. Garcia, "Comparison of Friction Models Applied to a Control Valve," *Control Engineering Practice*, Vol. 16, No. 10, pp. 1231-1243, October 2008.

[12] S. Sivagamasundari and D. Sivakumar, "Estimation of Valve Stiction Using Particle Swarm Optimization," *Sensors & Transducers Journal*, Vol. 129, pp. 149-162, 2011.

[13] A. Nikabadi and M. Ebadzadeh , "Particle Swarm Optimization Algorithms with Adaptive Inertia Weight : A Survey of the State of the Art and a Novel Method," *IEEE Journal of Evolutionalry Computation*, 2008.

[14] Y. Feng, G. F. Teng, A. X. Wang, and Y. M. Yao, "Chaotic Inertia Weight in Particle Swarm Optimization," *Proceedings of 2nd International Conference of Innovative Computing, Information and Control*, Kumamoto, 2007, p. 475.

[15] J. Xin, G. Chen, and Y. Hai., "A Particle Swarm Optimizer with Multistage Linearly-Decreasing Inertia Weight," *Proceedings of International Joint Conference on Computational Sciences and Optimization*, Sanya, Hainan, 2009, pp. 505-508.

[16] Y. Shi and R. C. Eberhart, "Empirical Study of Particle Swarm Optimization," *Proceeding of the 1999 Congress on Evolutionalry Computation*, Washington, DC, 2002, pp. 1945-1950.

[17] J. C. Bansal, P. K. Singh, M. Saraswat, A. Verma, S. S. Jadon, and A. Abraham, "Inertia Weight Strategies in Particle Swarm Optimization," *Proceedings of 3rd World Congress on Nature and Biologically Inspired Computing*, Salamanca, 2011, pp.640-647.

Pongsurachat Aksornsri received the BEng degree in Electrical Engineering from Chiangmai Universtiy, 2005. He graduated with a master degree in Control Systems and Instrumentation Engineering from King Monkut's University of Technology Thonburi, Bangkok, Thailand in 2015. His research interests include process monitoring for industrial applications.

Sarawan Wongsa received the BEng degree in Control Systems and Instrumentation from King Monkut's University of Technology Thonburi, Bangkok, Thailand in 1998, and the MSc and PhD degrees in Automatic Control and Systems Engineering from Sheffield University, United Kingdom, in 2002 and 2007, respectively. She is currently an assistant professor in the Department of Control Systems and Instrumentation Engineering at King Monkut's University of Technology Thonburi. Her research interests include intelligent fault detection and diagnosis.

Probabilistic Modelling of Variation in FGMOSFET Devices

Rawid Banchuin[1] and **Roungsan Chaisricharoen**[2], Non-members

ABSTRACT

The analytical probabilistic modelling of random variation in the drain current of a Floating-Gate MOSFET (FGMOSFET) induced by manufacturing process variations has been performed. Both triode and saturation region operated FGMOSFETs have been considered. The results have been found to be very efficient since they can accurately fit the probabilistic distributions of normalized random drain current variations of the candidate triode and saturation FGMOSFETs obtained using the 0.25μm level BSIM3v3 based Monte-Carlo SPICE simulations, where the variation of the saturation FGMOSFET has been found to be more severe. These results also satisfy the goodness of fit test at a very high level of confidence and more accurately than the results of the previous probabilistic modelling attempts.

Using our results, many statistical parameters, probabilities and the objective functions, which are useful in statistical/variability aware analysis and design involving FGMOSFETs can be formulated. The impact of drain current variation upon the design trade-offs can be studied. It has been found that the occurrence of the drain current variation is absolutely certain. Moreover, the analytical probabilistic modelling and computationally efficient statistical/variability aware simulation of FGMOSFET based circuits can also be performed.

Keywords: FGMOSFET, Probabilistic Modelling, Saturation Region, Statistical/variability Aware Analysis and Design, Triode Region

1. INTRODUCTION

FGMOSFET devices have been widely utilized in various analog/digital circuits and systems [1-14]. As with conventional MOSFETs, the performance of FGMOSFET based circuits and systems are deteriorated by circuit level random variations induced by variations in manufacturing processes, random dopant fluctuation, lithographic variation, variation in contact resistance and line edge roughness, among other factors. To handle these difficulties, the design of many FGMOSFET based circuits and systems has been done using the concept of statistical/variability aware design as proposed in numerous studies [4], [15-21].

The key circuit level parameter of both ordinary MOSFET and FGMOSFET devices is their drain current, I_D, because the variations in other parameters can be conveniently determined using the variation in I_D. Using ΔI_D as a basis, analytical modelling of the ΔI_D of a single transistor has been done in previous studies [22-23]. The obtained results are generic as they are applicable to any circuit and system. Modelling of the variation in any circuit level parameter can be done for statistical/variability aware analysis and design. Unfortunately, generic single transistor based studies have focused on ordinary MOSFET devices. Alternatively, most of the previous studies on the variability of FGMOSFET devices [24-32] have been done in a case by case manner focusing only on corresponding FGMOSFET based circuits. So, these results are not generic as they are applicable only to specific circuits. Later, analytical modelling of ΔI_D for a single FGMOSFET device was done [33-34]. However, the resulting model in [33] was expressed in terms of the variance of ΔI_D, which gives no information other than the magnitude of ΔI_D. Other useful information was not expressed, e.g., the probability of obtaining either a certain value or interval of ΔI_D and statistical parameters such as the mean and second moment. The model proposed in [34] is in terms of a probability density function of the normalized value of ΔI_D. This is able to yield much useful information apart from the variance after applying a probability density function. However, only a FGMOSFET operated in the saturation region has been considered while the short channel effects have been overlooked.

Hence, analytical probabilistic modelling of ΔI_D for a single FGMOSFET was performed in this research. The results were generated using a probability density function of $\Delta I_D/I_D$. This was chosen as it is a convenient dimensionless quantity. Unlike previous work [34], both triode and saturation region operated FGMOSFET devices were considered and their short channel effects were taken into account. Therefore, this work is more complete, detailed and accurate that earlier studies. Compared to the model proposed in [33], which deals only with variances, the

---

Manuscript received on July 27, 2016 ; revised on March 9, 2017.

Final manuscript received on April 2, 2017.

[1] The author is with Faculty of Engineering and Graduated School of Information Technology, Siam University, Bangkok, Thailand., E-mail: rawid_b@yahoo.com

[2] The author is with School of Information Technology, Mea Fah Luang University, Chiangrai, Thailand., E-mail: ch.roungsan@gmail.com

resulting models of this work can give more useful information as it is in terms of probability density functions. Many useful statistical parameters and probabilities have been formulated using our model, emphasizing this point. The modelling process of this work has taken variations in the manufacturing process into account regarding a single FGMOSFET. Therefore, the resulting models are generic and can be extensively applied to any FGMOSFET associated circuit or system. These models are also very accurate. They can fit the probabilistic distributions of $\Delta I_D/I_D$ for the candidate triode and saturation FGMOSFETs of both N-type and P type devices obtained using Monte-Carlo SPICE simulations based on BSIM3v3 at a 0.25μm level with high accuracy. In these cases, it was found that the saturation variation of a FGMOSFET is more severe than that of a triode device. They also satisfy the goodness of fit test at a 99% confidence level. As the formerly overlooked short channel effects have been now taken into account, our probabilistic model for a saturation FGMOSFET is even more accurate than that proposed in [34].

With the statistical parameters and probabilities obtained by using our modelling results, the impact of ΔI_D upon the design trade-offs associated with FGMOSFET devices can be studied and objective functions that support statistical/variability aware analysis and design can be formulated. It has also been mathematically shown that the occurrence of ΔI_D is absolutely certain, which emphasizes the necessity of our work. Moreover, analytical probabilistic modelling and computationally efficient statistical/variability aware numerical analysis of FGMOSFET based circuits can be performed. In the subsequent section, an overview of FGMOSFET devices will be briefly given.

2. PROBLEM OF CONVENTIONAL ONE-TAP MMSE-FDE FOR DFTS-OFDM SIGNAL

FGMOSFET is a special type of MOSFET device with an additional gate that is completely isolated within the oxide, namely a floating gate [11], [34]. A cross sectional view, symbol and equivalent circuit model of an N-type FGMOSFET with N inputs is depicted in Figs. 1-3 [11], [34]. If we let $\{i\} = \{1, 2, 3, \dots, N\}$, it can be seen that Ci denotes the capacitance between any i$^{\text{th}}$ input and the floating gate. Moreover, the capacitive coupling ratio of any i$^{\text{th}}$ input (k_i) [2], [11], [34], can be defined as:

$$k_i = \frac{C_i}{\sum_{i=1}^{N} C_i} \tag{1}$$

Since I_D of the FGMOSFET in any region can be obtained by simply replacing the gate to source voltage term in the drain current equation of an ordinary MOSFET device in that region with a floating gate to source voltage, V_{FGS} can be given by:

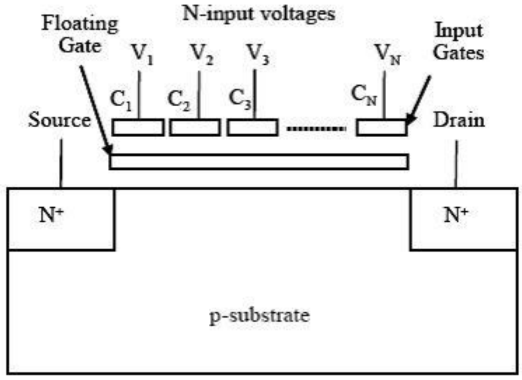


Fig.1: *A cross sectional view of an N-type N inputs FGMOSFET [11], [34].*

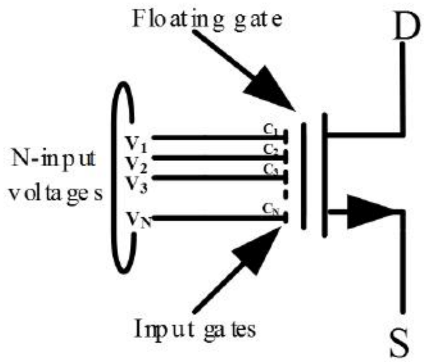


Fig.2: *A symbol of an N-type N inputs FGMOSFET [34].*

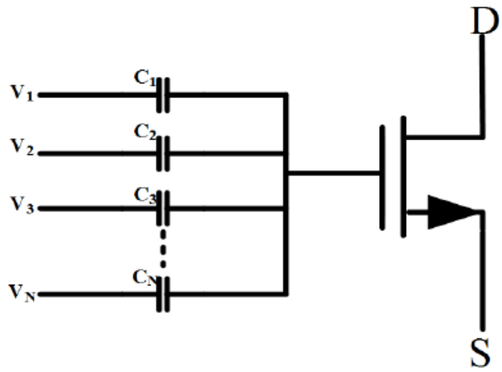


Fig.3: *An equivalent circuit model of an N-type N inputs FGMOSFET [34].*

$$V_{FGS} = \sum_{i=1}^{N} k_i V_i - V_S \tag{2}$$

where V_i and V_S denote the voltage at any i$^{\text{th}}$ input of the FGMOSFET and the source terminal voltage respectively. I_D of the triode and saturation region operated FGMOSFET with short channel effects taken into account can be respectively given as follows:

$$\begin{aligned} I_D =& \mu C_{ox}\frac{W}{L}[1-\theta(\sum_{i=1}^{N}k_iV_i - V_S - V_t)] \\ & \times [(\sum_{i=1}^{N}k_iV_i - V_S - V_t)V_{DS} - \frac{1}{2}V_{DS}^2] \end{aligned} \tag{3}$$

$$\begin{aligned} I_D =& \frac{\mu}{2}C_{ox}\frac{W}{L}[1-\theta(\sum_{i=1}^{N}k_iV_i - V_S - V_t)] \\ & \times (\sum_{i=1}^{N}k_iV_i - V_S - V_t)^2(1+\lambda V_{DS}) \end{aligned} \tag{4}$$

where $\mu, \lambda, \theta, C_{ox}, V_t, W$ and L represent the mobility of the carriers, channel length modulation coefficient, mobility degradation coefficient, gate oxide capacitance per unit area, threshold voltage, channel width and channel length, respectively. It is notable that the linearly approximated mobility degradation model [33] was used in the formulation of equations (3) and (4) for simplicity. In the subsequent section, our proposed analytical probabilistic modelling will be given.

3. THE ANALYTICAL PROBABILISTIC MODELING

Firstly, $\Delta I_D/I_D$ must be defined. By taking the manufacturing process variations into account, random fluctuations in device level parameters, e.g., μ, V_t, W and L, among others, exist and cause fluctuation in the drain current. Thus $\Delta I_D/I_D$ can be defined for any region of operation as:

$$\frac{\Delta I_D}{I_D} \triangleq \frac{I_D - I_{D,ideal}}{I_{D,ideal}} \tag{5}$$

where $I_{D,ideal}$ denotes the ideal drain current in which the physical level nonidealities have been neglected. In the following subsection, the analytical probabilistic modelling of a triode region FGMOSFET is presented.

3.1 The modelling of triode region operated FGMOSFET

It should be noted here that intrinsic manufacturing process variations are emphasized as they are directly caused by the physical limits of the device [35]. These are intrinsic to the basic technology [36]. Moreover, the key intrinsic manufacturing process variations are random dopant fluctuation and line edge roughness [36]. These induce threshold voltage variations [35], [36]. As a result, such device level random variations assume a Gaussian distribution and have been found to be dominant. Thus $\Delta I_D/I_D$ of the FGMOSFET in the triode region can be analytically given using equations (3), (5) and the physical level parameters used by [37] are as follows:

$$\begin{aligned} \frac{\Delta I_D}{I_D} =& \{\theta[1-\theta(\sum_{i=1}^{N}k_iV_i - V_S - V_t)]^{-1} - V_{DS}[(\sum_{i=1}^{N}k_iV_i - V_S - V_t)V_{DS} - \frac{1}{2}V_{DS}^2]^{-1}\} \\ & \times (V_t - qN_{sub}W_{dep}C_{inv}^{-1} - V_{FB} - 2\phi_F) \end{aligned} \tag{6}$$

where C_{inv}, N_{sub}, V_{FB}, W_{dep} and ϕ_F denote capacitance of the inversion layer, substrate dopant concentration, flat band voltage, width of the depletion layer and the Fermi potential. These are physical level parameters. Moreover, V_t now randomly fluctuates.

For deriving the probability density function of $\Delta I_D/I_D$ i.e., $f(\delta I_D/I_D)$, where $\delta I_D/I_D$ represents the corresponding sample variable, the often cited analytical model of device level random variation [37], [38] was adopted. As a result, $f(\delta I_D/I_D)$ can be given by equation (6) and the adopted model is as follows:

$$\begin{aligned} f(\frac{\delta I_D}{I_D}) =& \sqrt{\frac{3WLC_{inv}^2\{\theta[1-\theta(\sum_{i=1}^{N}k_iV_i - V_S - V_t)]^{-1} - V_{DS}[(\sum_{i=1}^{N}k_iV_i - V_S - V_t)V_{DS} - \frac{1}{2}V_{DS}^2]^{-1}\}^{-2}}{2\pi N_{sub}W_{dep}q^2}} \\ & \times \exp[-\frac{3WLC_{inv}^2\{\{\theta[1-\theta(\sum_{i=1}^{N}k_iV_i - V_S - V_t)]^{-1} - V_{DS}[(\sum_{i=1}^{N}k_iV_i - V_S - V_t)V_{DS} - \frac{1}{2}V_{DS}^2]^{-1}\}^{-1}(\delta I_D/I_D)\}^2}{2N_{sub}W_{dep}q^2}] \end{aligned} \tag{7}$$

In the next subsection, the modelling of a FGMOSFET device in the saturation region will be presented.

3.2 The modelling of saturation region operated FGMOSFET

For a FGMOSFET in the saturation region, $\Delta I_D/I_D$ can be given in a similar manner to that of a FGMOSFET in the triode region, but using equation (4) as a basis instead of (3). Thus we obtain:

$$\begin{aligned} \frac{\Delta I_D}{I_D} =& \{\theta[1-\theta(\sum_{i=1}^{N}k_iV_i - V_S - V_t)]^{-1} - 2(\sum_{i=1}^{N}k_iV_i - V_S - V_t)^{-1}\} \\ & \times (V_t - qN_{sub}W_{dep}C_{inv}^{-1} - V_{FB} - 2\phi_F) \end{aligned} \tag{8}$$

As a result, $f(\delta I_D/I_D)$ of the saturation FGMOSFET can be given by equation (8) and a similar approach to that of the triode FGMOSFET as follows:

$$f(\frac{\delta I_D}{I_D}) = \sqrt{\frac{3WLC_{inv}^2\{\theta[1-\theta(\sum_{i=1}^{N}k_iV_i - V_S - V_t)]^{-1} - 2(\sum_{i=1}^{N}k_iV_i - V_S - V_t)^{-1}\}^2}{2\pi N_{sub}W_{dep}q^2}} \tag{9}$$

$$3WLC_{inv}^2\{\{\theta[1-\theta(\sum_{i=1}^{N} k_iV_i-V_S-V_t)]^{-1}$$
$$\times \exp[-\frac{-2(\sum_{i=1}^{N} k_iV_i-V_S-V_t)^{-1}\}^{-1}(\delta I_D/I_D)\}^2}{2N_{sub}W_{dep}q^2}]$$

Since the short channel effects have now been considered, the resulting $f(\delta I_D/I_D)$ of the saturation FGMOSFET is more detailed and thus more accurate than that proposed in [34]. This will be shown in the subsequent section where verification of our analytical modelling results is presented.

4. VERIFICATION OF THE MODELING RESULTS

As a comparison to [34], our verification was performed based on the 0.25 μm level CMOS process technology of TSMC. For simplicity, the candidate FGMOSFET was chosen from two inputs types. It should be noted that the modelling of a FGMOSFET for simulation was done based on the equivalent FGMOSFET circuit in Fig. 3. Modelling of the core MOSFET was achieved using BSIM3v3. The simulation methodology proposed in [39] was adopted for solving the convergence problem. The necessary parameters were extracted by MOSIS. A nominal L of 0.25 μm along with a nominal W/L of 20/0.25 was chosen. Of course, both NMOS and PMOS technological bases were considered.

Similar to [34], verification was performed in both qualitative and quantitative aspects. In the qualitative sense, graphical plots of $f(\delta I_D/I_D)$ of triode and saturation FGMOSFETs, which are the results of the analytical probabilistic modelling, were compared to those of the candidate triode and saturation FGMOSFETs based upon the probability distribution of $\Delta I_D/I_D$, and $f'(\delta I_D/I_D)$ obtained from Monte-Carlo SPICE simulations with 3000 runs. Concerning the quantitative aspects, the Kolmogorov-Smirnof test, i.e., the KS-test, is a powerful goodness of fit measure [40], [41] that was used with a 99% confidence level. According to [40] and [41], the objective of the KS-test is to perform a comparison of the KS-test statistic, KS and the critical value, c, where it can be stated that any model can accurately fit its target data set if and only if $KS \leq c$. For this research, KS can be defined as:

$$KS = \max_{\delta I_D/I_D}[|F'(\frac{\delta I_D}{I_D})| - |F(\frac{\delta I_D}{I_D})|] \qquad (10)$$

where,

$$F(\frac{\delta I_D}{I_D}) = \int_{-\infty}^{\delta I_D/I_D} f(x)dx \qquad (11)$$

$$F'(\frac{\delta I_D}{I_D}) = \int_{-\infty}^{\delta I_D/I_D} f'(x)dx \qquad (12)$$

Since the confidence level of the test was 99%, c can be given by equation [41]:

$$c = \frac{1.63}{\sqrt{n}} \qquad (13)$$

where n denotes the number of runs for the Monte-Carlo SPICE analysis, i.e., 3000, as previously stated. Thus, $c = 0.0297596$. In the upcoming subsections, verification of the triode and saturation FGMOSFET based modelling results will be given.

4.1 Verification of the triode FGMOSFET based result

The graphical comparisons of $f(\delta I_D/I_D)$ for the triode region FGMOSFET given by equation (7) with $N = 2$ as a candidate FGMOSFET, has two inputs. $f'(\delta I_D/I_D)$ obtained using the candidate triode region FGMOSFET is depicted in Fig. 4 and Fig. 5 for and N type and P-type FGMOSFETs, respectively, where $\Delta I_D/I_D$ is dimensionless and is expressed as a percentage. It can be seen that there is strong agreement between $f(\delta I_D/I_D)$ values obtained from the analytical probabilistic modelling and $f'(\delta I_D/I_D)$ obtained from the candidate FGMOSFET.

Moreover, it can be seen that the resulting KS obtained using equation (7) with $N = 2$ for determining can be found as $KS = 0.01935$ and $KS = 0.01873$ for N type and P-type FGMOSFETs, respectively. Both were lower than $c = 0.0297596$, which means that our derived $f(\delta I_D/I_D)$ of the FGMOSFET in the triode region can fit the candidate triode FGMOSFET based data with 99% confidence. These KS-test results and the aforementioned strong agreement in the comparative plots verify the accuracy of our triode region analytical probabilistic modelling.

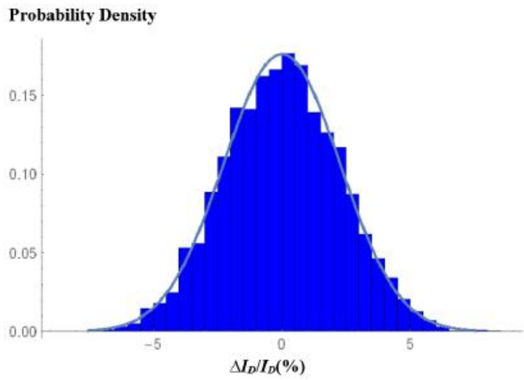


Fig.4: *N-type triode FGMOSFET based comparative plots of $f(\delta I_D/I_D)$ (line) and $f'(\delta I_D/I_D)$ (histogram).*

4.2 Verification of the saturation FGMOSFET based result

For the FGMOSFET in saturation, the comparative plots of $f(\delta I_D/I_D)$ given by equation (9) with

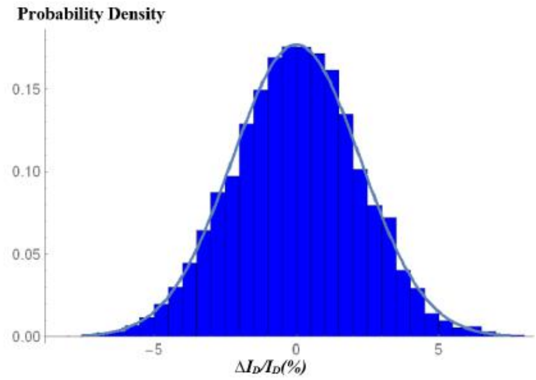


Fig.5: *P-type triode FGMOSFET based comparative plots of $f(\delta I_D/I_D)$ (line) and $f'(\delta I_D/I_D)$ (histogram).*

$N = 2$ and $f'(\delta I_D I_D)$ obtained using the candidate saturation FGMOSFET are depicted in Figs. 6 and 7 for N-type and P-type devices, respectively, where $\Delta I_D/I_D$ is expressed as a percentage. From these figures, strong agreement between $f(\delta I_D/I_D)$ and $f'(\delta I_D/I_D)$ can be observed. Moreover, it has been found that that the variation of the saturation FGMOSFET was more severe than that of the device in the triode region as indicated by the more dispersed distributions of $\Delta I_D/I_D$ of the saturation FGMOSFET.

Using equation (9) with $N = 2$ for determining $F(\delta I_D/I_D)$, the resulting KS values for N-type and P-type FGMOSFETs were found to be $KS = 0.02013$ and $KS = 0.01895$, respectively, which are both lower than $c = 0.0297596$. This means that our derived $f(\delta I_D/I_D)$ of the saturation FGMOSFET can fit the candidate saturation FGMOSFET based data with 99% confidence. The strong agreement in the comparative plots and the KS-test results verify the accuracy of our saturation region analytical probabilistic modelling. Since the above KS values are lower than those of previous modelling results [34], where $KS = 0.02823$ and $KS = 0.02619$ for N-type and P-type FGMOSFETs, respectively, it can be seen that our saturation region modelling results are more accurate.

5. THE APPLICATIONS

Unlike the previous variance term models [33], many statistical parameters and probabilities associated with ΔI_D can be formulated for both the triode and saturation region operated FGMOSFETs using our modelling results. This is because they are in terms of probability density functions. With these parameters and probabilities, it has been mathematically shown that the occurrence of ΔI_D is absolutely certain, the impact of ΔI_D to the design trade-offs associated with FGMOSFET can be studied and objective functions conducive to statistical/variability

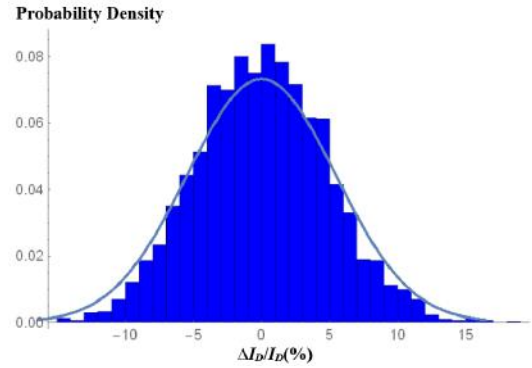


Fig.6: *N-type saturation FGMOSFET based comparative plots of $f(\delta I_D/I_D)$ (line) and $f'(\delta I_D/I_D)$ (histogram).*

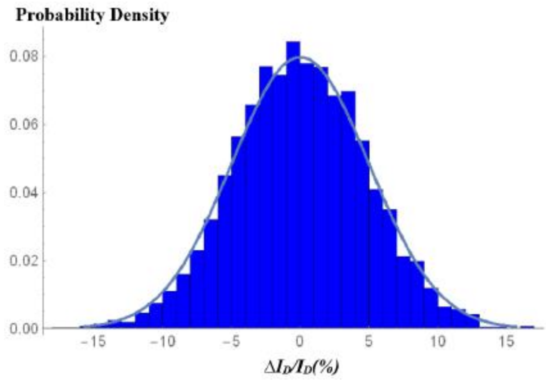


Fig.7: *P-type saturation FGMOSFET based comparative plots of $f(\delta I_D/I_D)$ (line) and $f'(\delta I_D/I_D)$ (histogram) .*

aware analysis and design can be obtained. Moreover, analytical probabilistic modelling and computationally efficient statistical/variability aware numerical simulation of the FGMOSFET based circuits can also be performed. These applications of our modelling results will be subsequently presented in this section.

5.1 Formulation of statistical parameters and the related objective functions

Using our $f(\delta I_D/I_D)$, statistical parameters of $\Delta I_D/I_D$ such as average, median and variance can be analytically obtained for FGMOSFETs in both the triode and saturation regions. To do so, conventional statistical mathematics must be applied. For example, the average of $\Delta I_D/I_D$ i.e. $\overline{\Delta I_D/I_D}$, i.e., its first moment, can be mathematically defined as:

$$\frac{\overline{\Delta I_D}}{I_D} = \int_{-\infty}^{\infty} \frac{\delta I_D}{I_D} f(\frac{\delta I_D}{I_D}) d\frac{\delta I_D}{I_D} \qquad (14)$$

After applying our $f(\delta I_D/I_D)$ given by equations (7) and (9) for triode and saturation FGMOSFETs, to equation (14), $\overline{\Delta I_D/I_D}$ was found to be zero for FGMOSFETs in both the triode and saturation regions.

Moreover, the median value of $\Delta I_D/I_D$, m can be determined by solving the following equation:

$$\int_{-\infty}^{\infty} f(\frac{\delta I_D}{I_D})d\frac{\delta I_D}{I_D} = \frac{1}{2} \quad (15)$$

After applying equations (7) and (9), it was found that $m = 0$ for the FGMOSFET in both regions.

Even though $\overline{\Delta I_D/I_D}$ and m were found to be zero, the variance of $\Delta I_D/I_D$ $\sigma^2_{\Delta I_D/I_D}$ was not. The value of $\sigma^2_{\Delta I_D/I_D}$ can be mathematically defined as:

$$\sigma^2_{\frac{\Delta I_D}{I_D}} = \int_{-\infty}^{\infty} \left(\frac{\delta I_D}{I_D} - \frac{\overline{\Delta I_D}}{I_D}\right)^2 f(\frac{\delta I_D}{I_D})d\frac{\delta I_D}{I_D} \quad (16)$$

By applying equations (7) and (9), $\sigma^2_{\Delta I_D/I_D}$ of the FGMOSFET in the triode and saturation regions can be respectively determined as:

$$\sigma^2_{\frac{\Delta I_D}{I_D}} = \frac{N_{sub}W_{dep}q^2}{3WLC^2_{inv}\{\theta[1-\theta(\sum_{i=1}^{N} k_iV_i - V_S - V_t)]^{-1} - V_{DS}[(\sum_{i=1}^{N} k_iV_i - V_S - V_t)V_{DS} - \frac{1}{2}V^2_{DS}]^{-1}\}^{-2}} \quad (17)$$

$$\sigma^2_{\frac{\Delta I_D}{I_D}} = \frac{N_{sub}W_{dep}q^2}{3WLC^2_{inv}\{\theta[1-\theta(\sum_{i=1}^{N} k_iV_i - V_S - V_t)]^{-1} - 2(\sum_{i=1}^{N} k_iV_i - V_S - V_t)^{-1}\}} \quad (18)$$

Since $\overline{\Delta I_D/I_D} = 0$, the second moment of $\Delta I_D/I_D$, $\overline{(\Delta I_D/I_D)^2}$ is mathematically defined as:

$$\overline{(\frac{\Delta I_D}{I_D})^2} = \int_{-\infty}^{\infty} (\frac{\delta I_D}{I_D})^2 f(\frac{\delta I_D}{I_D})d\frac{I_D}{I_D} \quad (19)$$

and was been found to be equal to $\sigma^2_{\Delta I_D/I_D}$. Thus $\overline{(\Delta I_D/I_D)^2}$ can also be given by equations (17) and (18) for triode and saturation FGMOSFETs, respectively.

Finally, the moment generating function of $\Delta I_D/I_D$, $M(u)$, is the summation of all moments of $\Delta I_D/I_D$ including the first and second ones [42]. It is mathematically defined as:

$$M(u) = \int_{-\infty}^{\infty} \exp(u\frac{\delta I_D}{I_D})f(\frac{\delta I_D}{I_D})d\frac{I_D}{I_D} \quad (20)$$

These can be respectively obtained for triode and saturation FGMOSFETs by applying equations (7) and (9) to (20) as follows:

$$M(u) = \exp[\frac{N_{sub}W_{dep}q^2}{6WLC^2_{inv}\{\theta[1-\theta(\sum_{i=1}^{N} k_iV_i - V_S - V_t)]^{-1} - V_{DS}[(\sum_{i=1}^{N} k_iV_i - V_S - V_t)V_{DS} - \frac{1}{2}V^2_{DS}]^{-1}\}^{-2}}] \quad (21)$$

$$M(u) = \exp[\frac{N_{sub}W_{dep}q^2u^2}{6WLC^2_{inv}\{\theta[1-\theta(\sum_{i=1}^{N} k_iV_i - V_S - V_t)]^{-1} - 2(\sum_{i=1}^{N} k_iV_i - V_S - V_t)^{-1}\}^{-2}}] \quad (22)$$

From equations (17) and (18), it can be seen that:

$$\sigma^2_{\Delta I_D/I_D} \alpha \frac{1}{WL} \quad (23)$$

for both the triode and saturation region operated FGMOSFETs. Thus, lowering the device area causes an increased ΔI_D as a penalty. This is an example of the impact of ΔI_D upon the design trade-offs involving FGMOSFETs. Using equationss (17) and (18), the minimum value of WL, $(WL)_{min}$ which dictates the minimum area of the transistor, can be obtained for the triode and saturation FGMOSFETs as follows:

$$(WL)_{\min} = \frac{N_{sub}W_{dep}q^2\{\theta[1-\theta(\sum_{i=1}^{N} k_iV_i - V_S - V_t)^{-1}] - V_{DS}[(\sum_{i=1}^{N} k_iV_i - V_S - V_t)V_{DS} - \frac{1}{2}V^2_{DS}]^{-1}\}^{-2}}{3C^2_{inv}\sigma^2_{\Delta I_D/I_D,\max}} \quad (24)$$

$$(WL)_{\min} = \frac{N_{sub}W_{dep}q^2\{\theta[1-\theta(\sum_{i=1}^{N} k_iV_i - V_S - V_t)^{-1}] - 2(\sum_{i=1}^{N} k_iV_i - V_S - V_t)^{-1}\}^{-2}}{3WLC^2_{inv}\sigma^2_{\Delta I_D/I_D,\max}} \quad (25)$$

where $\sigma^2_{\Delta I_D/I_{D,\max}}$ denotes the maximum acceptable value of $\sigma^2_{\Delta I_D/I_D}$.

Since $\sigma^2_{\Delta I_D/I_D}$ reflects the size of $\Delta I_D/I_D$, the statistical/variability aware design of any single FGMOSFET can be performed using the optimization scheme with the following objective function:

$$\min[\sigma^2_{\Delta I_D/I_D}] \quad (26)$$

where either equation (17) or (18) must be used according to the operating region of the FGMOSFET under consideration.

5.2 The formulation of the useful probabilities and objective functions

In order to do so, the cumulative distribution function of $\Delta I_D/I_D$, $F(\delta I_D/I_D)$ and the survival function [43] of $\Delta I_D/I_D$,$S(\delta I_D/I_D)$ must first be derived. It can be seen that $F(\delta I_D/I_D)$ is equivalent to the probability that $\Delta I_D/I_D \leq \delta I_D/I_D$ i.e.,

$\Pr\{\Delta I_D/I_D \leq \delta I_D/I_D\}$. This is also necessary for determining KS as shown in the previous section, and can be obtained using our $f(\delta I_D/I_D)$ as in equation (11). For a FGMOSFET in the triode and saturation regions with an arbitrary value of N, $F(\delta I_D/I_D)$ can be respectively given by applying equations (7) and (9) to (11) without specifying N as follows:

$$F(\frac{\delta I_D}{I_D})=\frac{1}{2}[1+erf(\frac{\sqrt{3WL}C_{inv}(\delta I_D/I_D)}{\sqrt{2N_{sub}W_{dep}}q\{\theta[1-\theta(\sum_{i=1}^{N}k_iV_i-V_S-V_t)]^{-1}-V_{DS}[(\sum_{i=1}^{N}k_iV_i-V_S-V_t)V_{DS}-\frac{1}{2}V_{DS}^2]^{-1}\}})] \tag{27}$$

$$F(\frac{\delta I_D}{I_D})=\frac{1}{2}[1+erf(\frac{\sqrt{3WLC}_{inv}(\delta I_D/I_D)}{\sqrt{2N_{sub}W_{dep}}q\{\theta[1-\theta(\sum_{i=1}^{N}k_iV_i-V_S-V_t)]^{-1}-2(\sum_{i=1}^{N}k_iV_i-V_S-V_t)^{-1}\}})] \tag{28}$$

where erf() denotes the error function and can be mathematically defined in terms of an arbitrary variable, y, as:

$$erf(y)=\frac{2}{\sqrt{\pi}}\int_0^y \exp(-u^{2)du} \tag{29}$$

Alternatively, $S(\delta I_D/I_D)$, which is equivalent to the probability that $\Delta I_D/I_D > \delta I_D/I_D$, i.e., $\Pr\{\Delta I_D/I_D > \delta I_D/I_D\}$, can be mathematically defined as:

$$S(\frac{\delta I_D}{I_D})=\int_{\delta I_D/I_D}^{\infty} f(x)dx \tag{30}$$

Thus, $S(\delta I_D/I_D)$ of a FGMOSFET in the triode and saturation regions can be found by respectively applying equations (7) and (9) to (30) as given by equations (31) and (32). Then, the useful probabilities and objective functions can be derived and used for statistical/variability aware analysis and design involving FGMOSFETs:

$$S(\frac{\delta I_D}{I_D})=1-\frac{1}{2}[1+erf(\frac{\sqrt{3WL}C_{inv}(\delta I_D/I_D)}{\sqrt{2N_{sub}W_{dep}}q\{\theta[1-\theta(\sum_{i=1}^{N}k_iV_i-V_S-V_t)]^{-1}-V_{DS}[(\sum_{i=1}^{N}k_iV_i-V_S-V_t)V_{DS}-\frac{1}{2}V_{DS}^2]^{-1}\}})] \tag{31}$$

$$S(\frac{\delta I_D}{I_D})=1-\frac{1}{2}[1+erf(\frac{\sqrt{3WLC}_{inv}(\delta I_D/I_D)}{\sqrt{2N_{sub}W_{dep}}q\{\theta[1-\theta(\sum_{i=1}^{N}k_iV_i-V_S-V_t)]^{-1}-2(\sum_{i=1}^{N}k_iV_i-V_S-V_t)^{-1}\}})] \tag{32}$$

Firstly, it will be shown that the probability that $\Delta I_D/I_D$, which is a continuous random variable, lies within a certain range given as $[a, b]$ and can be derived using $F(\delta I_D/I_D)$. This probability is denoted as $\Pr\{a \leq \Delta I_D/I_D \leq b\}$, and can be given in terms of $F(\delta I_D/I_D)$ as:

$$\Pr\{a \leq \frac{\Delta I_D}{I_D} \leq b\} = F(b) - F(a) \tag{33}$$

Thus, by applying equations (27) and (28) to (33), $\Pr\{a \leq I_D/I_D \leq b\}$ of a FGMOSFET in the triode and saturation regions can be respectively obtained as follows:

$$\Pr\{a\leq\frac{\Delta I_D}{I_D}\leq b\}=\frac{1}{2}[erf(\frac{\sqrt{3WL}C_{inv}b}{\sqrt{2N_{sub}W_{dep}}q\{\theta[1-\theta(\sum_{i=1}^{N}k_iV_i-V_S-V_t)]^{-1}-V_{DS}[(\sum_{i=1}^{N}k_iV_i-V_S-V_t)V_{DS}-\frac{1}{2}V_{DS}^2]^{-1}\}}) -erf(\frac{\sqrt{3WL}C_{inv}a}{\sqrt{2N_{sub}W_{dep}}q\{\theta[1-\theta(\sum_{i=1}^{N}k_iV_i-V_S-V_t)]^{-1}-V_{DS}[(\sum_{i=1}^{N}k_iV_i-V_S-V_t)V_{DS}-\frac{1}{2}V_{DS}^2]^{-1}\}})] \tag{34}$$

$$\Pr\{a\leq\frac{\Delta I_D}{I_D}\leq b\}=\frac{1}{2}[erf(\frac{\sqrt{3WLC}_{inv}b}{\sqrt{2N_{sub}W_{dep}}q\{\theta[1-\theta(\sum_{i=1}^{N}k_iV_i-V_S-V_t)]^{-1}-2(\sum_{i=1}^{N}k_iV_i-V_S-V_t)^{-1}\}}) -erf(\frac{\sqrt{3WLC}_{inv}a}{\sqrt{2N_{sub}W_{dep}}q\{\theta[1-\theta(\sum_{i=1}^{N}k_iV_i-V_S-V_t)]^{-1}-2(\sum_{i=1}^{N}k_iV_i-V_S-V_t)^{-1}\}})] \tag{35}$$

It can be seen that $\Pr\{a \leq \Delta I_D/I_D \leq b\}$ can be directly used for predicting the probability that $\Delta I_D/I_D$ lies within an arbitrary predetermined range. Moreover, the probability that the magnitude of $\Delta I_D/I_D$ does not exceed the allowable maximum value, which the resulting variation in the performance of FGMOSFET, can be determined using $\Pr\{a \leq \Delta I_D/I_D \leq b\}$. This probability is obviously useful in statistical/variability aware analysis and design involving FGMOSFETs.

To do so, we let the aforesaid maximum value be $|\Delta I_D/I_D|_{\max}$. Thus such probability denoted by $\Pr\{|\Delta I_D/I_D| \leq |\Delta I_D/I_D|_{\max}\}$, can be determined using $\Pr\{a \leq \Delta I_D/I_D \leq b\}$ with $a = -|\Delta I_D/I_D|_{\max}$ and $b = |\Delta I_D/I_D|_{\max}$. This is because $|\Delta I_D/I_D| \leq |\Delta I_D/I_D|_{\max}$ is equivalent to $-|\Delta I_D/I_D|_{\max} \leq \Delta I_D/I_D \leq |\Delta I_D/I_D|_{\max}$. As a result, $\Pr\{|\Delta I_D/I_D| \leq |\Delta I_D/I_D|_{\max}\}$ of the triode and saturation FGMOSFETs can be obtained using equations (34) and (35) as follows:

$$\Pr\{\left|\frac{\Delta I_D}{I_D}\right|\leq\left|\frac{\Delta I_D}{I_D}\right|_{\max}\}=\frac{1}{2}[erf(\frac{\sqrt{3WL}C_{inv}|\Delta I_D/I_D|_{\max}}{\sqrt{2N_{sub}W_{dep}}q\{\theta[1-\theta(\sum_{i=1}^{N}k_iV_i-V_S-V_t)]^{-1}}) \tag{36}$$

$$-V_{DS}[(\sum_{i=1}^{N} k_iV_i - V_S - V_t)V_{DS} - \tfrac{1}{2}V_{DS}^2]^{-1}\}$$

$$-erf(\frac{\sqrt{3WLC}_{inv}|\Delta I_D/I_D|_{\max}}{\sqrt{2N_{sub}W_{dep}}q\{\theta[1-\theta(\sum_{i=1}^{N} k_iV_i - V_S - V_t)]^{-1}}$$

$$-V_{DS}[(\sum_{i=1}^{N} k_iV_i - V_S - V_t)V_{DS} - \tfrac{1}{2}V_{DS}^2]^{-1}\})]$$

$$\Pr\{\left|\frac{\Delta I_D}{I_D}\right| \le \left|\frac{\Delta I_D}{I_D}\right|_{\max}\} = \tfrac{1}{2}[erf(\frac{\sqrt{3WLC}_{inv}|\Delta I_D/I_D|_{\max}}{\sqrt{2N_{sub}W_{dep}}q\{\theta[1-\theta(\sum_{i=1}^{N} k_iV_i - V_S - V_t)]^{-1}}$$

$$-2(\sum_{i=1}^{N} k_iV_i - V_S - V_t)^{-1}\}^{-2}) \tag{37}$$

$$-erf(\frac{\sqrt{3WLC}_{inv}|\Delta I_D/I_D|_{\max}}{\sqrt{2N_{sub}W_{dep}}q\{\theta[1-\theta(\sum_{i=1}^{N} k_iV_i - V_S - V_t)]^{-1}}$$

$$-2(\sum_{i=1}^{N} k_iV_i - V_S - V_t)^{-1}\}^{-2})]$$

Since maximizing $\Pr\{|\Delta I_D/I_D| \le |\Delta I_D/I_D|_{\max}\}$ yields the greatest likelihood of obtaining an acceptable magnitude of $\Delta I_D/I_D$, which is a goal of the statistical/variability aware design, the following objective function was found useful:

$$\max[\Pr\{|\Delta I_D/I_D| \le |\Delta I_D/I_D|_{\max}\}] \tag{38}$$

From equations (36) and (37), it can be seen that (38) can be satisfied for a FGMOSFET in both operating regions using the optimum values of controllable parameters, such as W, L and V_S.

Contrary to $\Pr\{a \le \Delta I_D/I_D \le b\}$, the probability that $\Delta I_D/I_D$ lies outside $[a,\ b]$ i.e., $\Pr\{(\Delta I_D/I_D < a) \vee (\Delta I_D/I_D < b)\}$, can be obtained using $S(\delta I_D/I_D)$as:

$$\Pr\{(\frac{\Delta I_D}{I_D} < a) \vee (\frac{\Delta I_D}{I_D} < b)\} = 1 + S(b) - S(a) \tag{39}$$

Thus, applying equations (31) and (32) to (39), $\Pr\{(\Delta I_D/I_D < a) \vee (\Delta I_D/I_D < b)\}$ of the triode and saturation FGMOSFETs can be respectively given by equationss (40) and (41). Using $\Pr\{(\Delta I_D/I_D < a) \vee (\Delta I_D/I_D < b)\}$ with $a = -|\Delta I_D/I_D|_{\max}$ and $b = |\Delta I_D/I_D|_{\max}$, the probability that $|\Delta I_D/I_D| > |\Delta I_D/I_D|_{\max}$ i.e., $\Pr\{|\Delta I_D/I_D| > |\Delta I_D/I_D|_{\max}\}$, can be obtained. This is because $|\Delta I_D/I_D| > |\Delta I_D/I_D|_{\max}$ is equivalent to the union of $\Delta I_D/I_D < -|\Delta I_D/I_D|_{\max}$ and $\Delta I_D/I_D > |\Delta I_D/I_D|_{\max}$, which are mutually exclusive events.

$$\Pr\{(\frac{\Delta I_D}{I_D} < a) \vee (\frac{\Delta I_D}{I_D} < b)\}$$

$$= 1 - \tfrac{1}{2}[erf(\frac{\sqrt{3WLC}_{inv}b}{\sqrt{2N_{sub}W_{dep}}q\{\theta[1-\theta(\sum_{i=1}^{N} k_iV_i - V_S - V_t)]^{-1}}$$

$$-V_{DS}[(\sum_{i=1}^{N} k_iV_i - V_S - V_t)V_{DS} - \tfrac{1}{2}V_{DS}^2]^{-1}\})$$

$$-erf(\frac{\sqrt{3WLC}_{inv}a}{\sqrt{2N_{sub}W_{dep}}q\{\theta[1-\theta(\sum_{i=1}^{N} k_iV_i - V_S - V_t)]^{-1}}$$

$$-V_{DS}[(\sum_{i=1}^{N} k_iV_i - V_S - V_t)V_{DS} - \tfrac{1}{2}V_{DS}^2]^{-1}\})] \tag{40}$$

$$\Pr\{(\frac{\Delta I_D}{I_D} < a) \vee (\frac{\Delta I_D}{I_D} > b)\} = 1 - \tfrac{1}{2}[erf(\frac{\sqrt{3WLC}_{inv}b}{\sqrt{2N_{sub}W_{dep}}q\{\theta[1-\theta(\sum_{i=1}^{N} k_iV_i - V_S - V_t)]^{-1}}$$

$$-2(\sum_{i=1}^{N} k_iV_i - V_S - V_t)^{-1}\}) \tag{41}$$

$$-erf(\frac{\sqrt{3WLC}_{inv}a}{\sqrt{2N_{sub}W_{dep}}q\{\theta[1-\theta(\sum_{i=1}^{N} k_iV_i - V_S - V_t)]^{-1}}$$

$$-2(\sum_{i=1}^{N} k_iV_i - V_S - V_t)^{-1}\})]$$

Therefore, $\Pr\{|\Delta I_D/I_D| > |\Delta I_D/I_D|_{\max}\}$ of the triode and saturation FGMOSFETs can be given using equations (38) and (39) as follows:

$$\Pr\{\left|\frac{\Delta I_D}{I_D}\right| > \left|\frac{\Delta I_D}{I_D}\right|_{\max}\}$$

$$= 1 - \tfrac{1}{2}[erf(\frac{\sqrt{3WLC}_{inv}|\Delta I_D/I_D|_{\max}}{\sqrt{2N_{sub}W_{dep}}q\{\theta[1-\theta(\sum_{i=1}^{N} k_iV_i - V_S - V_t)]^{-1}}$$

$$-V_{DS}[(\sum_{i=1}^{N} k_iV_i - V_S - V_t)V_{DS} - \tfrac{1}{2}V_{DS}^2]^{-1}\})$$

$$-erf(\frac{\sqrt{3WLC}_{inv}|\Delta I_D/I_D|_{\max}}{\sqrt{2N_{sub}W_{dep}}q\{\theta[1-\theta(\sum_{i=1}^{N} k_iV_i - V_S - V_t)]^{-1}}$$

$$-V_{DS}[(\sum_{i=1}^{N} k_iV_i - V_S - V_t)V_{DS} - \tfrac{1}{2}V_{DS}^2]^{-1}\})] \tag{42}$$

$$\Pr\{\left|\frac{\Delta I_D}{I_D}\right| \le \left|\frac{\Delta I_D}{I_D}\right|_{\max}\} = 1 - \tfrac{1}{2}[erf(\frac{\sqrt{3WLC}_{inv}|\Delta I_D/I_D|_{\max}}{\sqrt{2N_{sub}W_{dep}}q\{\theta[1-\theta(\sum_{i=1}^{N} k_iV_i - V_S - V_t)]^{-1}}$$

$$-2(\sum_{i=1}^{N} k_iV_i - V_S - V_t)^{-1}\}^{-2}) \tag{43}$$

$$-erf(\frac{\sqrt{3WLC}_{inv}|\Delta I_D/I_D|_{\max}}{\sqrt{2N_{sub}W_{dep}}q\{\theta[1-\theta(\sum_{i=1}^{N} k_iV_i - V_S - V_t)]^{-1}}$$

$$-2(\sum_{i=1}^{N} k_iV_i - V_S - V_t)^{-1}\}^{-2})]$$

The greatest probability of obtaining an acceptable magnitude of $\Delta I_D/I_D$ can be alternatively determined by minimizing $\Pr\{|\Delta I_D/I_D| > |\Delta I_D/I_D|_{\max}\}$. So, the following objective function can also be satisfied for both the triode and saturation FGMOSFETs using the optimum settings of controllable parameters. The following has been found useful:

$$\min[\Pr\{|\Delta I_D/I_D| > |\Delta I_D/I_D|_{\max}\}] \tag{44}$$

Finally, the probability of the occurrence of ΔI_D will be determined. Since the occurrence of ΔI_D yields $|\Delta I_D/I_D| > 0$, this probability is equal to the probability that $|\Delta I_D/I_D| > 0$ i.e., $\Pr\{\Delta I_D/I_D > 0\}$. As $|\Delta I_D/I_D| > 0$ is equivalent to the union of $\Delta I_D/I_D < 0$ and $\Delta I_D/I_D < 0$, which are mutually exclusive events, $\Pr\{|\Delta I_D/I_D| > 0\}$ can be determined using with $a = b = 0$. As a result, it was found by using equations (40) and (41) with $a = b = 0$, that $Pr\{\Delta I_D/I_D > 0\} = 1$ for both the triode and saturation FGMOSFETs. This means that the probability of the occurrence of ΔI_D is equal to 1 for a FGMOSFET in both operating regions. Thus, the occurrence of ΔI_D is absolutely certain. This emphasizes the necessity of our work and focuses on the certainly of ΔI_D.

5.3 Analytical probabilistic modelling of the FGMOSFET based circuit

Using our $f(\delta I_D/I_D)$, the analytical probabilistic modelling of a FGMOSFET based circuit is possible. This is because the key parameter of such a circuit depends upon the I_D values intrinsic to this FGMOSFET. Here, we let this parameter and its variation be respectively represented as X and ΔX. Thus, the analytical probabilistic modelling of a FGMOSFET based circuit can be conveniently performed by determining the probability density function of the normalized value of ΔX with respected to X, $\Delta X/X$, i.e., $g(\Delta X/X)$, where $\delta X/X$ is the sample variable of $\Delta X/X$. To obtain $g(\Delta X/X)$, the principle of transformation of random variables [42] must be applied. If we assume that a single FGMOSFET in the circuit affects X, $g(\Delta X/X)$ can be given as [34]:

$$g(\frac{\delta X}{X}) = f(\frac{\delta I_D}{I_D}(\frac{\delta X}{X}))\left[\left\|\frac{\partial(\delta I_D/I_D)}{\partial(\delta X/X)}\right\|\Bigg|_{\frac{\delta I_D}{I_D}\to\frac{\delta I_D}{I_D}(\frac{\delta X}{X})}\right]^{-1} \tag{45}$$

With the resulting $g(\delta X/X)$, the statistical parameters of $\Delta X/X$ and probabilities needed for statistical/variability aware design can be analytically determined in a similar manner to those of $\Delta I_D/I_D$. Thus, for example, the mean, variance, second moment of $\Delta X/X$ and the probability that the magnitude of $\Delta X/X$ does not exceed its maximum allowable value, $|\Delta X/X|_{\max}$, can be respectively given using the following equations:

$$\overline{\frac{\Delta X}{X}} = \int_{\infty}^{\infty} \frac{\delta X}{X} f(\frac{\delta X}{X}) d\frac{\delta X}{X} \tag{46}$$

$$\sigma^2_{\frac{\Delta X}{X}} = \int_{\infty}^{\infty} \left(\frac{\delta X}{X} - \overline{\frac{\delta X}{X}}\right)^2 f(\frac{\delta X}{X}) d\frac{\delta X}{X} \tag{47}$$

$$\overline{(\frac{\Delta X}{X})^2} = \int_{-\infty}^{\infty} (\frac{\delta X}{X})^2 f(\frac{\delta X}{X}) d(\frac{\delta X}{X}) \tag{48}$$

$$\Pr\{\left|\frac{\Delta X}{X}\right| \le \left|\frac{\Delta X}{X}\right|_{\max}\} = G(\left|\frac{\Delta X}{X}\right|_{\max}) - G(-\left|\frac{\Delta X}{X}\right|_{\max}) \tag{49}$$

where $G(|\Delta X/X|_{\max})$ and $G(-|\Delta X/X|_{\max})$ are respectively the cumulative distribution function of $\Delta X/X$ denoted by $G(\delta X/X)$ with $\delta X/X = |\Delta X/X|_{\max}$ and $\delta X/X = -|\Delta X/X|_{\max} \cdot G(\delta X/X)$ can be generally given by:

$$G(\frac{\delta X}{X}) = \int_{-\infty}^{\delta X/X} g(x)dx \tag{50}$$

As a case study using a practical FGMOSFET based circuit, we let X be the transconductance, G_m, of a FGMOSFET based voltage to current converter (VIC) [44]. The core circuit of this VIC is depicted in Fig. 8 [34], where M1 is a FGMOSFET and has been drawn in its equivalent circuit form.

Since M1, which operates in the saturation region, affects the performance of this FGMOSFET based VIC, G_m is equal to the transconductance of M1 and can be given in term its I_D value, i.e., I_{D1}, by equation (51). It can be seen that C_f and C_{in}, which act as the feedback and input capacitances, are actually the coupling capacitances of M1. As this FGMOSFET has two inputs and its source terminal has been grounded, I_{D1} can be given in terms of C_f and C_{in} by equation (52). Thus, the normalized random variation in G_m, $\Delta G_m/G_m$, can be formulated in terms of $\Delta I_{D1}/I_{D1}$ as given by equation (53). It should be noted that the short channel effects have been taken into account in the formulation of these equations.

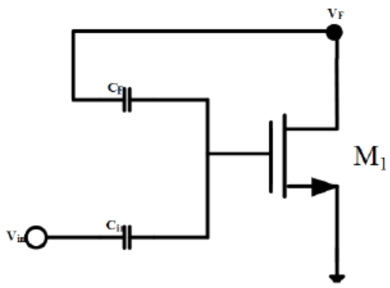


Fig.8: *The core circuit of FGMOSFET based VIC [34].*

$$\begin{aligned} G_m = &\frac{C_{in}}{C_{in}+C_f}\{2(\frac{C_{in}V_{in}}{C_{in}+C_f}+\frac{C_fV_f}{C_{in}+C_f}-V_t)^{-1} \\ &-\theta[1-\theta(\frac{C_{in}V_{in}}{C_{in}+C_f}+\frac{C_fV_f}{C_{in}+C_f}-V_t)]^{-1}\}I_{D1} \end{aligned} \tag{51}$$

$$\begin{aligned} I_{D1} = & \frac{\mu}{2} C_{ox} \frac{W}{L} [1 - \theta(\frac{C_{in} V_{in}}{C_{in} + C_f} + \frac{C_f V_f}{C_{in} + C_f} - V_t)] \\ & \times (\frac{C_{in} V_{in}}{C_{in} + C_f} + \frac{C_f V_f}{C_{in} + C_f} - V_t)^2 (1 + \lambda V_{DS}) \end{aligned} \tag{52}$$

$$\frac{\Delta G_m}{G_m} = \frac{\Delta I_{D1}}{I_{D1}} \tag{53}$$

As a result, the probability density function of $\Delta G_m/G_m$, $g(\delta G_m/G_m)$ can be analytically determined using our $f(\delta I_D/I_D)$ derived of the saturation FGMOSFET, i.e., equations (9), with $N = 2$, (45) and (53) as given by equation (54), where $\delta G_m/G_m$ is the sample variance of $\Delta G_m/G_m$. Using $g(\delta G_m/G_m)$, the useful statistical parameters of $\Delta G_m/G_m$ and probabilities can be analytically obtained. As an example, the variance of $\Delta G_m/G_m$, $\sigma^2_{\Delta G_m/G_m}$, which was analytically determined using equation (47) with $\Delta G_m/G_m$ and $g(\delta G_m/G_m)$, serve as $\Delta X/X$ and $g(\delta X/X)$, respectively, in this practical case study. They can be obtained as given by equation (55). Similar to equations (51)-(55), the short channel effects were included in the derivation of equations (54) and (55).

$$\begin{aligned} g(\frac{\delta G_m}{G_m}) = & \frac{(C_{inv}/q)\sqrt{3WL/2\pi N_{sub} W_{dep}}}{\{\theta[1-\theta(\frac{C_{in} V_{in}}{C_{in}+C_f} + \frac{C_f V_f}{C_{in}+C_f} - V_t)]^{-1}} \\ & -2(\frac{C_{in} V_{in}}{C_{in}+C_f} + \frac{C_f V_f}{C_{in}+C_f} - V_t)^{-1}\} \\ & \times \exp[\frac{3WL(C_{inv}\delta G_m/G_m)^2}{2N_{sub} W_{dep} q^2 \{\theta[1-\theta(\frac{C_{in} V_{in}}{C_{in}+C_f} + \frac{C_f V_f}{C_{in}+C_f} - V_t)]^{-1}} \\ & -2(\frac{C_{in} V_{in}}{C_{in}+C_f} + \frac{C_f V_f}{C_{in}+C_f} - V_t)^{-1}\}^{-2} \end{aligned} \tag{54}$$

$$\begin{aligned} \sigma^2_{\frac{\Delta G_m}{G_m}} = & \frac{N_{sub} W_{dep} q^2}{3WLC^2_{inv}\{\theta[1-\theta(\frac{C_{in} V_{in}}{C_{in}+C_f} + \frac{C_f V_f}{C_{in}+C_f} - V_t)]^{-1}} \\ & -2(\frac{C_{in} V_{in}}{C_{in}+C_f} + \frac{C_f V_f}{C_{in}+C_f} - V_t)^{-1}\}^{-2} \end{aligned} \tag{55}$$

5.4 Computationally efficient statistical/variability aware simulation of FGMOSFET based circuit

The statistical parameters of $\Delta X/X$ must be evaluated numerically without predetermining $g(\delta X/X)$. Among these parameters, $\sigma^2_{\Delta X/X}$ has been often cited as it directly measures the magnitude of variability in X. Traditionally, numerical determination of $\sigma^2_{\Delta X/X}$ must be performed using a Monte-Carlo simulation with numerous runs, numbering hundreds or thousands of runs.

Fortunately, determination of $\sigma^2_{\Delta X/X}$ in the small signal analysis [45] based simulation, where the required computational cost is significantly lower than that of a conventional Monte-Carlo simulation, becomes possible using our $f(\delta I_D/I_D)$. This is because $\sigma^2_{\Delta I_D/I_D}$ values are necessary inputs and can be known in advance via applying $f(\delta I_D/I_D)$ as discussed in the previous subsection. If we let the FGMOSFET based circuit under consideration be composed of M FGMOSFETs, $\sigma^2_{\Delta X/X}$ can be numerically determined using the following equation:

$$\sigma^2_{\frac{\Delta X}{X}} = \sum_{m=1}^{M} [\left(\frac{I_{Dm}}{X} S^X_{I_{Dm}}\right)^2 \sigma^2_{\frac{\Delta I_{Dm}}{I_{Dm}}}] \tag{56}$$

where I_{Dm}, $S^X_{I_{Dm}}$ and $\sigma^2_{\Delta I_{Dm}/I_{Dm}}$ denote the ideal drain current of an arbitrary m$^{\text{th}}$ FGMOSFET in the circuit in which all nonidealities have been neglected and the sensitivity of X with respect to I_{Dm}. Moreover, $\sigma^2_{\Delta I_{Dm}/I_{Dm}}$ stands for $\sigma^2_{\Delta I_D/I_D}$ of the m$^{\text{th}}$ FGMOSFET. It is notable that there exists no correlation terms in relationship (56). This is because $\Delta I_{Dm}/I_{Dm}$ values are random variations that exhibit no spatial correlations [37]. According to [45], $S^X_{I_{Dm}}$ which quantitatively represents the degree of dependency on I_{Dm} of X, can be mathematically defined as follows:

$$S^X_{I_{Dm}} = \frac{\partial X}{\partial I_{Dm}} \tag{57}$$

Since the whole set of $S^X_{I_{Dm}}$ values can be obtained using sensitivity analysis in which the circuit under consideration requires only one calculation for its solution [45], the computational effort can be significantly reduced compared to a conventional Monte-Carlo simulation. Finally, since $\sigma^2_{\Delta X/X}$ reflects the magnitude of variability in X, the following objective function can be satisfied using the optimum controllable parameters of the FGMOSFETs within the circuit.

$$\min[\sigma^2_{\Delta X/X}] \tag{58}$$

6. CONCLUSIONS

In this research, an analytical probabilistic modelling of ΔI_D of a FGMOSFET caused by manufacturing process variations has been proposed. The resulting models are in terms of a probability density function of $\Delta I_D/I_D$. Unlike [34], both the triode and saturation region operated FGMOSFETs were considered and the short channel effects taken into account. The resulting models have been found to be very accurate as they can fit the probabilistic distributions of $\Delta I_D/I_D$ obtained from Monte-Carlo SPICE simulations of the 0.25 μm level BSIM3v3 based candidate triode and saturation FGMOSFETs with very high accuracy. It was found that the variation of a saturation FGMOSFET is more severe. Moreover, the KS-tests were satisfied at a 99% confidence level. It was found that our resulting model for the saturation FGMOSFET was more accurate than those previously proposed [34].

Unlike the previous variance term models [33], statistical parameters and probabilities associated with ΔI_D can be formulated for both the triode and saturation region operated FGMOSFETs using our results. With such parameters and probabilities, it was found that the occurrence of ΔI_D is absolutely certain, which highlights the necessity of our work. Moreover, the impact of ΔI_D upon the design trade-offs associated with FGMOSFETs can be studied. Many objective functions that are conducive to the statistical/variability aware analysis and design, can be derived. Analytical probabilistic modelling and reduced computational effort for statistical/variability aware simulation of the FGMOSFET associated circuit can also be performed. Therefore our proposed probabilistic modelling of a FGMOSFET has been found to be useful for statistical/variability aware analysis and design of various FGMOSFET based circuits and systems. These have applications in many areas, e.g., signal processing [2], [3] dosimetry systems, [5], [6], biomedical engineering [7] and neural networks [21], [31], among others.

ACKNOWLEDGEMENT

The first author would like to acknowledge Mahidol University, Thailand, for their online database service, which served as a primary information resource.

References

[1] G. Serrano, and P. Hasler, "A Floating Gate DAC Array," *Proceeding of the 2004 International Symposium on Circuits and Systems (ISCAS 04)*, pp. 357-360, 2004.

[2] M. Drakaki, G. Fikos and S. Siskos, "Analog Signal Processing Circuits Using Floating Gate MOS Transistors," *Proceedings of the International Conference on Technology and Automation (ICTA 05)*, pp. 322-327, 2005.

[3] S. Sharma, S.S. Rajput, L.K. Mangotra and S.S. Jamuar, "Applications of A New FGMOS Based CCII in Low Voltage Analog Filters," *Indian Journal of Engineering and Materials Sciences*, Vol. 11, No. 5, pp. 397-400, 2004.

[4] B. Ramasubramanian, "A New Wide Input-Range FGMOS Based Four Quadrant Multiplier with Electrical Error Correction," *Proceeding of the 2007 International Symposium on Circuits and Systems (ISSCS 07)*, pp. 1-4, 2007.

[5] Y. Wang, Y. Wang, G. Tarr and K. Iniewski, "A Temperature, Supply Voltage Compensated Floating-Gate MOS Dosimeter Using VTH Extractor," *Proceeding of the 5th International Workshop on System-on-Chip for Real-Time Applications (IWSOC 05)*, pp 176-179, 2005.

[6] Y. Wang, G. Tarr and Y. Wang, "A Novel Fully Integrated Floating-Gate MOSFET Radiation Dosimeter Using VTH Extractor," *Proceeding of the 2nd Annual IEEE Northeast Workshop on Circuits and Systems (NEWCAS 04)*, pp 9-12, 2004.

[7] P. Mejía-Chávez, J. C. Snchez-García, J. Velázquez-López, "Differential Difference Amplifier FGMOS for Electrocardiogram Signal Acquisition," *Proceeding of the 2011 8th International Conference on Electrical Engineering Computing Science and Automatic Control (CCE 011)*, pp. 1-5, October 2011.

[8] G. Kapur, S. Mittal, C. M. Markan and V.P. Pyara, "Analog Field Programmable CMOS Current Conveyor," *Proceeding of the 2012 Students Conference on Engineering and Systems (SCES 12)*, pp. 1-6, 2012.

[9] F. Khateb, N. Khatib and D. Kubánek, "Novel Ultra-Low-Power Class AB CCII+ Based on Floating-Gate Folded Cascode OTA," Circuits Systems and Signal Processing, Vol. 31, No. 2, pp. 447-464, 2012.

[10] K. Gupta, M. Bhardwaj, B.P. Singh and R. Choudhary, "Design of Low Power Low Cost True RMS-to-DC Converter," *Proceeding of the 2012 Second International Conference on Advanced Computing & Communication Technologies (ACCT 12)*, pp. 364-367, 2012.

[11] R. Pandey and M. Gupta, "FGMOS Based Voltage-Controlled Grounded Resistor," *Radio Engineering*, Vol. 19, No. 3, pp. 455-459, 2010.

[12] M. Gupta and R. Pandey, "Low-Voltage FGMOS Based Analog Building Blocks," *Microelectronics Journal*, Vol. 42, No. 6, pp. 903-912, 2011.

[13] M. Gupta and R. Pandey, "FGMOS Based Voltage-Controlled Resistor and Its Applications," *Microelectronics Journal*, Vol. 41, No. 1, pp. 25-32, 2010.

[14] V Suresh Babu, P. S. Haseena and M. R. Baiju, "A Floating Gate MOSFET Based Current Reference with Subtraction Technique," *Proceeding of the 2010 IEEE Computer Society Annual Symposium on VLSI (ISVLSI 10)*, pp. 206-209, 2010.

[15] S. Siskos, "FGMOSFET Based Built-In Current Sensor for Low Supply Voltage Analog and Mixed-Signal Circuits Testing," *Proceeding of the 2010 IEEE Computer Society Annual Symposium on VLSI (ISVLSI 10)*, pp. 259-264, 2010.

[16] M-C J. Antonio, G-C Lizeth and G-C Felipe, "Floating-Gate MOSFET Parallel Analog Network for Assignment Problems," *Proceeding of the 2010 7th International Conference on Electrical Engineering Computing Science and Automatic Control (CCE 10)*, pp. 556-559, 2010.

[17] J. M. A. Miguel, A. J. Lopez-Martin, L. Acosta, J. Ramrez-Angulo, and R. G. Carvajal, "Using Floating Gate and Quasi-Floating Gate Techniques for Rail-To-Rail Tunable CMOS Transconductor Design," *IEEE Transactions on*

Circuits and Systems I: Regular Papers, Vol. 58, No. 7, pp. 1604-1614, 2011.

[18] A. J. Lopez-Martin, J. Ramrez-Angulo, R. G. Carvajal, and L. Acosta, "CMOS Transconductors with Continuous Tuning Using FGMOS Balanced Output Current Scaling," *IEEE Journal of Solid-State Circuits*, Vol. 43, No. 5, pp. 1313-1323, 2008.

[19] V. Srinivasan, G. Serrano, C. M. Twigg, and P. Hasler, "A Floating-Gate-Based Programmable CMOS Reference," *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 55, No. 11, pp. 3448-3456, 2008.

[20] Y. L. Wong, M. H. Cohen and P. A. Abshire, "A Floating-Gate Comparator with Automatic Offset Adaptation for 10-bit Data Conversion," *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 52, No. 7, pp. 1316-1326, 2005.

[21] Y. L. Wong, P. Xu, and P. Abshire, "Ultra-Low Spike Rate Silicon Neuron," *Proceeding of the 2007 IEEE Biomedical Circuits and Systems Conference (BioCAS 07)*, pp. 95-98, November 2007.

[22] K. Hasegawa, M. Aoki, T. Yamawaki, S. Tanaka, "Modelling Transistor Variation Using α-Power Formula and Its application to Sensitivity Analysis on Harmonic Distortion in Differential Amplifier," *Analog Integrated Circuits and Signal Processing*, Vol. 72, No. 3, pp. 605-613, 2011.

[23] H. Masuda, T. Kida and S. Ohkawa, "Comprehensive Matching Characterization of Analog CMOS Circuits," *IEICE Transaction on Fundamental of Electronics, Communications and Computer Sciences*, Vol. E92-A, No. 4, pp. 966-975, 2009.

[24] S. Vlassis, S. Siskos, "Current-Mode Non-Linear Building Blocks Based on Floating-Gate Transistors," *Proceeding of the 2000 International Symposium on Circuits and Systems (ISCAS 00)*, pp. 521-524, 2000.

[25] C. Y. Kwok and H. R. Merhrvarz, "Low Voltage and Mismatch Analysis of Quadruple Source Coupled Multi-input Floating-Gate MOSFET Multiplier with Offset Trimming," *Analog Integrated Circuits and Signal Processing*, Vol. 26, No. 2, pp. 141-156, 2001.

[26] S. Vlassis and S. Siskos, "Design of Voltage-Mode and Current-Mode Computational Circuits Using Floating-Gate MOS Transistors," *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 51, No. 2, pp. 329-341, 2004.

[27] A. El mourabit, P. Pittet, G. N. Lu, "A wide-linear range subthreshold OTA based on FG-MOS transistor," *Proceedings of the 2004 11th IEEE International Conference on Electronics, Circuits and Systems (ICECS 04)*, pp. 17-20, 2004.

[28] A. El mourabit, P. Pittet, G. N. Lu, "A low voltage highly linear CMOS OTA," *Proceedings. The 16th International Conference on Microelectronics (ICM 04)*, pp. 700-703, 2004.

[29] S. Vlassis and S. Siskos, "Differential-Voltage Attenuator Based on Floating-Gate MOS Transistors and Its Applications," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, Vol. 48, No. 11, pp. 1372-1378, 2001.

[30] Y. Zhai and P. A. Abshire, "Adaptive Log Domain Filters for System Identification Using Floating Gate Transistors," *Analog Integrated Circuits and Signal Processing*, Vol. 56, No. 1, pp. 23-36, 2008.

[31] V. Suresh Babu, Salini Devi R., Ambika Sekhar and M. R. Baiju, "FGMOSFET Circuit for Neuron Activation Function and Its Derivative," *Proceeding of the 2009 4th IEEE Conference on Industrial Electronics and Applications (ICIEA 09)*, pp. 739-744, 2009.

[32] J. Alfredsson and S. Aunet, "Trade-offs for high yield in 90 nm subthreshold floating-gate circuits by Monte-Carlo simulations," *Proceedings of IFIP VLSI-SOC Conference 2008 (VLSI-SOC 08)*, pp. 1-4, 2008.

[33] R. Banchuin, "Analytical Model of Random Variation in Drain Current of FGMOSFET," *Active and Passive Electronic Components*, Vol. 2015, pp. 1-12, 2015.

[34] R. Banchuin and R. Chaisricharoen, "The probabilistic modeling of random variation in FG-MOSFET," *Proceedings of the 2016 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON'16)*, pp. 1-5, 2016.

[35] Y. Ye, S. Gummalla, C.-C.Wang, C. Chakrabarti and Y. Cao,"Random Variability Modeling and Its Impact on Scaled CMOS Circuits," *Journal of Computational Electronics*, Vol. 9, No. 3, pp. 108-113, 2010.

[36] S. Bhunia and S. Mukhopadhyay, *Low Power Variation Tolerant and Design in Nanometer Silicon*, Springer Science+Business Media, LLC, New York, 2011, ch. 1.

[37] K. Takeuchi, A. Nishida and T. Hiramoto, "Random Fluctuations in Scaled MOS Devices," *Proceeding of the 2009 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD 09)*, pp. 79-85, 2009.

[38] T. Mogami, "Perspective of CMOS Technology and Future Requirement," *Proceedings of SPIE, the International Society for Optical Engineering*, pp. 774802-1-774802-9, 2010.

[39] J. Ramirez-Angulo, G. Gonzlrlez-Altamirano and S. C. Choi, "Modelling Multiple-Input Floating-Gate Transistors for Analog Signal Pro-

cessing," *Proceeding of the 1997 International Symposium on Circuits and Systems (ISCAS 97)*, pp. 2020-2023, 1997.

[40] T. Altiok and B. Melamed, *Simulation Modelling and Analysis with ARENA*, Elsevier, Inc., United States, 2007, ch. 8.

[41] S.A. Klugman, H.H Panjer and G.E. Willmot, *Loss Models: From Data to Decisions*, John Wiley and Sons, Inc., United States, 2008, ch. 13.

[42] W.W. Hines, D.C. Montgomery, D.M. Goldsman and C.M. Borror, *Probability and Statistics in Engineering*, John Wiley and Sons, Inc., United States, 2003, ch. 3.

[43] C. Forbes, M. Evans, N. Hastings and B. Peacock, *Statistical Distributions*, 4^{th} ed., John Wiley and Sons, Inc., New York, 2011, ch. 2.

[44] A.S. Medina-Vazquez, J. de la Cruz-Alejo, F. Gomez-Castaneda and J. A. Moreno-Cadenas, "Low-Voltage Linear Transconductor and A memory Current Using The MIFGMOS Transistor," *International Journal of Electronics*, Vol. 96, No. 9, pp. 895-914, 2009.

[45] G. Cijan, T. Tuma and A. Burmen, "Modelling and Simulation of MOS Transistor Mismatch," *Proceeding of the 6th Eurosim Congress on Modeling and Simulation (EUROSIM 07)*, pp. 1-8, 2007.

Rawid Banchuin received the B.Eng. degree in electrical engineering from Mahidol University, Bangkok, Thailand in 2000, the degree of M.Eng. in computer engineering and Ph.D. in electrical and computer engineering from King Mongkut's University of Technology Thonburi, Bangkok, Thailand in 2003 and 2008 respectively. He was with the Department of Electrical Engineering, Rajamangala Institute of Technology, North-Bangkok Campus since 2003. At the present, he is with the Faculty of Engineering and Graduated School of Information Technology, Siam University, Bangkok, Thailand. Asst. Prof. Dr. Rawid Banchuin is a member of IEEE and council of engineer (Thailand) and has joined the technical program committee of International ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI-NCON), since 2017.

Roungsan Chaisricharoen received B.Eng., M.Eng. and Ph.D. degrees from the department of computer engineering, King Mongkut's University of Technology Thonburi, Bangkok, Thailand. He is an assistant professor at the school of information technology, Mae Fah Luang University, Chiang Rai, Thailand. His research interests are computational intelligence, analog circuits and devices, wireless networks and optimization techniques.

Studying Peripheral Vascular Pulse Wave Velocity Using Bio-impedance Plethysmography and Regression Analysis

Ji-Jer Huang[1], **Yang-Min Huang**[2], and **Aaron Raymond See**[3], Non-members

ABSTRACT

In this study, a simple bioimpedance plethysmography method was employed to measure the pulse wave velocity (PWV) from the radial artery in the wrist to the middle finger. Subsequently, electrocardiography was combined with the bioimpedance method to calculate the PWV from ECG and pulse waves to the middle finger. Experiments were conducted by employing cuffs that temporarily blocks blood flow to produce observable changes in the PWV. Statistical results indicated that temporary blockage of blood flow did not influence the PWV of typical healthy people. Moreover, multiple regression analysis was adopted and analyzed to establish a regression equation for estimating two types of PWV and its relevance with other physiological parameters. Multiple regression analysis indicated that the abdomen circle and height are independent predictors of wfPWV ($r = 0.893$), systolic blood pressure (SBP) and diastolic blood pressure (DBP) are independent predictors of tfPWV ($r = 0.898$). Correlation analysis showed the wfPWV is significantly associated with tfPWV ($r = 0.770$, $p<0.01$).

Keywords: Bioimpedance, Pulse Wave Velocity, Multiple Regressions

1. INTRODUCTION

According to a report released by the World Health Organization in 2014, cardiovascular disease, cancers, chronic respiratory diseases, diabetes, and other non-communicable diseases accounted for 37%, 27%, 8%, 4%, and 24%, respectively, of all causes of death among people under the age of 70 years in 2012 [1]. The report elucidated the relevant threat of cardiovascular diseases to the health of the general population in modern society. Moreover, vascular sclerosis is an irreversible disease. Once the blood vessels are affected by aging, one can only maintain its current state and prevent further deterioration. The process is irreversible and it is impossible to have the former flexibility of the blood vessels. This loss of flexibility can be attributed to the gradual proliferation of atherosclerotic plaque that thickens the vascular wall resulting in the hardening of the walls. Consequently, it narrows the blood vessels causing reduced blood flow or the rupturing of plaques that form thrombus leading to blocked arteries, tissues or organs that may lead to ischemic necrosis. In the event that the carotid artery or the cerebral artery is obstructed, it can result in a stroke while blocked coronary arteries can issue to myocardial infarction. Therefore, prevention is better than cure especially in terms of the hardening of blood vessels. However, many are unable to perform regular check-ups due to their busy lifestyle. And once the symptoms are aggravated, it would be difficult to restore the original health state of the person. In addition to diagnosing the illness, a check-up at this time may control the disease and inhibit further deterioration.

Preventive measures for cardiovascular diseases would be beneficial for early detection of symptoms and have real-time cardiovascular monitors can be adopted to detect cardiovascular disease and estimate the severity of arteriosclerosis. A specific form of arteriosclerosis called atherosclerosis is known as the hardening of the arteries which begins at the intima of the artery [2]. Currently, various methods have been developed for the detection of the degree of arteriosclerosis. A particular method for analyzing the situation is the pulse wave velocity (PWV) method. PWV is by definition the propagation velocity of the arterial pulse and is also the distance traveled by a wave divided by the time required for a wave to travel that distance. PWV measurement involves either invasive or non-invasive methods. Invasive measurement methods include angiography, fiberoptic angioscope [3], intravascular ultrasound (IVUS) [4] and MRI [5]. This type of measurement can be used to directly investigate the presence of fat accumulation or intimal hyperplasia on the artery. However, it is likely to cause fear and health risks to the patient. Furthermore, it is also very time-consuming and requires the operator to have professional knowledge and training to mete out an appropriate judgment. Due to the inherent inconvenience and limitations of the invasive method, current clinical medicine practice uses non-invasive measurement method to assess the degree of early vascular disease[6]. Non-invasive measure-

Manuscript received on August 16, 2016 ; revised on March 27, 2017.

Final manuscript received on April 22, 2017.

[1,2,3] The authors are with Department of Electrical Engineering, Southern Taiwan University of Science and Technology, Tainan, Taiwan, E-mail: jjhuang@stust.edu.tw, ma020227@stust.edu.tw and aaron.r.see@gmail.com

ment methods include ultrasound [7], cuff [8], pressure sensing, infrared sensing [9], pulse rate, among others. Many studies have found that vascular wall elasticity is affected by personal health factors or disease [10, 11], and the measured aortic pulse-wave velocity (PWV) changes due to the elasticity of the arterial wall. In hardened arteries, PWV is faster, but in more elastic arteries, PWV is reduced. Arterial elasticity has been associated with various cardiovascular diseases [12]; therefore, PWV parameters may serve as a useful indicator in the assessment of human arteries. Consequently, current clinical medical practitioners frequently employ the non-invasive measurement method (e.g., ultrasound, cuff, and infrared sensing methods) to estimate early vascular lesions.

The most frequently used non-invasive approach is the cuff method, for which a cuff is used to measure the pulse activities of the upper arm and ankle and calculate brachial-ankle PWV and ankle-brachial index. A previous study [13] indicated that brachial-ankle PWV measurement is a valid approach featuring high reproducibility and reported that the measured PWV of patients with cardiovascular diseases are substantially higher than those of healthy people [14,15]. However, the difficulty of estimating vessel length creates problems in practically applying this method. Currently, there exist numerous researchers on PWV, but most are directed toward the study of baPWV or aortic PWV with limited literature discussing peripheral blood vessels. Subsequently, prior to the hardening of the aorta, peripheral vascular blood flow should show changes. Therefore, deducing the health of the aorta by long-term monitoring of the peripheral blood vessels will be good news for patients with cardiovascular disease. When blood flows into blood vessels, aside from producing mechanical expansion effects functionally, a change in the bioelectrical impedance is generated. Bioimpedance measuring is a measurement technique that is useful because it is cost-effective and portable and poses no radioactive injury to patients. Research teams have measured PWV and heart rate by employing the bioimpedance technique at the forearm [16], indicating that bioimpedance can be used to measure PWV. For these reasons, the current research utilized changes in bioelectrical impedance by measuring the pulse rate through the bioelectrical impedance changes from the arm radial artery to the finger. Simultaneously, the obtained rate and physiological parameters (such as anthropometric parameters, heart and blood pressure) underwent regression analysis to obtain an estimated formula for understanding the relation between PWV and physiological parameters.

2. SYSTEM ARCHITECTURE

2.1 Bioimpedance Plethysmography Measuring System

Fig. 1 presents the bioimpedance plethysmography measuring system that includes the bioimpedance circuit, cuff control, pressure sensing circuit, and electrocardiography (ECG). The received physiological data are converted using an analogue-to-digital converter and subsequently transmitted to the computer; the user interface and data storage function was developed using National Instruments (NI) LabWindows and signal analysis to obtain PWV parameters was accomplished using the Matlab software.

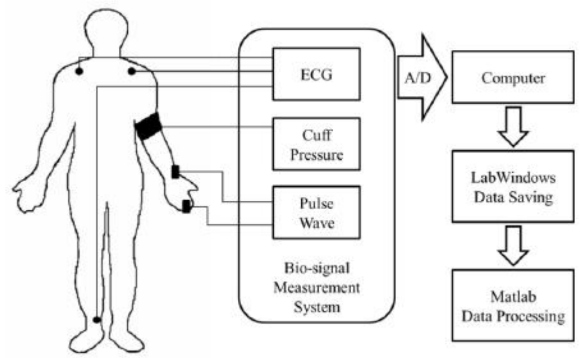


Fig.1: *System architecture of physiological signal measurement.*

2.2 Bioimpedance Circuit and Electrode

When arteries pulsate from the heart's pumping of blood, the vascular caliber changes. The change causes an increase in intravascular blood volume that alters the bio- impedance. Therefore, vessel pulsation can be considered as a parameter to determine bioimpedance signal that varies as blood flow changes. The bioimpedance electrodes were constructed using eight copper electrodes with the following dimensions: size: 10×3 mm, thickness: 0.2 mm, and inter-electrode separation: 3 mm. These were affixed to the radial artery and the middle finger to measure PWV as shown in Fig.2. A quad-electrode method was used to create constant current within two electrodes and apply the created current into the participants. The voltage differences between the other two electrodes affixed to the participants were also measured [8]. In this experiment, a constant current source was operated at 50 kHz frequency and a 1 mA current. One of two outer electrodes was inserted into the constant current source and the other was used as the constant current sink. The two middle electrodes were used to measure the changes in pulse wave impedance.

The electric signals received by the electrodes were transmitted to an instrumentation amplifier (AD620, Analogue Devices) with a high common-mode rejection ratio to amplify differential signals. The AD620 operational amplifier is advantageous because of the following properties: high input impedance, high common-mode rejection ratio, and low noise. The measuring alternating current signal was fed to the

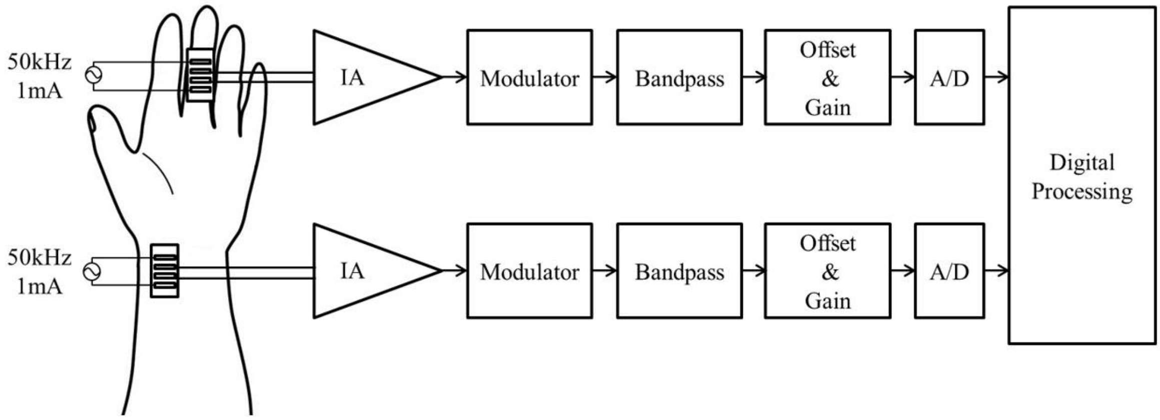


Fig.3: *Block diagram of the bioimpedance circuit.*

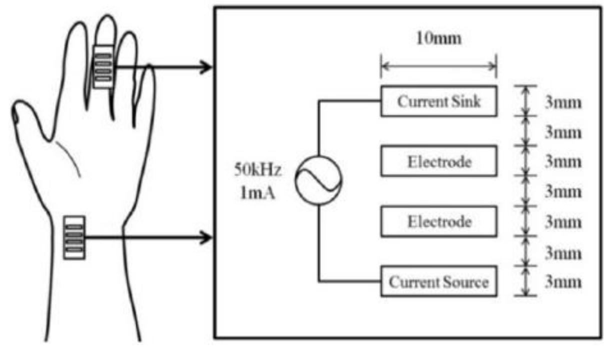


Fig.2: *Schematic diagram of a bioimpedance electrode.*

modulator, and then the direct current signal was obtained. The signal was sent to the amplifier; thus, the pulsations of arteries reflected the magnitude of the voltage modulated. Subsequently, the bandpass filter was used to filter all noise other than the pulse frequency. The measurement system used a high-pass filter with a cut-off frequency of 0.48 Hz and a low-pass filter with a cut-off frequency of 5.3 Hz. Since the electric noise frequency of 60-Hz was higher than the cut-off frequency range of the low pass filter, no band exclusion filter was used in this study. After the aforementioned circuits were used, level adjustment and circuit amplifiers were implemented to transmit signals to a 12bits analogue-to-digital converter to capture the digital data. Fig. 3 presents the circuit block diagram.

2.3 Cuff Pressure Sensing Circuit

The cuff pressure sensing circuit detects air pressure changes in the cuff and is measured with the aid of a differential pressure sensing element (SCC05DN, Honeywell Sensing and Control Co., Ltd). A pressure sensing component was adapted to measure cuff pressure changes to reduce the influence of temperature. The sensor also incorporates a temperature compensation circuit designed using the Wheatstone bridge principle. When the pressure was applied to the sensor, changes were generated in the electrical resistance of the bridge as shown in Fig. 4. The generated voltage difference is the output voltage and is defined as

$$V_{os} = V_B \cdot \frac{2\Delta R}{R} \tag{1}$$

where V_{os} indicates offset voltage (the output voltage was 0 when the pressure was applied), and V_B indicates offset bridge voltage.

Electrical resistance is directly proportional to pressure, and the output voltage formula is as follows

$$V_o = s \cdot p \cdot V_B + V_{os} \tag{2}$$

where V_o, s, and p denote output voltage, sensitivity coefficient, and pressure, respectively.

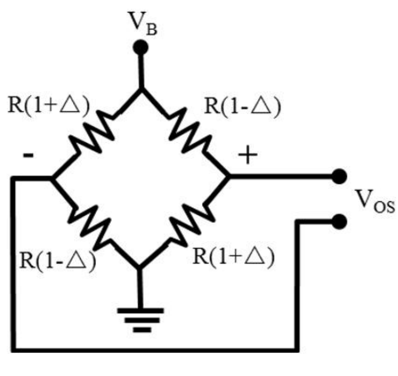


Fig.4: *The cuff pressure sensing circuit diagram.*

2.4 Electrocardiogram Measurement Instrument

The Ultraview SL2200, manufactured by Spacelabs Healthcare, was selected to measure and capture ECG signals. This equipment can output measured physiological signals through the analog output method, facilitate back-end data processing. The aforementioned data received by the circuit were stored by the DAQ card USB-6351, manufactured by National Instruments, in the computer, and the DAQ card. The DAQ card has 16 bits ADC resolution that can accommodate 8 differential mode or 16 single-ended connections. Its sample rates are 1.25MHz and 1.0MHz for single channel and multichannel, respectively. In our system, the sampling rate was set at 10 kHz, and data were recorded into a raw data file, and simultaneously displayed, in order to ensure the authenticity of the signals. After the measurement was completed, the raw data file was loaded onto customized Matlab program for back-end signal processing.

2.5 Signal Processing

Before data were processed, the arm length, shoulder length, and length from the wrists to middle fingers of the participants were inputted for PWV calculation as shown in Fig. 5. After the participant data were read, signals were processed using a digital filter, and the pulse signal was processed using a third-order Butterworth low-pass filter with a cut-off frequency of 5.3 Hz. The ECG used a third-order high-pass filter and a low-pass filter with cut-off frequencies of 0.1 Hz and 20 Hz, respectively. The digital filter can further eliminate the motion noise. Subsequently, the trough of the pulse wave was captured. The R and T waves of the ECG were then obtained, and the PWV and heart rate were calculated. Finally, the results were shown on the screen, and the PWV was recorded for statistical analysis.

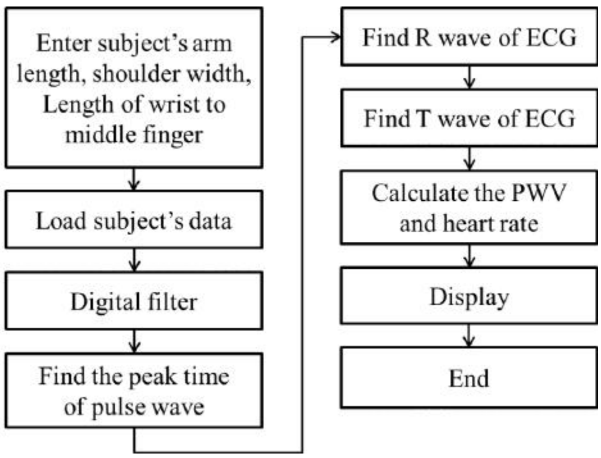


Fig.5: *Block diagram of signal processing.*

2.6 Graphic User Interface

Fig. 6 displays the graphic user interface (GUI) developed using the LabWindows software (National Instruments). The far left boxes display the channel parameters and sample rate that can be adjusted accordingly. In order to ensure that signal is not distorted, the sample rate is set at 10k Hz. Data are obtained and simultaneously stored in data files, which is then loaded by the system to the Matlab program for back-end analysis. The graphic interface on the left displays two electrical impedance pulse waves and graphic interfaces on the right exhibits the ECG signal and cuff pressure.

3. RESULTS AND DISCUSSION

3.1 Experimental Design

Demographic information of the participants was obtained at the beginning of the experiment that includes height, weight, body mass index (BMI), body fat, length from wrist to the middle finger, arm length, shoulder width, waistline, and blood pressure. These demographic data were then combined with PWV parameters for statistical analysis. Table 1 provides the demographic information of the ten participants in this experiment. After the demographic information was measured, the participants were asked to rest for 5 minutes, after which the measurement electrodes were affixed to the participants' bodies. The total measurement time was 5 minutes. During the first 2 minutes, the PWV and ECG signals of the participants were collected while the participants were in a relax state. Thereafter, the cuffs were inflated until the participant's blood flow was blocked; this was maintained for 5 seconds before the cuff was deflated, releasing the cuff pressure. Subsequently, participant PWV and ECG signals were measured for 2 minutes. Fig. 7 presents the flow chart of this experiment.

Table 1: *Basic physiological parameters of the participants.*

	Mean±SD	Max/Min
Height(cm)	169.5±4.27	176/163
Weight(kg)	67.29±7.98	81.8/54.6
Arm length(cm)	72.85±2.53	76/68
Shoulder width(cm)	44.4±2.33	48/41
Waistline (cm)	91.5±4.77	97.5/83
SBP(mmHg)	121.2±6.88	131/111
DBP(mmHg)	71.1±7.61	83/60
BMI(kg/m^2)	23.53±2.43	78/56
Body fat(%)	18.63±4.06	27.65/19.34
Age(years)	22.9±1.79	27/21
Length of wrist to middle finger(cm)	16.41±1.16	18/14

3.2 PWV Analysis

After data were measured, 2-minute of raw data of pulse wave and ECG were collected, before the cuffs were inflated and after they were deflated, for analysis. Fig. 8 shows measuring electrodes and the image of the fixed electrode on the hand. Several defined distances are shown in Fig. 9. This experiment defined three types of PWV parameters: (1) the PWV

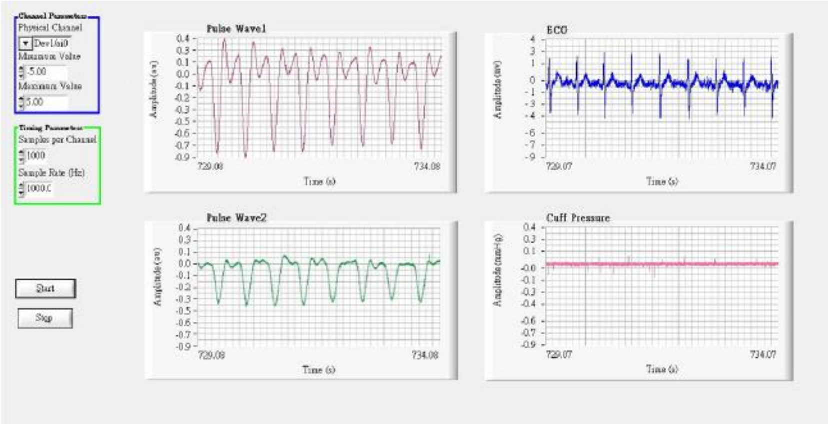


Fig. 6: *Data display and storage interface using LabWindows.*

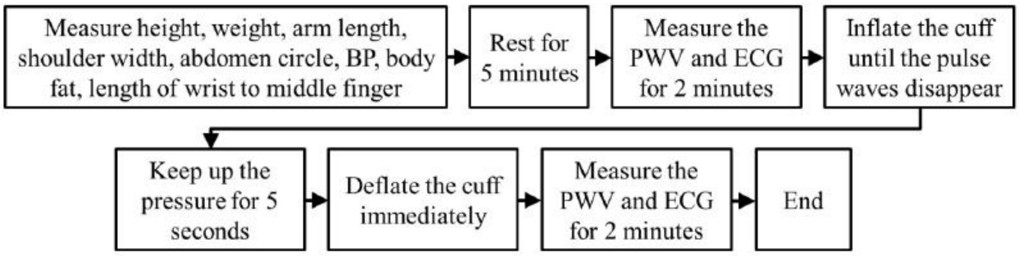


Fig. 7: *Experiment flowchart.*

from the radial artery in the wrist to the middle finger (wfPWV); (2) the PWV from the ECG R wave to the middle finger (rfPWV); and (3) the PWV from the ECG T wave to the middle finger (tfPWV) which are computed as follows

$$\mathrm{wfPWV} = \mathrm{D_{wf}}/\Delta \mathrm{T_{wf}} \tag{3}$$

where $\mathrm{D_{wf}}$ refers to the distance from the electrode of the radial artery in the wrist to the electrode of the middle finger, and $\Delta\mathrm{T_{wf}}$ indicates the pulse wave time difference from the radial artery to the middle finger shown in equation (3) and

$$\mathrm{rfPWV} = \mathrm{D_{rf}}/\Delta \mathrm{T_{rf}} \tag{4}$$

Since the distance from the aorta to the middle finger was difficult to measure, the half the shoulder width plus arm length equals the distance defined as $\mathrm{D_{rf}}$, and $\Delta\mathrm{T_{rf}}$ represents the time difference of the peak ECG R wave to the middle finger showed in equation (4). The definition of tfPWV is as follows

$$\mathrm{tfPWV} = \mathrm{D_{tf}}/\Delta \mathrm{T_{tf}} \tag{5}$$

where $D_{tf} = D_{rf}$, and ΔT_{tf} indicates the time difference of the peak ECG T wave to the middle finger. Fig. 10 presents the pulse wave of the radial artery in the wrist, pulse wave of the middle finger, ECG, and a waveform with a length of approximately 3 seconds.

Table 2 shows the mean, standard deviation, max-

(a) (b)

Fig. 8: *(a) Measuring electrodes, (b) Photo of fixed electrode on the wrist and finger.*

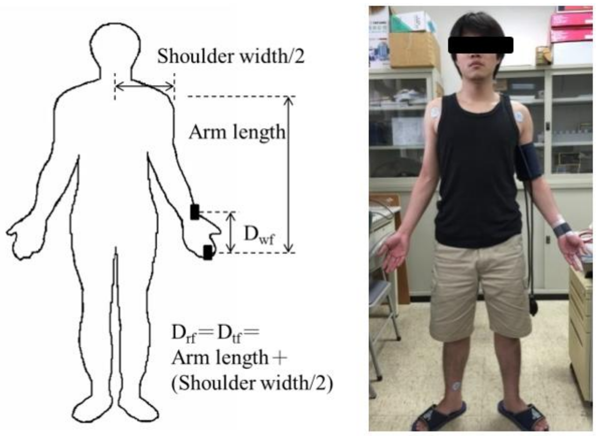


Fig. 9: *Defined distances of D_{wf}, D_{rf}, and D_{tf}.*

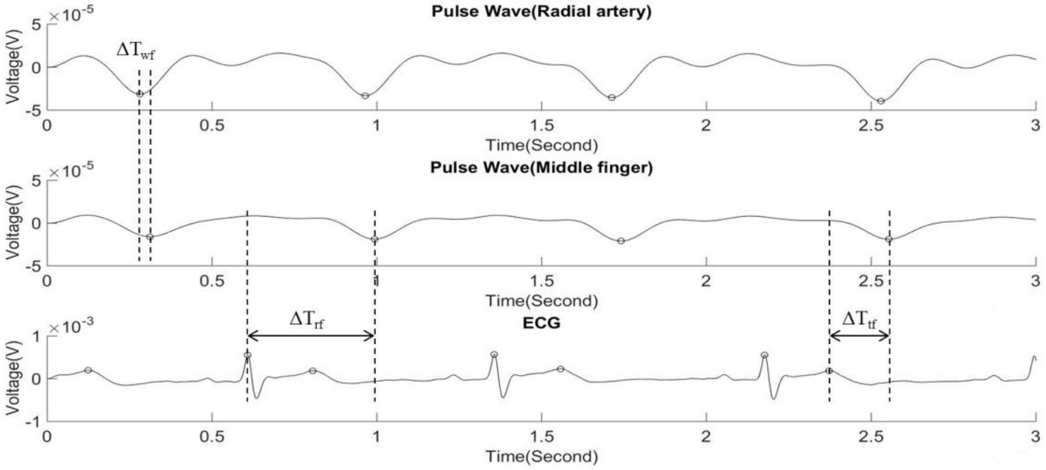


Fig.10: *Measured pulse waves and ECG.*

imum value and minimum value of the three types of PWV of the participants before the cuffs were inflated and after they were deflated. According to Table 2, the numerical value of rfPWV was lower than that of wfPWV. Furthermore, the rfPWV distance was the length from the heart to the middle finger, representing the path from the aorta to the peripheral artery. The cardiac action corresponding to the ECG signal must be determined in order to evaluate whether rfPWV and tfPWV were substantial. According to medical literature, the phase in which the R wave occurs is the time immediately before the heart enters the systolic phase, during which the heart has not yet contracted. In other words, the heart does not pump blood at this phase. T-waves occur immediately after the heart completes the systolic phase and immediately before it enters the diastolic phase. During this phase, the aorta pumps blood systemically because of the recent contraction. The pulse wave measured in this study could be verified when the troughs between two pulse waves occurred after the T wave. Because the definition of PWV is the speed at which a pulse wave occurs, the time reference point defined by rfPWV is unreasonable. Therefore, rfPWV was discarded, and only wfPWV and tfPWV were used for analysis. The results of this analysis were as follows.

Table 2: *Three types of PWV before cuff inflation and after cuff deflation.*

Status	Mean±SD	Max/Min
wfPWV before inflating	5.33±1.18	7.0/3.31
wfPWV after deflating	5.22±1.25	7.38/3.15
rfPWV before inflating	2.28±0.14	2.47/2.04
rfPWV after deflating	2.29±0.14	2.47/2.03
tfPWV before inflating	5.36±0.79	7.24/4.45
tfPWV after deflating	5.36±0.92	7.68/4.55

In this experiment, statistical analysis was used to determine whether PWV differed significantly before inflation and after deflation, and the paired sample t test was adopted to compare the results. The paired sample t-test employs data of the same individual from a small sample (sample number < 30), which were measured at two-time points. The confidence interval was 95%. According to Table 3, wfPWV and tfPWV were not significant. In other words, a temporary blood flow blockage caused no significant difference in the PWV of healthy people.

Multiple regression analysis was conducted to elucidate whether the physiological parameters (e.g., height and weight) influenced PWV and could be used to estimate PWV. Multiple regression analysis is a statistical method that uses several parameters to predict reaction variables (i.e., PWV in this study). However, these physiological parameters were correlated to PWV. Thus, the stepwise selection was selected as the regression model. This model introduces one variable at a time and examines all previously selected variables after selecting a new variable into the equation to ensure that the variables retain statistical significance. Therefore, the regression equation for PWV can be obtained by using these variables. Before cuff inflation, PWV was equal to the PWV measured at an ordinary state. Thus, in the subsequent evaluations, PWV obtained before cuff inflation was adopted to conduct statistical analysis. After analysis, the variables selected for wfPWV were waistline and height, with $R = 0.893$ and $R^2 = 0.798$. The regression equation is as follows

$$\text{wfPWV} = -0.195 \cdot X_{ab} + 0.168 \cdot X_h - 5.315 \quad (6)$$

where X_{ab} and X_h denote waistline and height, respectively. The variables selected for tfPWV were

Table 3: *Paired sample t-test for wfPWV and tfPWV.*

Status	Mean	SD	Std. Error	Lower	Upper	t	df	Sig.(two-tailed)
wfPWV before inflating wfPWV after deflating	0.114	0.485	0.153	-0.233	0.462	0.745	9	0.475
tfPWV before inflating tfPWV after deflating	0.001	0.274	0.087	-0.195	0.196	0.007	9	0.995

systolic blood pressure (SBP) and diastolic blood pressure (DBP), with R = 0.898 and $R^2 = 0.806$. The regression equation was calculated to be

$$\text{tfPWV} = 0.109 \cdot \mathrm{X_{SBP}} - 0.05 \cdot \mathrm{X_{DBP}} - 4.226 \quad (7)$$

X_{SBP} represents SBP and X_{DBP} denotes DBP.

Table 4 presents the correlation of wfPWV and tfPWV with various physiological parameters and indicates that wfPWV and tfPWV are significantly correlated. Several physiological parameters were also significant. SBP was significantly correlated to wfPWV and tfPWV. In addition, after the multiple regression analysis was conducted, two physiological parameters were used as the input variables for the equation. According to the regression model selected in this study, even though the other parameters could be surmised to be correlated to PWV, the multiple regression analysis indicated that they were not significant after these variables were incorporated. Moreover, the other physiological parameters may exhibit severe collinear relationships, which may influence the results of the regression analysis.

Table 4: *Correlation coefficient table for wfPWV and tfPWV.*

Parameters	wfPWV	tfPWV
wfPWV	1	0.77**
tfPWV	0.77**	1
Height	0.445	0.354
Weight	-0.333	-0.273
Arm length	0.426	0.778*
Shoulder width	0.08	0.092
Waistline	-0.699*	-0.389
SBP	0.673*	0.728*
DBP	-0.34	-0.152
Heart rate	-0.252	-0.763
BMI	-0.624	-0.43
Body fat	-0.648*	-0.6
Age	-0.104	0.045
Length of wrist to middle finger	0.395	0.31

4. CONCLUSIONS

In this study, pulse waves were measured from the finger, indicating that pulse waves can be measured from peripheral areas of the body in addition to the radial artery in the wrist. The pulse waves measured from these two areas were combined with the ECG to calculate the three types of PWV (i.e., wfPWV, rfPWV, and tfPWV). Considering the stages of a cardiac cycle, this study determined that the time defined by rfPWV was inappropriate for evaluating PWV. For the statistical analysis, a paired sample t-test was employed, and temporary blood flow blockage was shown to have no substantial influence on the PWV of healthy people. Moreover, a multiple regression analysis was employed to establish a regression equation that can estimate the two types which are wfPWV and rfPWV and determine the correlation between the physiological parameters (e.g., height and weight) and PWV. Results showed that the independent variables for the regression equation of wfPWV are waistline and height (r = 0.893). While the independent variables for the regression equation of tfPWV are systolic and diastolic blood pressure (r = 0.898). In addition, the correlation between other physiological parameters such as height and PWV was shown. wfPWV and tfPWV both have significant correlation (r = 0.770, $p < 0.01$) and are also associated with waist circumference, systolic blood pressure, and body fat (r = -0.699, r = 0.673 and r = -0.648, $p < 0.05$). tfPWV and arm length (r = 0.778, $p < 0.01$), systolic blood pressure and heart rate are correlated (r = 0.728 and r = -0.763, $p < 0.05$).

During the experiment, action noise strongly influenced signal quality. Thus, reduction or elimination of action noise is necessary to improve the measurement system. Moreover, in this experiment, only the PWV of healthy people was measured. Future experiments should incorporate the measurement data of patients with cardiovascular diseases for comparison with the data of healthy people in order to increase the number of reference data that can be applied for diagnosing peripheral vascular arteriosclerosis.

ACKNOWLEDGEMENT

This work is partly supported by the National Ministry of Science and Technology, R.O.C., under the MOST Grant No. MOST 104-2221-E-218-045-.

References

[1] Global status report on noncommunicable diseases 2014, Geneva: World Health Organization, 2015.

[2] C. Lahoz and J. M. Mostaza, "Atherosclerosis as a systemic disease," *Rev. Esp. Cardiol.*, Vol. 60, No.2, pp. 184-195, 2007.

[3] F. Litvack, W. S. Grundfest, M. E. Lee, R. M. Carroll, R. Foran, A. Chaux, G. Berci,H. B. Rose, J. M. Matloff, J. S. Forrester., "Angioscopic visualization of blood vessel interior in animals and humans," *Clin Cardiol*, Feb, Vol. 8, No.2, :pp. 65-70. 1985.

[4] J. E. Kwon et al., "Relationship between coronary artery plaque composition by virtual histology intravascular ultrasound analysis and brachial-ankle pulse wave velocity in patients with coronary artery disease," Coron Artery Dis., Vol. 22, No.8, pp. 565-569, Dec. 2011.

[5] D. A. Woodrum, A. J. Romano, A. Lerman, U. H. Pandya, D. Brosh, P. J. Rossman, L. O. Lerman, and R. L. Ehman, "Vascular wall elasticity measurement by magnetic resonance imaging.," Magn. Reson. Med., Vol. 56, pp. 593-600. 2006.

[6] M. F. O'Rourke et al., "Clinical applications of arterial stiffness; definitions and reference values," *Am. J. Hypertens.*, vol. 15, pp. 426-444, May. 2002.

[7] J. Calabia et al., "Doppler ultrasound in the measurement of pulse wave velocity: agreement with the Complior method," *Cardiovasc Ultrasound*, 9:13, 2011.

[8] J. Sugawara et al., "Brachial-ankle pulse wave velocity: an index of central arterial stiffness?," *J. Hum. Hypertens.*, Vol. 19, No.5, pp. 401-406, May 2005.

[9] L. Yang et al., "Study of Pulse Wave Velocity Noninvasive Detecting Instrument Based on Radial Artery and Finger Photoplethysmography Pulse Wave," *IITAW'08*, pp. 705-708, Dec., 2008.

[10] J. C. Bramwell and A. V. Hill, "The velocity of the pulse wave in man," *Proc. R. Soc. Lond. (Biol)*, Vol. 93, pp. 298-306, 1922.

[11] K. Hayashi, H. Handa, S. Nagasawa, A. Okumura, and K. Moritake, "Stiffness and elastic behavior of human intracranial and extracranial arteries," *Journal of biomechanics*, Vol. 13, No.2, pp. 175-184, 1980.

[12] E. D. Lehmann, "Pulse wave velocity as a marker of vascular disease," *The Lancet*, Vol. 348, pp. 744, 1996.

[13] A. Yamashina et al., "Validity, reproducibility, and clinical significance of noninvasive brachial-ankle pulse wave velocity measurement," *Hypertens.Res.*, Vol. 25, pp. 359-364, May. 2002.

[14] M. Munakata et al., "Utility of automated brachial ankle pulse wave velocity measurements in hypertensive patients," *Am. J. Hypertens.*, Vol. 16, pp. 653-657, Aug. 2003.

[15] H. J. Park et al., "The relationship between the acute changes of the systolic blood pressure and the brachial-ankle pulse wave velocity," *Korean J. Intern. Med.*, Vol. 22, No. 3, pp. 147-151, Sep. 2007.

[16] M. C. Cho, J. Y. Kim and S. H. Cho, "A Bioimpedance Measurement System for Portable Monitoring of Heart Rate and Pulse Wave Velocity Using Small Body Area," *Circuits and Systems*, pp.3106-3109, 2009.

Ji-Jer Huang received the B.Sc. degree in electrical engineering in 1992 from Feng Chia University. He received the M.S. and PhD degrees both in biomedical engineering in 1994 and 2001 from the National Cheng Kung University, Tainan, Taiwan. He did research in the field of the optoelectronic instrument at Instrument Technology Research Center, National Applied Research Laboratories before 2006. His current research interests include electrical impedance imaging, development of noninvasive biomedical measurement technologies and microcomputer based biomedical instrumentation.

Yang-Min Huang received the B.S. and M.S. degrees both in Electrical Engineering in 2011 and in Biomedical Engineering in 2015 from Southern Taiwan University of Science and Technology, Tainan, Taiwan. He is an electrical engineer to research about measurement technologies at JNC Technology since 2016.

Aaron Raymond See See was born in Manila, the Philippines. He received the B.S. degree in Electronics and Communications Engineering from De La Salle University - Manila, the Philippines in 2006, and the M.S. and PhD degrees in Electrical Engineering with a Major in Biomedical Engineering from Southern Taiwan University of Science and Technology (STUST), Taiwan in 2014. He did his postdoctoral research in the field of Neuroscience at the Brain Research Center in National Tsing Hua University. In 2015, he was Deputy Project Manager at SMOBIO Inc. and subsequently as Marketing Manager at Tong Fang Watch Co., before joining STUST in 2016. He is currently serving as a as Project Assistant Professor in the Department of Electrical Engineering. His research interests are in biomedical signal processing, neuroscience, assistive devices and wearable devices.

Output Voltage Control of the SP Topology IPT System based on Primary Side Controller Operating at ZVS

Supapong Nutwong[1], **Anawach Sangswang**[2],
Sumate Naetiladdanon[3], and **Ekkachai Mujjalinvimut**[4], Non-members

ABSTRACT

This paper presents a technique to control the output voltage of series-parallel (SP) topology inductive power transfer (IPT) system by using only a controller, located on the primary side. This reduces the cost, size, complexity and loss of the system compared with conventional IPT dual-side controllers. Asymmetrical duty cycle control (ADC) of full-bridge inverters has been used to control the DC output voltage to its designed value. In addition, a zero voltage switching (ZVS) operation can be obtained at all power levels by varying the switching frequency of the inverter. Theoretical analysis has been performed through mutual inductance coupling model and verified by computer simulation. Experimental results of the circular magnetic structure IPT system with adjustable air-gap confirm the validity of the proposed controller. The system efficiency is improved throughout the operation and the improvement becomes obvious as the output power is decreased.

Keywords: IPT System, SP Topology, Primary Side Controller, ADC Control, ZVS Operation

1. INTRODUCTION

Recently, inductive power transfer (IPT) technique has received much interest and widely adopted. This technique allows the contactless power transfer from the source to the load over a relatively large air gap via magnetic fields. Due to the use of inductive coupling, there is no direct physical and electrical contact between the electric source and load, which makes them safer and more convenient compared with the conventional conductive power transfer. This new technology has been used successfully in many applications such as material handling system, public transport system, EV battery charger, consumer electronics and medical implantable devices [1-3]. Fig. 1 shows a typical IPT system. The primary and secondary compensations are used to compensate the reactive power required by the coils. By doing that, the power transfer capability or system efficiency can be increased. Thus, compensation capacitors may be connected in series (S) or parallel (P) to the primary coil, secondary coil, or both coils. The commonly used compensation topologies are series-series (SS), series-parallel (SP), parallel-series (PS), and parallel-parallel (PP) as illustrated in Fig. 2 [4-6]. Parallel compensation in a primary side as in PP and PS requires an additional inductor to form a current source, which increase the cost and complexity of the system. The key benefits of the SP topology compared with the SS topology are that the inverter current and the voltage stress in primary capacitor are lower, for the same output voltage [7]. The SP topology is selected in this paper.

To control the DC output voltage of the IPT system, the secondary side controller is required to control the switch mode DC-DC converter located on the load side, also known as the conventional IPT dual-side controller [8,9] which increases the cost, size, complexity, and loss of the system. This may cause excessive heat and cannot be used in many applications such as implantable medical devices [10]. To overcome mentioned drawbacks, the output voltage control of the IPT system using only a controller located on the primary side has been introduced with the SS topology [11-13]. For SP topology, a load variation may result in the non-ZVS operation if the fixed frequency control is adopted [14]. The turn-on switching from the non-ZVS operation unavoidably affects the inverter efficiency. This is due to the change in the reflected reactance in the primary circuit, related to the primary resonant frequency. Because of its simple control and easy implementation, the ADC control is a common technique chosen in many full-bridge inverters. However, non-ZVS operation may occur during the adjustment of the duty cycle of the inverter voltage.

This paper presents an output voltage control method of SP topology IPT systems using only the primary side controller where the variable frequency ADC control of a full-bridge inverter is applied. The DC output voltage can be controlled to its designed

Manuscript received on August 15, 2016 ; revised on April 4, 2017.
Final manuscript received on June 6, 2017.
[1,2,3,4] The authors are with Department of Electrical Engineering, King Mongkut's University of Technology Thonburi, Bangkok, Thailand., E-mail: supapong.nut@mail.kmutt.ac.th, anawach.san@kmutt.ac.th, sumate.nae@kmutt.ac.th and ekkachai.muj@kmutt.ac.th

value by adjusting the duty cycle of the inverter voltage while the ZVS operation is maintained by varying the switching frequency. The removal of the DC-DC converter and controller in the secondary circuit reduces the cost, size and complexity of the system. The efficiency can be improved over the fixed frequency ADC control. Theoretical analysis of the non-ZVS mode due to load and angle variations has been performed, and verified by computer simulation and experimental results.

2. THEORETICAL ANALYSIS

The non-ZVS operation due to the load and alpha angle (α) variations has been analyzed in this section based on the mutual inductance coupling model of the SP topology IPT system. In addition, the minimum switching frequency that achieve ZVS operation for a given load resistance and α angle has been introduced.

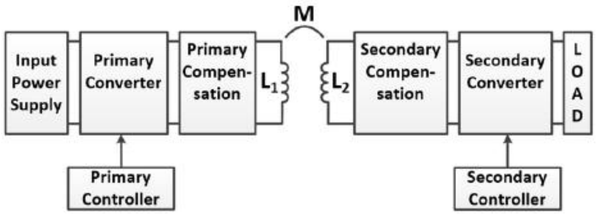


Fig.1: *Typical inductive power transfer (IPT) system.*

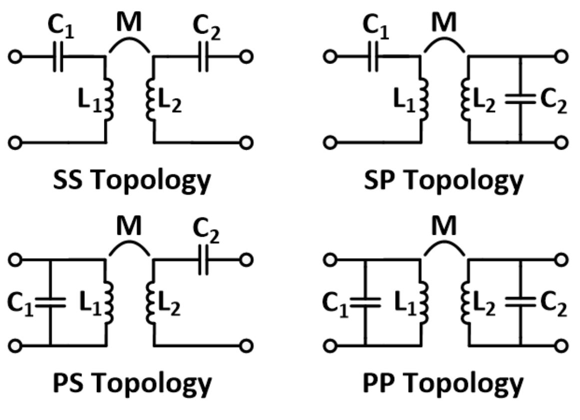


Fig.2: *Commonly used compensation topologies.*

2.1 Mutual inductance coupling model

The commonly used model for IPT system with mutual inductance coupling model is illustrated in Fig. 3 (a) [15]. The primary circuit is supplied by a sinusoidal voltage source (V_1). The compensation capacitors, C_1 and C_2, are connected in series with the primary coil and in parallel with the secondary coil, respectively, to form an SP topology. The coupled coil has mutual inductance M and self-inductance L_1 and L_2. The winding resistance of both coils are denoted as R_1 and R_2. R_L is the load resistance. The induced voltage in both coils are shown in series with the coil's inductance.

Since $R_2 \ll \omega L_2$, the secondary coil impedance $\vec{Z}_2 \approx j\omega L_2$. Using the source transformation, the resistance R_2 is combined with the load resistance R_L as,

$$R'_L = \frac{R_L(R_2^2 + X_{L2}^2)}{R_2 R_L + (R_2^2 + X_{L2}^2)} \quad (1)$$

The system is transformed from Fig. 3 (a) to 3 (b). At resonant frequency, the load voltage can be obtained by,

$$\vec{Z}_L = \frac{M\vec{I}_1 R'_L}{L_2} \quad (2)$$

Thus, the secondary current in Fig. 3 (a) is calculated as,

$$\vec{I}_2 = \frac{j\omega_{02} M\vec{I}_1 - \vec{V}_L}{R_2 + j\omega_{02} L_2} \quad (3)$$

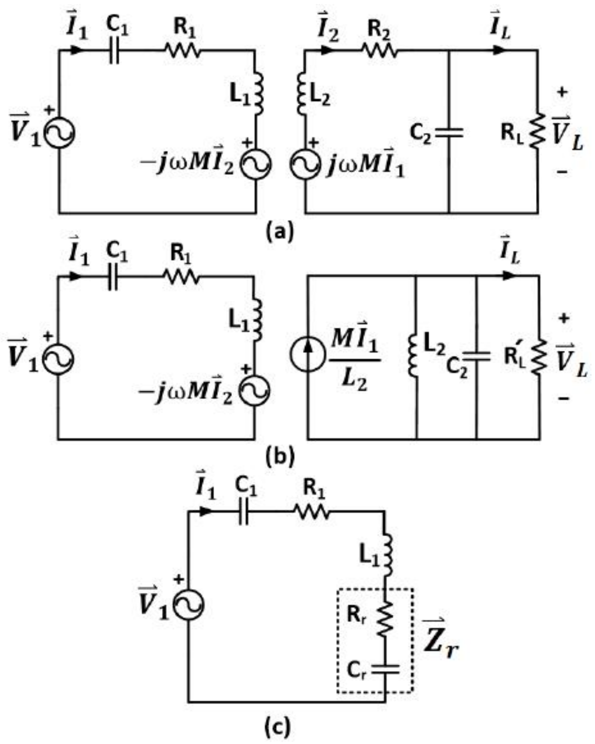


Fig.3: *Mutual inductance coupling model of the SP topology IPT system.*

The reflected impedance referred to the primary circuit can be obtained by dividing the primary induced voltage by the primary current as [15],

$$\vec{Z}_r = \frac{-j\omega_{02} M\vec{I}_2}{\vec{I}_1} \quad (4)$$

Substituting (3) into (4) resulted in,

$$\vec{Z}_r = \frac{\omega_{02}^2 M^2 (R_2 + R_L') - j\omega_{02} M^2 (\omega_{02}^2 L_2 - R_L' R_2 / L_2)}{R_2^2 + \omega_{02}^2 L_2^2} \quad (5)$$

From (5), the reflected resistance and capacitance which are connected in series with the primary winding as shown in Fig. 3 (c) can be defined in (6) and (7), respectively.

$$R_r = \frac{\omega_{02}^2 M^2 (R_2 + R_L')}{R_2^2 + \omega_{02}^2 L_2^2} \quad (6)$$

$$X_{cr} = \frac{\omega_{02} M^2 (\omega_{02}^2 L_2 - R_L' R_2 / L_2)}{R_2^2 + \omega_{02}^2 L_2^2} \quad (7)$$

From Fig. 3 (c), the primary resonant frequency and the reflected capacitance can be calculated as,

$$\omega_{01} = \frac{1}{\sqrt{L_1 \left(\frac{c_1 c_r}{c_1 + c_r} \right)}} \quad (8)$$

and,

$$C_r = \frac{1}{\omega_{02} X_{cr}} \quad (9)$$

From (7) - (9), the load variation affects the value of the reflected capacitance. This causes the primary resonant frequency to change. The ZVS and non-ZVS operating regions are illustrated in Fig. 4. If the switching frequency (fs) is fixed as indicated by the solid line, the non-ZVS operation will occur when the load resistance is decreased. This is the case where the switching frequency becomes lower than the primary resonant frequency. In other words, the system will operate in non-ZVS mode if the load resistance is decreased from its designed value of 50 Ω. In addition, the non-ZVS operation is rather sensitive to the reduction of the load resistance or with the increment of the magnetic coupling coefficient (k).

2.2 Asymmetrical duty cycle (ADC) control

The ADC control is one of commonly used controls for a full-bridge inverter. Gate signals and output voltage waveforms obtained from the ADC control are shown in Fig. 5. The gate signals V_{G4} and V_{G2} are identical to V_{G1} and V_{G3} signals, respectively. Each pair of the switches operates in the complementary manner. Thus, the alpha angle (α) of the inverter output voltage (V_{inv}) is controlled by adjusting the duty cycle (D) of V_{G1}. The amplitude and phase of the fundamental component (v_1) of the inverter output voltage, can be obtained by

$$\left| \vec{V}_1 \right| = \frac{4V_{DC}}{\pi} \cos\left(\frac{\alpha}{2} \right) \quad (10)$$

and

$$\phi_{v1} = \frac{\alpha}{2} \quad (11)$$

where ϕ_{v1} is a phase difference between the inverter voltage and its fundamental component as shown in Fig. 5. Taking the rising edge of an inverter voltage as a reference, the voltage v_1 always leads v_{inv} or $\phi_{v1} \geq 0$. The primary circuit current is represented by i_1 as shown in Fig. 5 where ϕ_{i1} is a phase difference compared with the inverter voltage. From (2) and (10), the relationship between the output voltage (V_L) and the duty cycle (D) of the inverter voltage can be expressed as,

$$\left| \vec{V}_1 \right| = \frac{4V_{DC} M R_L' \cos[(0.5 - D) \times 180°]}{\pi L_2 (R_1 + R_r)} \quad (12)$$

Clearly, the magnitude of the output voltage is at its maximum at D = 0.5 and decreases with the duty cycle. By adjusting the duty cycle, the magnitude of the output voltage can be controlled. From Fig. 3 (c), the input impedance as seen from the voltage source is calculated as,

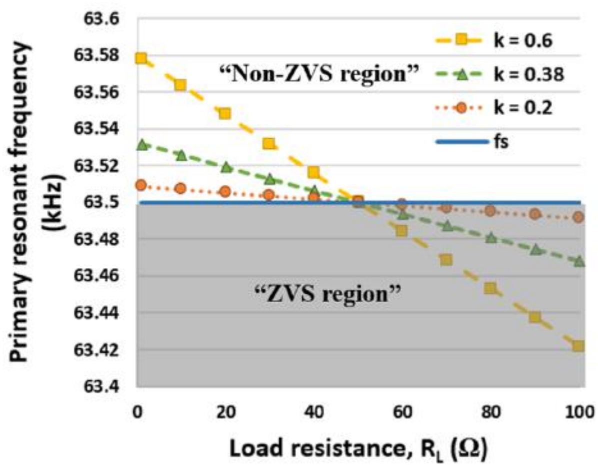


Fig.4: *Non-ZVS operation due to load variation.*

$$\vec{V}_{in} = (R_1 + R_r) + j(X_{L1} - X_{C1} - X_{cr}) \quad (13)$$

The phase angle is obtained by,

$$\theta_{\vec{Z}_{in}} = \tan^{-1}\left(\frac{\mathrm{Im}\{\vec{Z}_{in}\}}{\mathrm{Re}\{\vec{Z}_{in}\}} \right) \quad (14)$$

In a typical operation, most IPT systems are connected with high quality factor (Q) circuit. The inverter voltage and primary current can be approximated by the fundamental components, (v_1) and (i_1). Thus, ϕ_{i1} can be approximately estimated as,

$$\phi_{i1} = \phi_{v1} - \theta_{\vec{Z}_{in}} \quad (15)$$

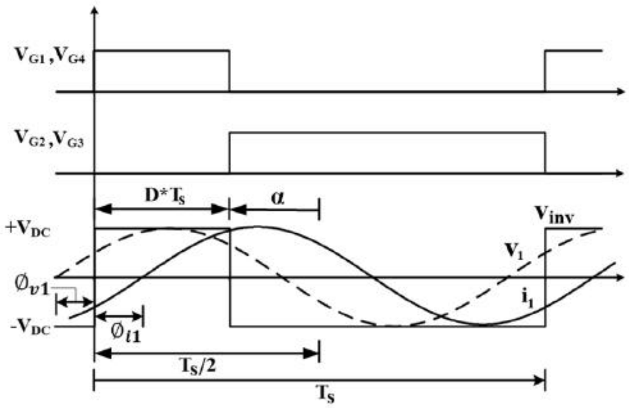


Fig.5: Gate signals and output voltage waveform of the full-bridge inverter with ADC control.

In Fig. 5, the condition for a ZVS operation is that the primary current lags the inverter voltage or,

$$\phi_{i1} \leq 0 \tag{16}$$

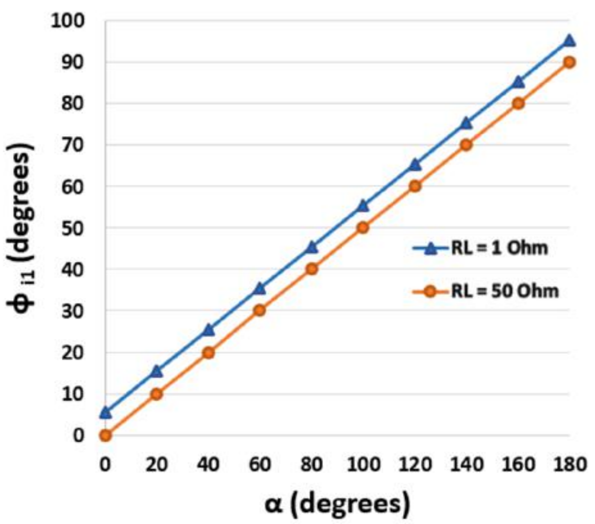


Fig.6: Non-ZVS operation due to α angle variation.

The plot of ϕ_{i1} against the α angle is shown in Fig. 6 where the targeted load resistance is at 50 Ω. As seen from the plot, primary current will lead the inverter voltage ($\phi_{i1} > 0$) when α angle is higher than 0° which cause the non-ZVS operation. The phase difference ϕ_{i1} is increased with the α angle. When the load resistance is decreased from its designed value, ϕ_{i1} becomes greater than 0° even at $\alpha = 0°$. For a given α angle and load resistance, the minimum switching frequency that achieve the ZVS condition ($\phi_{i1} = 0°$) can be calculated as,

$$f_{s,\min} = \frac{A + \sqrt{A^2 + 4\left\{C_1\left(L_1 - \frac{M^2}{L_2}\right)\right\}}}{4\pi\left[C_1\left(L_1 - \frac{M^2}{L_2}\right)\right]} \tag{17}$$

and

$$A = \frac{C_1 M^2 R_L \tan\left(\frac{\alpha}{2}\right)}{L_2^2} \tag{18}$$

As seen from the plot of $f_{s,\min}$ against α angle in Fig. 7, to maintain the ZVS operation, the switching frequency must be increased from the primary resonant frequency after the α angle is increased. The $f_{s,\min}$ is increased with the α angle. Moreover, high value of the load tends to require a higher switching frequency to achieve ZVS operation.

3. VERIFICATION

To verify the proposed control and analysis, a computer simulation is performed. Fig. 8 shows a circuit used in the simulation consisted of a DC voltage source, full-bridge inverter, coupled coils, compensation capacitors and a resistive load. The circuit para-

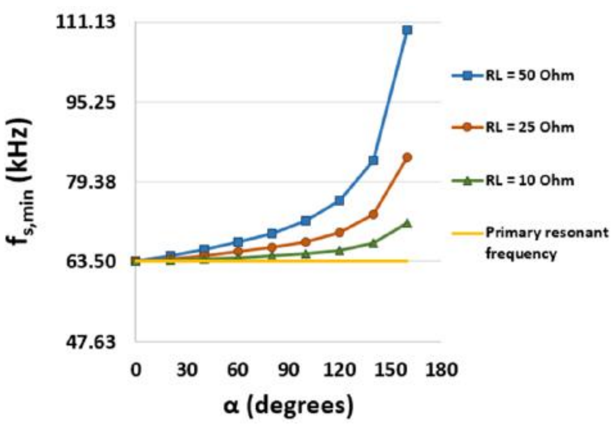


Fig.7: Minimum switching frequency to achieve a ZVS operation.

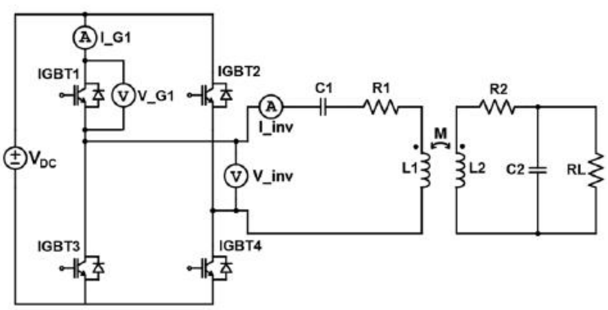


Fig.8: Simulation circuit.

meters used in the simulation are listed in Table 1. Simulation waveforms of the IGBT voltage (V_G1), IGBT current (I_G1), inverter voltage (V_inv) and inverter current (I_inv) at various load resistance and α angle are illustrated in Figs. 9 - 13. In Fig. 9, when the load resistance is at the designed value (50Ω) and the α angle is set to 0°, the ZVS operation is achieved. The primary current is essentially in-phase with the inverter voltage causing a phase difference ϕ_{i1} to be at 0°. The system is operated at the primary resonant frequency. When the load resistance is decreased to 1 Ω while α and f_s remain unchanged, the non-ZVS operation occurs as displayed in Fig. 10. The primary current slightly leads the inverter voltage and ϕ_{i1} becomes positive. This is an agreement with the results shown in Fig. 4. When the angle is increased from 0° to 90°, non-ZVS operation also occurs as illustrated in Fig. 11. This agrees with the results observed in Fig. 6. Fig. 12 depicts a non-ZVS operation when RL is decreased from 50Ω to 10Ω and α is increased from 0° to 135° while f_s is fixed at 63.5 kHz. By increasing the switching frequency to 66.68 kHz, as obtained from (17), the system can be restored to operate under ZVS condition as shown in Fig. 13. This verifies the previous theoretical analysis.

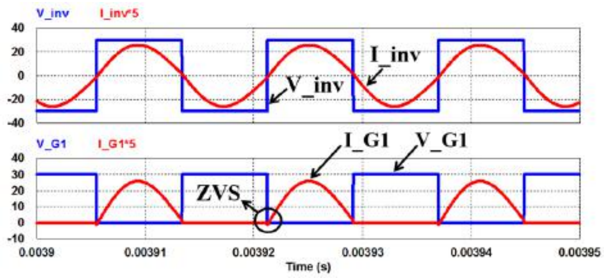


Fig.9: *Simulation results of a ZVS operation (R_L= 50 Ω, α = 0° and f_s = 63.5 kHz).*

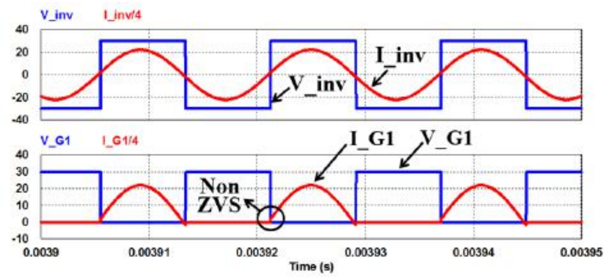


Fig.10: *Simulation results of a non-ZVS operation (R_L= 1 Ω, α = 0° and f_s = 63.5 kHz).*

4. PROPOSED CONTROLLER

From the previous section, the non-ZVS operation occurs when the load resistance is decreased from the designed value or when α angle is increased from 0°

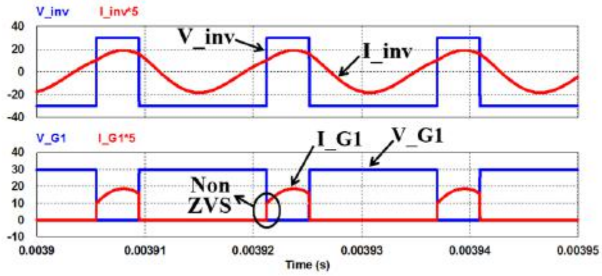


Fig.11: *Simulation results of a non-ZVS operation (R_L= 50 Ω, α = 90° and f_s = 63.5 kHz).*

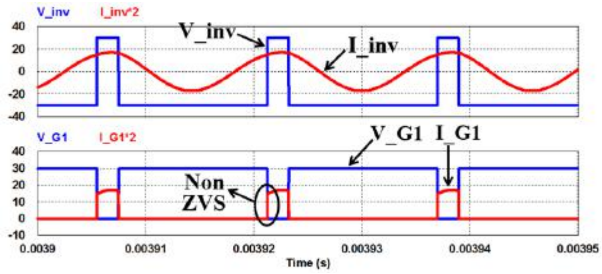


Fig.12: *Simulation results of a non-ZVS operation (R_L= 10 Ω, α = 135° and f_s = 63.5 kHz).*

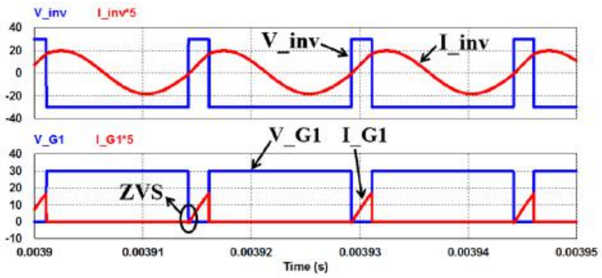


Fig.13: *Simulation results of a ZVS operation (R_L= 10 Ω, α = 135° and f_s = 66.68 kHz).*

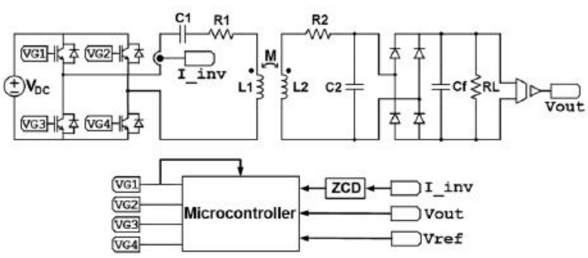


Fig.14: *Overall system of the proposed controller.*

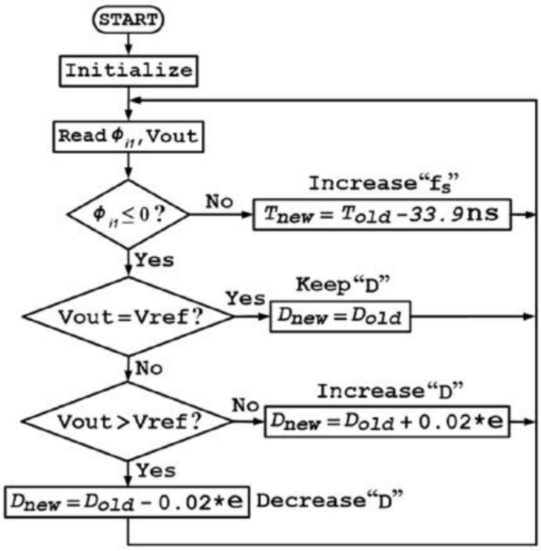


Fig.15: *Flowchart for the microcontroller.*

while the system is operating at a fixed frequency. However, by increasing the switching frequency, the inverter is operated under ZVS mode. Therefore, the controller proposed in this paper is based on a variable frequency control. Fig. 14 shows the overall system of the proposed controller. The primary circuit is supplied by a full-bridge inverter to produce the high frequency voltage. The induced voltage on the secondary side is then rectified using a full-wave rectifier. The aim of the proposed controller is to control the DC output voltage at the desired value under ZVS mode of operation. The variable frequency asymmetrical duty cycle (ADC) control of the full-bridge inverter is used in the proposed method. To regulate the DC output voltage, the α angle of the inverter voltage (V_{inv}) is controlled by adjusting the duty cycle of the gate signal V_{G1}. The switching frequency of the inverter circuit is adjusted for the ZVS operation. As shown in Fig. 14, signals being fed back to the controller are the DC output voltage (V_{out}), inverter current (I_{-inv}) and gate signal (V_{G1}). The phase of the inverter current is obtained by a zero current detection (ZCD) circuit. The gate signal VG1 represents the phase of the inverter voltage. The flowchart of the proposed controller is shown in Fig. 15. The phase difference between the inverter voltage and current (ϕ_{i1}) is obtained by using an input capture function of the microcontroller. The analog to digital function is used to obtain the value of the DC output voltage (V_{out}). First, a ZVS condition ($\phi_{i1} \leq 0$) is checked. If the system is operating under the non-ZVS mode, the controller will increase the switching frequency until the ZVS mode is achieved. Then, output voltage condition is checked. If DC output voltage is lower than the reference voltage (V_{ref}), the controller increases the duty cycle of V_{G1} to increase the output voltage. On the other hand, if the DC output voltage is higher than the reference value, the controller will decrease the duty cycle of V_{G1} to reduce the DC output voltage to the desired value. The process of the duty cycle adjustment stops when the output voltage is equal to the reference voltage.

The steps for duty cycle and switching frequency adjustments, as shown in Fig. 15, are empirically obtained where the optimum response is taken into consideration. The switching period of the inverter is denoted by "T" while "e" is an error from the comparison between the output and reference voltages. Since the primary quality factor (Q_1) is relatively high, its bandwidth is narrow. The change in frequency is very sensitive to the phase difference between the inverter voltage and current (ϕ_{i1}). The minimum step of the switching period variation is chosen at 33.9 ns, which is the smallest achievable step of the chosen microcontroller. Larger values of the frequency step may cause an increase in ϕ_{i1}, resulted in sacrificed system efficiency and output power. To improve the system response, the changing step for the duty cycle is dependent on the error (e). A large changing step causes an overshoot in the response. If the changing step is too small, it may take longer for the response to reach the steady state.

5. EXPERIMENTAL RESULTS

To validate the proposed controller, an experimental setup, as shown in Fig. 16, has been implemented. Experiments of output voltage control due to step change in reference voltage and air gap have been performed.

5.1 Design of a coupled coil

The design objective is to send an output power (P_{out}) of 200 Watt to a 20 Ohm load resistance (R_L) with

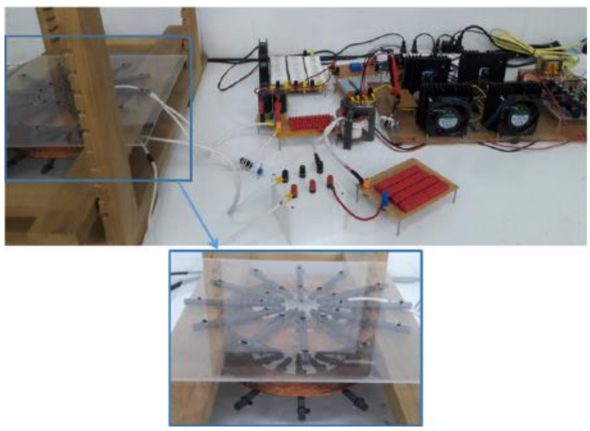


Fig.16: *Experimental setup.*

90% efficiency (η) from a 30 Volt DC input voltage (VDC) with the switching frequency (fs) at 63.5 kHz. The design constraints are:

Table 1: *Measured circuit parameters at 6 cm air-gap.*

Parameters	Value
L_1	104.1 μH
L_2	105.8 μH
M	40.07 μH
k	0.38
R_1	0.25Ω
R_2	0.25Ω

1) Primary coil is identical to secondary coil ($L_1 = L_2$ and $R_1 = R_2$).
2) The system is operating at secondary resonant frequency ($f_s = f_{02}$).
3) Quality factor of the primary circuit is high ($Q_1 > 10$).

From the design objective and constraints, coupled coils are designed as follows,

At first, the required magnitude of primary current is calculated from,

$$\left|\vec{I}_1\right| = \frac{\pi P_{out}}{2\eta V_{DC}} \tag{19}$$

$$= \frac{\pi(200)}{2(0.9)(30)} = 11.64 \text{ A}$$

From (19), the diameter of a conduction wire is selected for the designed current. The required magnetic coupling can be obtained by using the calculated $\left|\vec{I}_1\right|$ as,

$$k = \sqrt{\frac{2P_{out}}{\left|\vec{I}_1\right|^2 R_L}} \tag{20}$$

$$= \sqrt{\frac{2(200)}{(11.64)^2(20)}} = 0.384$$

Next, the minimum value of the primary coil's inductance for high primary quality factor can be calculated as,

$$L_{1,\min} = \frac{20V_{DC}}{\pi^2 f_{02}\left|\vec{I}_1\right|} \tag{21}$$

$$= \frac{20(30)}{\pi^2(63.5k)(11.64)} = 82.25\ \mu\text{H}$$

The exact value of the coil's inductance is dependent on the air gap. The more air gap distance, the greater the inductance needed to achieve the required magnetic coupling. The maximum value of coil's resistance with the primary quality factor higher than 10 can be obtained by using the calculated $L_{1,\min}$ as,

$$R_{1,\max} = \frac{2(P_{out}/\eta - P_{out})}{\left[\left|\vec{I}_1\right|^2 + k^2\left|\vec{I}_1\right|^2(1 - R_L/2\pi f_{02}L_{1,\min})\right]} \tag{22}$$

From (22), the calculated value of $R_{1,\max}$ is 0.31 Ω. From all the calculated design values, the coupled coil can be implemented. Due to the advantage of scalable function, the coupled coil with circular magnetic structure has been used in this paper. The ferrite arrangement has been selected from the highest value of ferrite utilization as introduced in [16]. The litz-wire made from 50 of AWG 31 wires is used as the conduction wire. The primary and secondary coils are made identical and consist of 16 turns of litz-wire. Each coil has 12 pieces of ferrite bars attached, as shown in Fig. 16. The measured circuit parameters at 6 cm air gap are listed in Table 1. The load resistance used in the experiment is 20 Ω. The secondary resonant frequency is 63.5 kHz. The DC input voltage is maintained at 30 V.

5.2 Output voltage control

Experimental results of output voltage control under a step change of the reference voltage are shown in Figs. 17 - 20. For comparison purpose, the results from both fixed and variable frequency ADC control are shown in both ZVS and non-ZVS operations. Experimental results of the DC output voltage control with a step change in the reference voltage from 63 to 40 V are shown in Figs. 17 and 18, for fixed and variable frequency ADC control, respectively. At the beginning of operation, the DC output voltage is set to 63 V where $\alpha = 0°$ and $f_s = 63.5$ kHz. The ZVS operation is achieved since the primary current is inphase with the inverter voltage as shown in Figs. 17 (b) and 18 (b). Then, the reference voltage is changed to 40 V. The DC output voltage is decreased to 40 V for both fixed and variable frequency ADC control by automatically decreasing the duty cycle of an inverter voltage. In Fig. 17 (c), the non-ZVS operation occurs as seen from the leading phase of an inverter current compared with the inverter voltage. In fact, the α angle is greater than 0° and the switching frequency is fixed. However, by the use of the proposed variable frequency ADC control, the ZVS operation can be obtained by automatically increasing the switching frequency. This results in a lagging phase of the primary current as shown in Fig. 18 (c). Figs. 19 and 20 show the step response of DC output voltage with a step increase in the reference voltage from 30 to 50 V. At the beginning of operation where DC output voltage is set to 30 V, the non-ZVS operation occurs as shown in Figs. 19 (b) and 20 (b). This is due to $\alpha > 0°$ and fs is fixed at 63.5 kHz. Then, V_{ref} is instantaneously increased to 50 V. The DC output voltage is increased

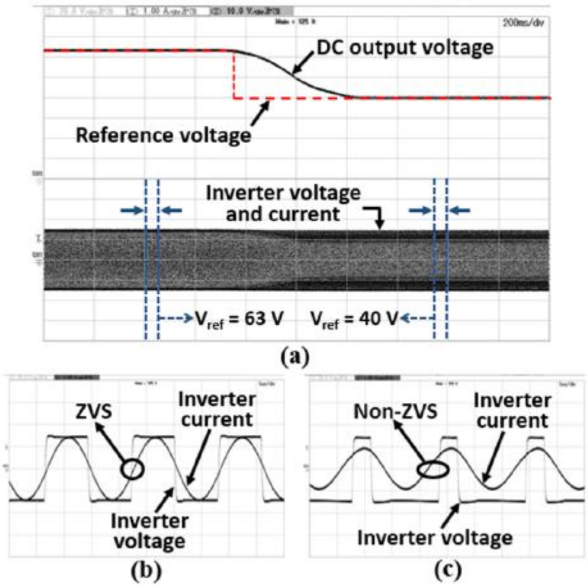


Fig.17: *(a) Experimental results of non-ZVS operation due to output voltage control with fixed frequency (Step decrease in output voltage reference from 63 to 40 V).*
(b)Magnified version of inverter voltage and current under V_{ref} = 63 V.
(c)Magnified version of inverter voltage and current under V_{ref} = 40 V.

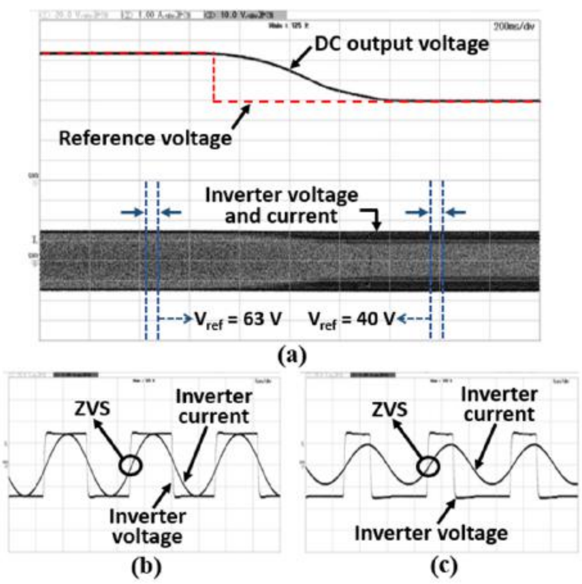


Fig.18: *(a) Experimental results of ZVS operation due to output voltage control with variable frequency (Step decrease in output voltage reference from 63 to 40 V).*
(b)Magnified version of inverter voltage and current under V_{ref} = 63 V.
(c)Magnified version of inverter voltage and current under V_{ref} = 40 V.

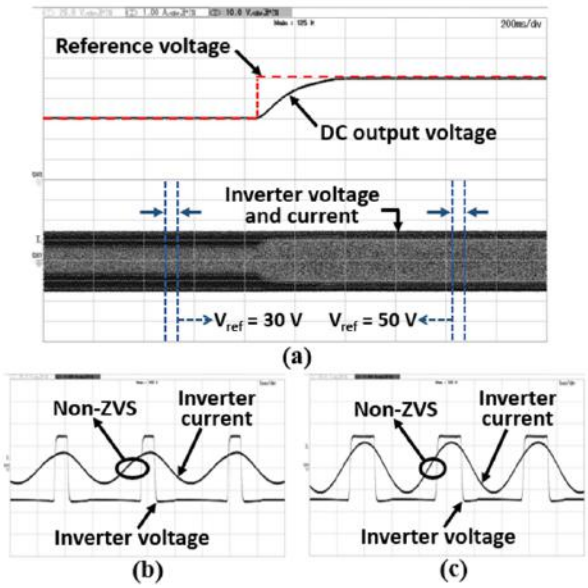


Fig.19: *(a)Experimental results of non-ZVS operation due to output voltage control with fixed frequency (Step increase in output voltage reference from 30 to 50 V).*
(b)Magnified version of inverter voltage and current under V_{ref} = 30 V.
(c)Magnified version of inverter voltage and current under V_{ref} = 50 V.

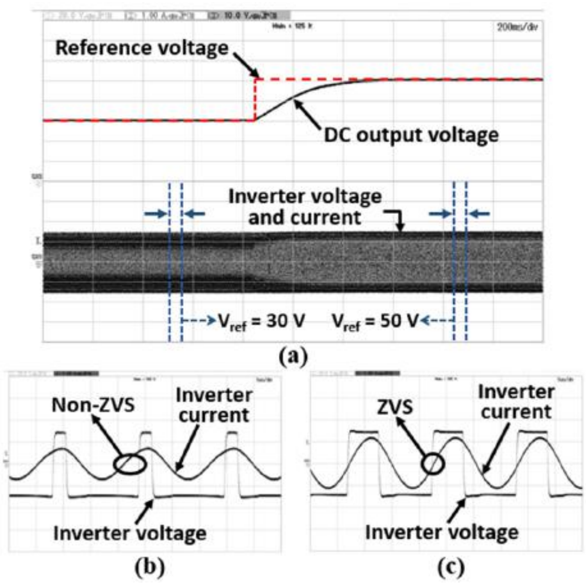


Fig.20: *(a)Experimental results of ZVS operation due to output voltage control with variable frequency (Step increase in output voltage reference from 30 to 50 V).*
(b)Magnified version of inverter voltage and current under V_{ref} = 30 V.
(c)Magnified version of inverter voltage and current under V_{ref} = 50 V.

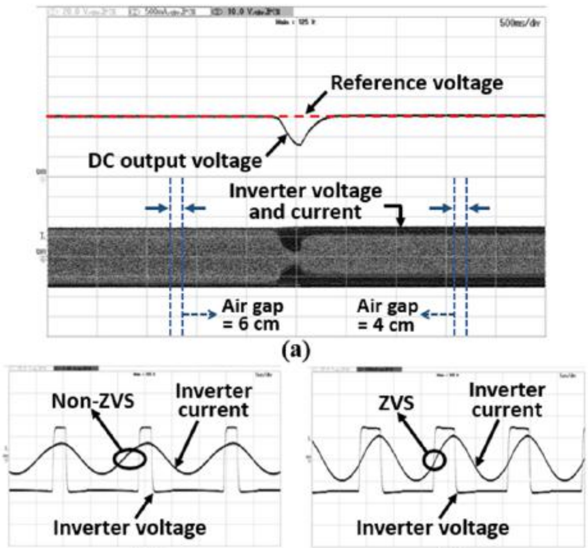


Fig.21: *(a)Experimental results of ZVS operation due to output voltage regulation with variable frequency (Step decrease in air gap from 6 to 4 cm). (b)Magnified version of inverter voltage and current under Air gap = 6 cm. (c)Magnified version of inverter voltage and current under Air gap = 4 cm.*

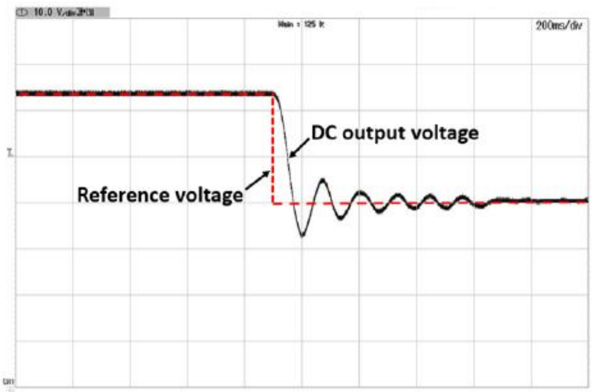


Fig.22: *Experimental results of output voltage response when changing step of duty cycle in the microcontroller is increased.*

to 50 V by varying the duty cycle of the inverter voltage. The non-ZVS operation is observed in the system with fixed frequency ADC control as shown in Fig. 19 (c). With the proposed variable frequency ADC control, a ZVS operation can be obtained by automatically increasing the switching frequency as seen in Fig. 20 (c).

The change in air gap causes the magnetic coupling (k) to change which results in the output voltage variation. However, with the proposed controller, the output voltage can be regulated. For example, if the air gap is decreased then, the coupling coefficient k is increased. This causes the output voltage to be reduced. From the flowchart in Fig. 15, the duty cycle must be increased to adjust the output voltage while keeping the system operation under ZVS mode through switching frequency variation. The experimental results of DC output voltage regulation for a case of air gap distance change is shown in Fig. 21. The reference voltage is set to 30 V. As the coefficient k is increased from 0.38 to 0.52 by decreasing the air gap from 6 cm to 4 cm. The output voltage is decreased at the beginning. Then, it is restored to 30 V with ZVS operation by automatically increasing the duty cycle and the switching frequency. Thus, from all the results as shown in Figs. 17 - 21, the proposed controller can be validated.

Fig. 22 shows the output voltage response when the duty cycle step is increased from 0.02 to 0.2. The output voltage response has higher overshoot and longer settling time compared with the results in Fig. 18. This is due to the inappropriate duty cycle changing step. As stated earlier, the empirical method has been used to select the changing step for simplicity. An optimal controller design and system stability analysis will be carried out as a future work.

The efficiency comparison of the ADC control with fixed and variable frequency for the same angle and output power is shown in Fig. 23. The efficiency of a variable frequency control is higher than the fixed frequency control for all α angle due to a lower turn-on switching loss. With $\alpha = 0°$, both controls have the same efficiency at 74.3%. Once the output power is decreased, indicated by an increase in the α angle, the difference becomes obvious. At $\alpha = 135°$, the efficiency of the proposed variable frequency control is 13.7% higher.

6. CONCLUSION

The output voltage control of the SP topology IPT system using a primary side controller is proposed in this paper. The absence of the DC-DC converter and secondary side controller reduces the cost, size and complexity of the system. The ADC control method is implemented on a full-bridge inverter to control the DC output voltage through the inverter voltage. In addition to the ZVS operation, the switching frequency of the inverter circuit is also varied. Theoretical analysis of the non-ZVS mode due to load and α angle variations has been performed based on mutual inductance coupling

model. Simulation and experimental studies are performed. Experimental results of the DC output voltage control due to reference voltage and magnetic coupling variations validate the proposed controller. The system efficiency is improved and become obvious as the output power is reduced. The optimal controller design and system stability analysis will be carried out in our future work.

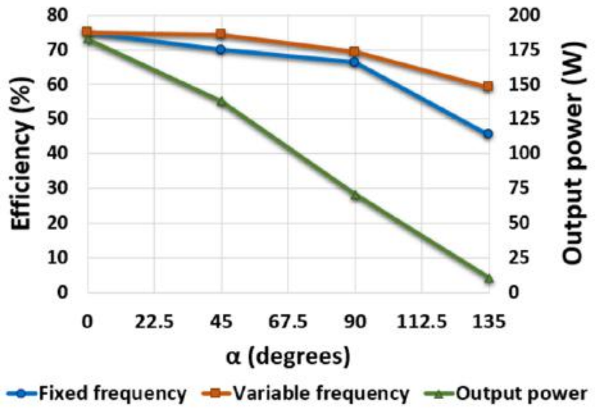


Fig.23: *Experimental results of the efficiency comparison between fixed and variable frequency control.*

ACKNOWLEDGEMENT

This work was supported by the Petchra Pra Jom Klao Doctoral Scholarship, King Mongkut's University of Technology Thonburi (KMUTT), Thailand.

References

[1] G. A. Covic and J. T. Boys, "Inductive power transfer," *Proc. IEEE*, vol. 101, no. 6, pp. 1276-1289, 2013.

[2] I. Mayordomo, T. Drager, P. Spies, J. Bernhard, and A. Pflaum, "An overview of technical challenges and advances of inductive wireless power transmission," *Proc. IEEE*, vol. 101, no. 6, pp. 1302-1311, 2013.

[3] F. Musavi and W. Eberle, "Overview of wireless power transfer technologies for electric vehicle battery charging," *IET Power Electron.*, vol. 7, no. 1, pp. 60-66, 2014.

[4] C. S. Wang, O. H. Stielau, and G. A. Covic, "Load models and their application in the design of loosely coupled inductive power transfer systems," *in Proc. Int. Conf. Power System Technology*, vol. 2, pp. 1053-1058, 2000.

[5] C. S. Wang, G. A. Covic, and O. H. Stielau, "Power transfer capability and bifurcation phenomena of loosely coupled inductive power transfer systems," *IEEE Trans. Ind. Electron.*, vol. 51, no. 1, pp. 148-157, 2004.

[6] S. Nutwong, A. Sangswang, and S. Naetiladdanon, "Design of the wireless power transfer system with uncompensated secondary to increase power transfer capability," *in Proc. 8th IET Int. Conf. Power Electron. Mach. Drives (PEMD)*, pp. 1-5, 2016.

[7] C. Duan, C. Jiang, A. Taylor, and K. Bai, "Design of a zero-voltage-switching large-air-gap wireless charger with low electric stress for electric vehicles," *IET Power Electron.*, vol. 6, no. 9, pp. 1742-1750, 2013.

[8] T. Diekhans and R. W. De Doncker, "A dual-side controlled inductive power transfer system optimized for large coupling factor variations and partial load," *IEEE Trans. Power Electron.*, vol. 30, no. 11, pp. 6320-6328, 2015.

[9] H. H. Wu, A. Gilchrist, K. D. Sealy, and D. Bronson, "A high efficiency 5 kW inductive charger for EVs using dual side control," *IEEE Trans. Ind. Informat.*, vol. 8, no. 3, pp. 585-595, 2012.

[10] P. Si, A. P. Hu, S. Malpas, and D. Budgett, "A frequency control method for regulating wireless power to implantable devices," *IEEE Trans. Biomed. Circuits Syst.*, vol. 2, no.1, pp. 22-29, 2008.

[11] H. L. Li, A. P. Hu, and G. A. Covic, "A power flow control method on primary side for a CPT system," *in Proc. IEEE, The International power electronics conference (IPEC)*, pp. 1050-1055, 2010.

[12] K. Aditya and S. S. Williamson, "Advanced controller design for a series-series compensated inductive power transfer charging infrastructure using asymmetrical clamped mode control," *in Proc. IEEE, Applied Power Electronics Conference and Exposition (APEC)*, pp. 2718-2724, 2015.

[13] B. C. Kim, K. Y. Kim, S. Ramachandra, A. Khandelwal, and B. H. Lee, "Wireless lithium-ion battery charging platform with adaptive multi-phase rapid-charging strategy," in Proc. IEEE, Energy Conversion Congress and Exposition (ECCE), pp. 3087-3091, 2015.

[14] S. Nutwong, A. Sangswang, and S. Naetiladdanon, "Output voltage control of the SP topology IPT system using a primary side controller," *in Proc. 13th Int. Conf. on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, pp. 1-5, 2016.

[15] C. S. Wang, O. H. Stielau, and G. A. Covic, "Design considerations for a contactless electric vehicle battery charger," *IEEE Trans. Ind. Electron.*, vol. 52, no. 5, pp. 1308-1314, 2005.

[16] M. Budhia, G. A. Covic, and J. T. Boys, "Design and optimization of circular magnetic structures for lumped inductive power transfer systems," *IEEE Trans. Power Electron.*, vol. 26, no. 11, pp. 3096-3108, 2011.

Supapong Nutwong was born in Bangkok, Thailand. He received the B.E. and M.E. degree in Electrical Engineering from King Mongkut's University of Technology Thonburi (KMUTT), Bangkok, Thailand, in 2007 and 2011 respectively. He is presently working toward his D.E. degree in Electrical and Information Engineering Technology, KMUTT. His current research interests include power electronics and control for wireless power transfer (WPT) and induction heating (IH) applications.

Sumate Naetiladdanon received his B.Eng. degree from Chulalongkorn University, Bangkok, Thailand, in 1995, his M.S. degree from Rensselaer Polytechnic Institute, Troy, NY, USA., in 1999, and his Ph.D. degree from Osaka University, Osaka, Japan, in 2006. He is presently a Lecturer in the Department of Electrical Engineering, King Mongkut's University of Technology Thonburi (KMUTT), Bangkok, Thailand. His current research interests include power electronics for power quality improvement and renewable energy.

Anawach Sangswang was born in Bangkok, Thailand. He received the B.Eng. degree from the King Mongkut's University of Technology Thonburi (KMUTT), Bangkok, in 1995 and the M.S. and Ph.D. degrees from Drexel University, Philadelphia, Pennsylvania, USA, in 1999 and 2003, respectively. He is currently an Assistant Professor with the Department of Electrical Engineering, KMUTT. His research interests include stochastic modeling, digital control of power electronic converters, and power system stability.

Ekkachai Mujjalinvimut was born in Bangkok, Thailand, on September 17, 1985. He received the B.E., M.E. and D.E. degrees in the Department of Electrical Engineering from King Mongkut's University of Technology Thonburi (KMUTT), Bangkok, Thailand, in 2007, 2009 and 2016, respectively. He is currently a lecturer with the Department of Electrical Engineering, KMUTT. His current research interests include switched mode power supplies, applications of nonlinear control theory, and digital control.

Enhanced Running Spectrum Analysis for Robust Speech Recognition Under Adverse Conditions: Case Study on Japanese Speech

George Mufungulwa[1], **Hiroshi Tsutsui**[2], **Yoshikazu Miyanaga**[3], and **Shin-ichi Abe**[4], Non-members

ABSTRACT

In real environment, many noises degrade the performance of Automatic Speech Recognition (ASR) systems. In addition, in case of similar pronunciations, it is not easy to realize high accuracy of recognition rate. From this point of view, our work envisaged an enhanced processing algorithm into speech modulation spectrum as Running Spectrum Analysis (RSA). It is also adequately applied to observed speech data. In the envisaged method, a modulation spectrum filtering (MSF) method directly modifies the observed cepstral modulation spectrum by Fourier transform of the cepstral time frequency. The method and experiments carried out for various passbands had favorable results that showed the improvement of about 1-4 % recognition accuracy as compared with current conventional methods.

Keywords: MFCC, HMM, ASR, RSF, RSA

1. INTRODUCTION

The fundamental stages in speech recognition are speech feature extraction and feature matching. Various speech features, including ones from linear prediction coding (LPC) [1-4], time-varying linear prediction coding (TVLPC) [5], mel frequency cepstral coefficients (MFCC) [6-9] among others, have been used to model speech recognition either singularly or collectively in improving speech recognition accuracies. MFCC, which is based on spectral content of the signal and can be considered as one of the standard method for feature extraction [10] is opted for use in our study.

Speech recognition systems often suffer from multiple sources of variability due to corrupted speech signal features [11]. In compensating for distortions, most speech recognizers use normalization methods and noise filtering techniques in conjunction with voice activity detection (VAD) techniques. Improved accuracy in noise robust speech recognition can be realized by processing speech using running spectrum filtering (RSF)[12, 13], for example. The downside, is high computation costs and high demand on memory.

In recent past, several typical methods relating to the use of modulation spectrum features for noisy speech recognition have been developed [14–16]. Running spectrum analysis (RSA) is not only an effective technique for reduction of noise on the modulation spectrum domain (MSD)[17] but it can also be deployed to realize ideal processing [18].

Although running spectrum analysis (RSA) is a well known method focusing on modulation spectrum, it has mostly been applied for automatic continuous speech recognition [19]. Furthermore, in speech communication, its application has been mainly focused on frequency components in the range of 2-8 Hz because this range contains the dominant components of the amplitude envelope of speech [20][21]. Modulation frequency band higher than 8 Hz can be regarded as miscellaneous noise components or such unnecessary speech components for recognition as speaker's characteristics such as tone, pronunciation, etc [22].

However, this work presented a novel noise-robust feature extraction framework that leveraged the technique of RSA on isolated phrase recognition. This work was envisaged with the goal to enhance RSA for the purpose of achieving higher recognition accuracy for both male and female, similar and non-similar pronunciation Japanese speech phrases under noisy conditions. Robust speech features realized using this method can be required in many applications, including modelling for analysis/synthesis and recognition of isolated utterances with "Listen/Not-Listen" states. Situations in which this method can be applied include tasks that require human machine interface such as automatic call processing in telephone networks and query based information systems such as voice dictation, stock price quotations, [23] among others. Authors assume that the proposed method performance relates with gender just as recognition accuracy can be influenced by the signal-to-noise ra-

Manuscript received on April 3, 2017 ; revised on May 12, 2017.

Final manuscript received on June 6, 2017.

[1,2,3] The authors are with Graduate School of Information Science and Technology, Hokkaido University, Kita 14, Nishi 9, Kita-ku, Sapporo 060-0814, Japan., E-mail: mufungulwac@gmail.com, hiroshi.tsutsui@ist.hokudai.ac.jp and miya@ist.hokudai.ac.jp

[4] The author is with Vehicle Information and Communication System Center (VICS Center), Nittochi Kyobashi Bldg., 8F, 2-5-7 Kyobashi, Chuo-ku, Tokyo, 104-0031, Japan., E-mail: s-abe@vics.or.jp

tio (SNR) which the authors aim to ascertain.

In this study, the work applied running spectrum analysis (RSA) on modulation spectrum for noise robust speech recognition of adequately selected frequency components. The noise effect was dealt with filtering the range of frequency components, 1-7 Hz, 1-15 Hz, 1-35 Hz and 1-40 Hz in the modulation spectrum domain. Further, it is argued that the expected speech recognition accuracy can be improved when modulation spectrum filtering (MSF) directly modify the cepstral modulation spectrum (CMS) [16] which is specifically referred to as the Fourier transform of the cepstral time sequence.

Although hidden Markov modelling (HMM) based approaches require training in automatic speech recognition (ASR) systems, the HMM method has been widely used. Since there are several noise reduction methods and speech enhancement methods against any noises, almost all of ASR systems using HMM and noise reduction can show higher accuracy of speech recognition rate than that given by a conventional standard HMM based ASR.

The rest of the paper is organized as follows. In Section 2, the proposed system is explained. In Section 3, performance of proposed method is evaluated. In the same section, experimental conditions are explained and the results stated. Section 4 discusses the results and in Section 5 which is the conclusion compares the enhanced RSA over the RSF.

2. PROPOSED SYSTEM

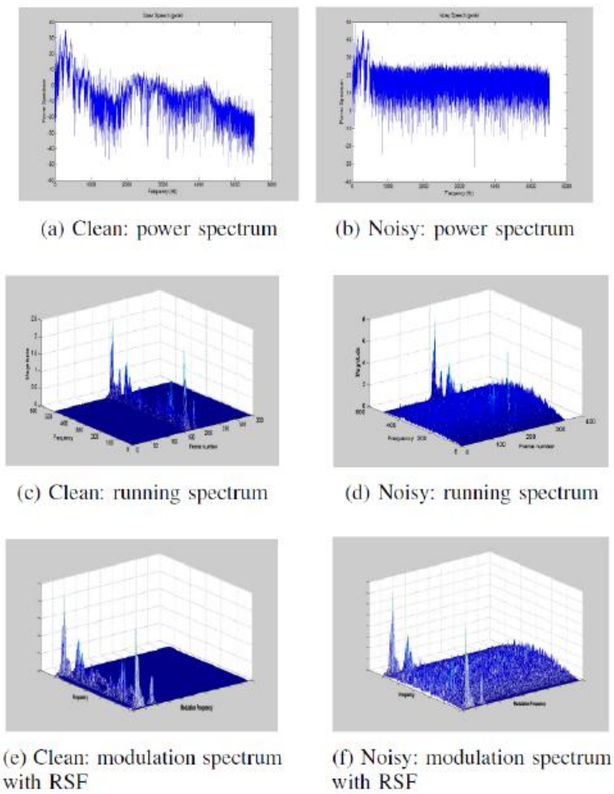
The motivation of this study is to evaluate the effectiveness of the enhanced running spectrum analysis (RSA), which is explained later, as it compares with running spectrum filtering (RSF). RSA is the processing of speech over modulation spectrum domain. Linguistically dominant factors of the speech signal may occupy different parts of the modulation spectrum than do some non-linguistics factors such as steady additive noise [24]. A proper processing of modulation spectrum of speech may improve quality of noisy speech. Investigations on possibilities of the modulation spectrum domain for enhancement of noisy speech [25][26] support the dominance of modulation spectrum components in the vicinity of 2-8 Hz in speech communication.

We now explain the effect of noise in running and modulation spectrum domains.

For standard speech information processing, the frame concept has been applied. The 256 sample point length frame is first defined and using this frame, a short time speech waveform is extracted. For the short time speech waveform, a speech power spectrum is calculated as a typical speech analysis. The frame is shifted with 128 points and then many short time speech waveforms can be obtained. Running spectrum is defined as the time trajectory in frequency domain. It consists of many speech power spectra given from short time frames. The modulation spectrum is defined as the spectrum in time varying of short-time running spectrum.

Figures 1(a) and 1(b) show the power spectra of clean speech and speech with additive white noise at 10 dB SNR for a Japanese phrase /genki/. Both spectra are calculated from short time speech waveforms. These figures indicate that the dynamic range on a power spectrum of a noisy speech is smaller than that of a clean spectrum. In addition, some of the power spectrum characteristics are unobservable under noisy conditions. Figure 1(c) shows the running spectrum of clean speech while Figure 1(d) shows the running spectrum of noisy speech of the same phrase /genki/. There are three axes, i.e., frequency axis, frame number axis and power amplitude axis.

When we observe the data on the frame number axis, the frequency is fixed to a specific value, its data can be recognized in the time domain. They can be applied by using fast Fourier transform (FFT). After such FFT is applied to all frequencies, we can get new 3-d data in the modulation spectrum domain. Modulation spectrum of the noisy signal is shown in Figure 1(f) and the modulation spectrum of the clean speech is shown in Figure 1(e).



(a) Clean: power spectrum
(b) Noisy: power spectrum
(c) Clean: running spectrum
(d) Noisy: running spectrum
(e) Clean: modulation spectrum with RSF
(f) Noisy: modulation spectrum with RSF

Fig.1: *Power spectra of (a) clean speech, and (b) noisy speech phrase /genki/ with white noise at 10 dB SNR. Running spectrum of (c) clean speech and (d) noisy speech. Modulation spectrum of (e) clean speech and (f) noisy speech with RSF*

Figure 2 shows the proposed system for which results and analysis are presented in Section 3. The left side of the figure shows the processes for male speakers while the right side of the same figure shows processes for female speakers. For each gender case, two output models for similar pronunciation (SP) and non similar pronunciation (NSP) respectively are realized. In the proposed system, there are four different kinds of filtering in RSA. The optimal filtering of RSA is applied for male and female speakers, SP and NSP.

In Figure 2, noisy speech at different signal-to-noise ratio (SNR) is input into a short-term energy (STE) based VAD for the purpose of retaining speech segments with sufficient energy while eliminating segments classified as noisy as well as silent. As in the case of training, the speech features are extracted using the standard MFCC as spectral analysis. A HMM based automatic speech recognition (ASR) system is utilized for testing. The gender of speaker (male or female) as well as the speech type, SP or NSP for each gender case are decided. This process results in four outputs; male SP, male NSP, female SP and female NSP, respectively. For each gender and speech type combination, the speech signal is passed through a voice activity detection (VAD) process in order to retain segments with speech activity or segments with high energy while eliminating segments with background noise or the ones with less energy prior to feature extraction.

Figure 3 shows the feature extraction process using fast Fourier transform (FFT) based MFCC with running spectrum filtering (RSF) for log spectra as a noise reduction technique.

In figure 3, it is shown that in order to obtain mel cepstrum, speech data is initially pre-emphasized and the pre-emphasized speech waveform in time domain is frame-blocked and windowed with a pre-defined analysis window. Later, fast Fourier transform (FFT) is computed. The magnitude of the output is then weighted by a series of mel filter frequency responses whose center frequencies and bandwidth roughly match those of auditory critical band filters [27]. The FFT bins are later combined so that each filter has unit weight. From the weighted sums of all amplitudes of signals, a vector is obtained by logarithmic amplitude compression computation. RSF is then applied before transforming the result to MFCC parameter by discrete cosine transform (DCT).

The performance of most if not all speech/audio processing methods is crucially dependent on the robustness of the extracted speech features. The accuracy of automatic speech recognition remains one of the important research challenges [23]. Most current feature extraction methods are still vulnerable against certain noises such as car noise [28].

Figure 4 shows the MFCC feature extraction process with running spectrum analysis (RSA). After spectral analysis, RSA is applied to realize the modulation spectrum. After which stage the process is as explained under feature extraction with RSF. In both cases, the features are trained into HMM, respectively.

In this paper, different types of enhanced RSA were selected for male and female speakers under noisy conditions.

During our preliminary study, among the RSA type (c) and type (f) were found to be better performers for male NSP and for SP respectively. Our study have also shown that, for example, in the case of female NSP, RSA with type (h) is better performer at high noise while type (c) and type (d) perform better at low noise. Similarly, for female SP, RSA with type (c) and type (h) were found to be better performers at high noise while type (d) performed better at less noise, respectively. The candidate of results with male or female speech are selected based on the maximum likelihood of HMM. Under noisy conditions different types of RSA show different performance for male and female speakers.

The proposed RSA differs from the one discussed in [19], for example. The former focuses on modulation frequency range of 2-8 Hz. However, in this study we evaluate the performance of several RSA types shown in Table 1. Table 1 shows 8 RSA passband specifications whose different sets of values are given as examples of filtering. In the modulation spectrum, it is possible to see the frequency range of the power concentration for each phrase and thereby help to decide which RSA type is most suitable. Each passband has a low cut-off frequency (LCF), and a high cut-off frequency (HCF). The difference between the two frequencies represents the number of frequency components over the modulation spectrum domain that are to be processed. In this way, we aim to determine the performance of new RSA over that of RSF by changing parameters such as; i) the number of frequency components (7, 15, 30, or 40 components), ii) the type of speaker (male or female), and iii) the signal-to-noise ratio (SNR) (10 dB, 15 dB, or 20 dB).

Table 1: *RSA passband specifications*

RSA Type	LCF (Hz)	HCF (Hz)
(a)	1	7
(b)	1	15
(c)	1	35
(d)	1	40
(e)	0.5	7
(f)	0.5	35
(g)	0.1	7
(h)	0.1	35

3. EXPERIMENTAL RESULTS

3.1 Objectives of the Experiments

The first objective of the experiments is to compare the performance of the proposed enhanced RSA

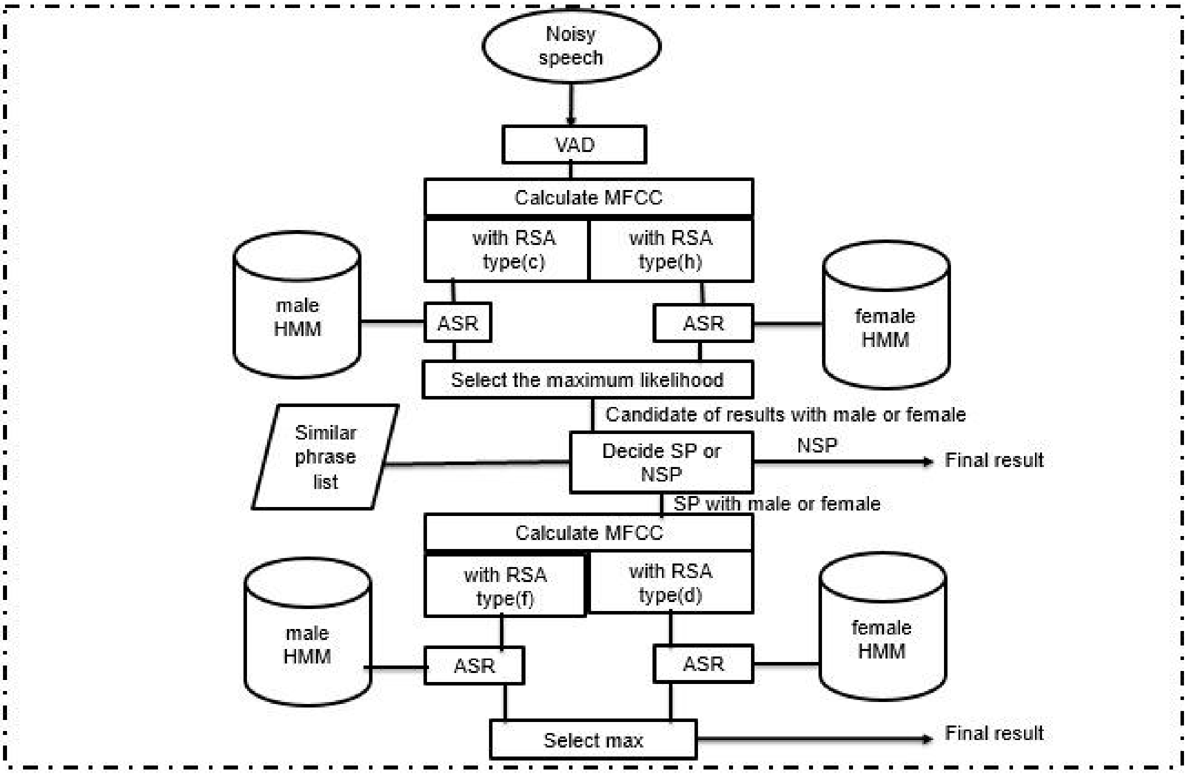


Fig.2: *Proposed system.*

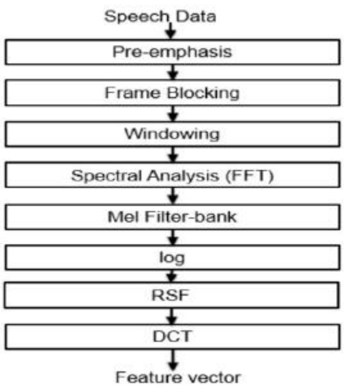


Fig.3: *Feature extraction with RSF.*

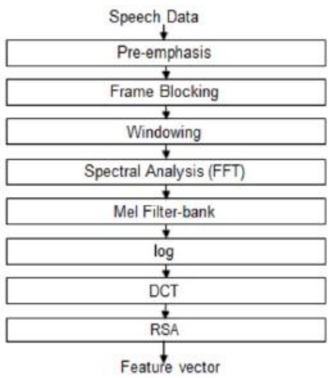


Fig.4: *Feature extraction with RSA.*

to that of RSF on similar and non-similar Japanese pronunciation phrases. The second objective is to evaluate how the performance relates to gender. The main method used for speech enhancement is filtering. We have evaluated the adaptability of our proposed RSA over modulation spectrum and compared its results to those of RSF. In this study, RSF is employed to act as the basis for comparing the tendency and to determine better performing RSA types at the given SNR for both gender.

3.2 Simulation parameters and conditions of experiments

Table 2 shows the simulation parameters.

Training sets of 30 male speakers and 30 female speakers, each speaker uttering 6 similar phrases and 100 Japanese common phrases, respectively, and each phrase repeated 3 times, are used for the front-end feature extraction and 32-states isolated phrase hidden Markov modeling (HMM) in training. Testing sets consisting of 10 male speakers and 10 female speakers (not used in training), with each speaker uttering 6 similar phrases and 100 Japanese common phrases and each phrase repeated 3 times respectively are utilized.

The speech sample is 11.025 KHz and 16-bit quantization. Frame-by-frame, 38-dimensional FFT based MFCC feature vectors are extracted after pre-emphasis and Hanning windowing. In the testing

Table 2: *The condition of speech recognition experiments*

Parameter name	Parameter value/type
Sampling	11.025 kHz (16-bit)
Frame length	23.2 ms (256 samples)
Shift length	11.6 ms (128 samples)
Pre emphasis	$1-0.97z^{-1}$
Windowing	Hanning window
Speech Feature vectors	$b_i(i=1,\ldots,12)$ $\Delta b_i(i=0,\ldots,12)$, $\Delta^2 b_i(i=0,\ldots,12)$,
Training Set	30 male , 30 female 3 utterances each
Tested Set	10 male, 10 female, 3 utterance each
Acoustic Model	32-states isolated phrase HMMs
Noise varieties	4 types from NOISEX-92 (white,pink, HF radio channel, babble)
SNR	10 dB, 15 dB, 20 dB
Filtering methods	RSF, RSA,

stage, 10 dB, 15 dB, and 20 dB of the 4 types of noises are artificially added to the original speech. We compare the performance of proposed enhanced RSA of specified passbands to those by RSF under 4 types of noises; white, pink, HF channel and babble noises in MATLAB (R2014a) software. Under the stated conditions, we measure the average recognition rates for 10 speakers on RSF and 8 enhanced RSA passband specifications given as Types (a) to Type (h) at 10 dB 15 dB, and 20 dB SNR.

Table 3 shows the average recognition accuracy for 100 Japanese common male speech phrases. Table 4 shows the average recognition accuracy for Japanese similar pronunciation male speech phrases. Table 5 shows the average recognition accuracy for 100 Japanese common female speech phrases. Table 6 shows the average recognition accuracy for Japanese similar pronunciation female speech phrases.

3.3 Simulation results and analysis

Analysis is carried out for the Japanese common and similar phrases databases. We use gender (male and female) and 4 SNR (at 10 dB, 15 dB, and 20 dB) as variables. Results analysis focuses on the performance of the enhanced RSA types on the various acoustic measures. The 4 kinds of noises used in the experiments are based on Signal Processing Information Base (SPIB) noise data measured in field by Speech Research Unit (SRU) at Institute for Perception-TNO, Netherlands, United Kingdom, under the project number 2589-SAM (Feb. 1990) In this paper the model formulation is as follows: the model uses FFT based MFCC coefficients consisting of 38-dimensional feature vectors. The 38-parameter feature vector consisting of 12 cepstral coefficients (without the zero-order coefficient) plus the corresponding 13 delta and 13 acceleration coefficients is given by $[b_1 b_2 \ldots b_{12} \Delta b_0 \Delta b_1 \ldots \Delta b_{12} \Delta^2 b_0 \Delta^2 b_1 \ldots \Delta^2 b_{12}]$ where b_i, Δb_i and $\Delta^2 b_i$, are MFCC, delta MFCC and delta-delta MFCC, respectively.

3.4 Results Explanations

In Table 3 at 10 dB SNR, RSA with type (c) performs better (76.6 %) compared to RSF (72.5 %). At 15 dB SNR, RSA with type (c) performs better (90.1 %) compared to RSF (87.6 %). RSA with type (c) performs better (94.9 %) than RSF (92.8 %) at 20 dB SNR.

RSA with type (c) (1 35) performs better than RSA with type (a). For RSA with type (c), the recognition accuracy results decline (from 76.6 % to 72.6 % for type (c) and type (f) and (h), respectively) with increase in bandwidth (for (c)(1 35), (f) (0.5 35), and (h) (0.1 35)).

Overall, RSA with type (c) (1 35) performs better at the given SNR.

In Table 4 RSA with type (f) performs better (69 %) than RSF (58 %) at 10 dB SNR. RSA with types (f) and (h) perform better (67 %) than RSF (60 %) at 15 dB SNR. RSA with types (f) and (h) perform much better (73 %) than RSF (66 %) at 20 dB SNR.

At 10 dB, increase in bandwidth from RSA with type (f)(0.5 35) to RSA with type (h)(0.1 35) there is a slight decline in recognition accuracy of 1 % (from 69 % to 68 %). On the other hand, at 15 dB and 20 dB SNR similar increase in bandwidth of RSA with type (f)(0.5 35) to that of RSA with type (h) (0.1 35) shows no change in results, both at 67 % and 73 % respectively.

Overall, RSA with type (f) (0.5 35) performs better.

In Table 5 at 10 dB SNR, RSA with type (h) performs better (58.7 %) than RSF (56.3 %). RSA with type (h) is a better performer (82.7 %) among the new RSA and is better than RSF (79.9 %) at 15 dB SNR. RSA with types (c) and (d) are better performers (91.1 %) among the new RSA and their performance is better compared to RSF (89.1 %) at 20 dB SNR.

Generally, RSA with a 35 frequency component range shows a better performance than RSA with a 7 frequency component range.

For RSA with a 35 frequency component range, the recognition accuracy results increases from 55.8 % to 57.6 % and later to 58.7 % at 10 dB SNR and from 80.8 % to 82.3 % and later to 82.7 % at 15 dB SNR for RSA with type (c) (1 35), RSA with type (f) (0.5 35) and RSA with type (h) (0.1 35),respectively. At 20 dB SNR, there is a slight decline in accuracy from 91.1 % to 90.5 % for RSA with type (c) (1 35) and both RSA with types (f) (0.5 35) and (h) (0.1 35) respectively.

RSA with type (h) (0.1 35) performs better at <

Table 3: *Average recognition accuracy(%) for 100 Japanese common male speech phrases*

	Avg(%) for 4 Noises		
	10 dB	15 dB	20 dB
RSF	72.5	87.6	92.8
RSA:Type(a)	69.3	83.5	88.5
RSA:Type(b)	74.0	87.0	91.3
RSA:Type(c)	76.6	90.1	94.9
RSA:Type(d)	76.5	89.9	94.8
RSA:Type(e)	66.4	81.2	86.5
RSA:Type(f)	72.6	87.2	92.7
RSA:Type(g)	66.9	81.2	86.4
RSA:Type(h)	72.6	87.2	92.7

Table 4: *Average recognition accuracy(%) for Japanese similar pronunciation male speech phrases*

	Avg(%) for 4 Noises		
	10 dB	15 dB	20 dB
RSF	58	60	66
RSA:Type(a)	57	61	61
RSA:Type(b)	63	65	71
RSA:Type(c)	65	66	68
RSA:Type(d)	65	66	70
RSA:Type(e)	62	63	67
RSA:Type(f)	69	67	73
RSA:Type(g)	55	56	61
RSA:Type(h)	68	67	73

Table 5: *Average recognition accuracy(%) for 100 Japanese common female speech phrases*

	Avg(%) for 4 Noises		
	10 dB	15 dB	20 dB
RSF	56.3	79.9	89.1
RSA:Type(a)	51.5	75.9	84.4
RSA:Type(b)	56.3	80.3	89.4
RSA:Type(c)	55.8	80.8	91.1
RSA:Type(d)	55.3	80.5	91.1
RSA:Type(e)	55.0	80.2	88.2
RSA:Type(f)	57.6	82.3	90.5
RSA:Type(g)	55.5	80.3	88.2
RSA:Type(h)	58.7	82.7	90.5

Table 6: *Average recognition accuracy(%) for Japanese similar pronunciation female speech phrases*

	Avg(%) for 4 Noises		
	10 dB	15 dB	20 dB
RSF	55	62	71
RSA:Type(a)	60	67	70
RSA:Type(b)	60	67	70
RSA:Type(c)	62	63	73
RSA:Type(d)	58	66	75
RSA:Type(e)	60	62	69
RSA:Type(f)	57	64	69
RSA:Type(g)	62	62	69
RSA:Type(h)	59	64	68

20 dB SNR while RSA with types (c) (1 35) and (d)(1 40) perform better at $>$ 15 dB SNR.

In Table 6 RSA with types (c) and (h) show better performance (64 %) among RSA schemes and are better than RSF (57 %) at 10 dB SNR. At 15 dB SNR, RSA with type (d) performs better (72 %) than other RSA schemes and better than RSF (68 %). RSA with type (d) is a better performer (77 %) among the RSA schemes and equally performs better than RSF (75 %) at 20 dB SNR. Generally, RSA with a 35 frequency component range shows a better performance than RSA with a 7 frequency component range.

For RSA with a 35 frequency component range, the recognition accuracy shows a tendency of decline from 64 % to 62 % at 10 dB SNR and a decline from 71 % to 69 % at 15 dB SNR and from 78 % to 76 % at 20 dB SNR for RSA with type (c) (1 35) and RSA with type (f) (0.5 35),respectively.

3.5 Analysis

Conventionally, RSF is a bandpass filter in a system that reduces the amplitudes of signal components that lie outside a given frequency range. It only lets through components within a band of frequencies. Bandpass filters are particularly useful for analysing the spectral content of signals. The proposed RSA simulates bandpass filtering by processing selected frequency components in modulation spectrum domain.

Experimental results show that the proposed RSA performs better than conventional RSF. In the case of Japanese common speech phrases for male speaker in Table 3, new RSA with type (c) (1 35) produce better results while for Japanese similar pronunciation male speech phrases in Table 4, new RSA with type (f) (0.5 35) show better performance among the evaluated specifications.

In the case of Japanese common female speech phrases in Table 5, the proposed RSA with type (h) (0.1 35) show better results while for Japanese similar pronunciation female speech phrases in Table 6, the proposed RSA with type (c) (1 35) and RSA with type (g) (0.1 7) at 10 dB, the new RSA with type (a) (1 7) and with type (b) (1 15) at 15 dB, and the RSA with type (d) at > 15 dB SNR perform better, respectively.

Based on the experimental results, for male NSP we found the most effective method to be RSA with type (c) (1 35) at all SNR under consideration while for male SP RSA with type (f) (0.5 35) was better at > 10 dB SNR. In the case of female speaker, the results indicate that for NSP the most effective method is RSA with type (h) (0.1 35) at < 20 dB SNR, while at > 15 dB SNR, RSA with type (d) (1 40) show better performance. For SP, RSA with type (h) (0.1 35) is better at < 15 dB SNR while RSA with type (d) (1 40) performs better at > 10 dB SNR.

4. DISCUSSION

In this section, we discuss the findings of our experiments. We show the positive contributions in applying the proposed enhanced RSA types with high frequency components on isolated speech recognition. By using a different number of frequency components, we mimic bandpass filtering to isolate each frequency region of the signal in turn so that we can measure the energy in a selected region. The same process is applied both on male and female speech recognition. Table 7 shows the average improvement on recognition accuracy for the better performers at each SNR.

Table 7: *Average recognition improvement(%)*

	Avg improvement(%)		
	10 dB	15 dB	20 dB
Male, NSP	4.1	2.5	2.1
Male, SP	11	7	7
Female, NSP	2.4	2.8	2.0
Female, SP	7	4	2

Both, the speech type (NSP and SP) and SNR (at 10 dB, 15 dB, and 20 dB) tend to have an influence on performance of proposed method hence the difference in results. The results indicate that proposed enhanced RSA depends on the input signal. Although in each speaker and speech categories there is a enhanced RSA type that shows a superior performance. Both the wide band and narrow band perform differently on male and female speech phrases. For instance, male SP has 11 % improvement at 10 dB compared to 7 % for female SP. Our proposed method shows improved performance on male SP compared to female SP (11 %, 7 %, 7 %, versus 7 %, 4 %, 2 %,) at 10 dB, 15 dB, and 20 dB, respectively. On the other hand, results for male NSP versus female NSP are given as (4.1 %, 2.5 % 2.1 % versus 2.4 %, 2.8 %, and 2.0 %), respectively. It has been observed that under the experimental conditions, male NSP is better than female NSP at 10 dB , while female NSP is slightly better than male NSP at 15 dB.

The accuracy of a speech recognition system can be defined as the percentage of time that the recognizer correctly identifies an input utterance. Recognition errors can be generally classified as misrecognitions or as nonrecognition errors. The tendency of differences in recognition accuracy between male and female can be attributed to many factors including user characteristics(age, sex), the language (vocabulary size), and the channel and environment (noise), for example, among many others [29]. The more varied the group of speakers using the system, the more challenging the recognition process. It is more difficult for a speaker-independent system to recognize accurately both male and female speakers.

The most limiting problem of larger vocabulary sizes is the corresponding decrease in recognizer accuracy. This refers to the total number of different phrases the speech recognizer is able to identify. Therefore, the tendency of differences in recognition accuracy between the 100 Japanese phrases and the Japanese similar pronunciation phrases is due to the differences in sizes of databases. A smaller database (of similar pronunciation phrases) has an increased chance of better recognition accuracy compared to a much larger database (of 100 Japanese phrases), in this case. In the latter, increased number of misrecognitions and false recognitions are often recorded as a result compared to in the former.

5. CONCLUSION

The paper proposes to use running spectrum analysis (RSA) with certain passbands for noisy speech recognition. Performances of speech recognition for Japanese short phrases are compared with those by running spectrum filtering (RSF). Experiments are conducted for various passbands, and the results show an advantage over RSF method.Filtering is optimized as in the case of RSA.

Theoretical analysis indicates the proposed RSA bandpass schemes are less complex to realize and experimental results demonstrate the effectiveness of the proposed approach in improving the robustness of automatic isolated phrase recognition.

From the experimental results it has been demonstrated that the use of RSA with high frequency components, particularly the ones in the range of (0.5 35), for example can be useful in ASR. In this study, RSA on a 35 frequency component range shows a better performance than RSA on a 7 frequency component range used in other related research study. Under noisy conditions different types of RSA show different performance for male and female speakers. It has also been discovered that in the case of male speakers system performance is influenced mostly by the RSA type while that of female speakers, the performance relies mostly on SNR. In future we plan to evaluate our proposed method on recognizing children's speech and develop a recognition system that can distinguish between a child voice and that of an elderly person.

ACKNOWLEDGEMENT

The authors would like to thank Raytron, inc, Japan for fruitful discussions. This study is supported in parts by the Japan Science and Technology Agency for A-Step Program (AS2416901H).

References

[1] M. Watanabe, H. Tsutsui and Y. Miyanaga,"Robust speech recognition for similar pronunciation phrases using MMSE under noise environments," *Proc. 13th International Symposium on Communications and Information Technologies (ISCIT)*, Surat Thani, pp.802-807, 2013.

[2] J. Tierney, "A study of LPC analysis of speech in additive noise," *IEEE Trans. on Acoustic., Speech, and Signal Process.*, vol. ASSP-28, no. 4, pp. 389-397, Aug. 1980.

[3] S. Kay, "Noise compensation for autoregressive spectral estimation," *IEEE Trans. on Acoust., Speech, and Signal Process.*, vol. ASSP-28, no. 3, pp. 292-303, Jun 1980.

[4] P. B. Patil, "Multilayered network for LPC based speech recognition," *IEEE Transactions on Consumer Electronics*, vol. 44, no. 2, pp. 435-438, May 1998.

[5] Mark G. Hall, Alan V. Oppenheim, and Alan S. Willsky,"Time-varying parametric modeling of speech," *Signal Processing*, vol. 5, pp. 267-285, 1983.

[6] S. Tanweer, A. Mobin and A. Alam,"Analysis of Combined Use of NN and MFCC for Speech Recognition," *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, vol. 8, no. 9, 2014.

[7] L. Muda, M. Begam and I. Elamvazuthi, "Voice Recopgnition Algorithm using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques," *Journal of Computing*, vol. 2, no. 3, pp. 138-143, 2010.

[8] C. Ittichaichareon, S. Suksri, and T. Yingthawornsuk, "Speech Recognition Using MFCC," *International Conference on Computer Graphics, Simulation and Modelling (ICGSM2012)*, pp. 135-138, 2012.

[9] Anjali Bala, Abhijeet Kumar, Niddhika Birla,"Voice command recognition system based on MFCC and DTW," *International Journal of Engineering Science and Technology*, vol. 2, no. 12, pp. 7335-7342, 2010.

[10] Petr Motliček, "Feature Extraction in speech coding and recognition," *Report of PhD research internship in ASP Group*, OGI-OHSU, 2002,

[11] K. Yao, K. K. Paliwal and S. Nakamura, "Model-based noisy speech Recognition with Environment Parameters Estimated by noise adaptive speech Recognition with prior," *EUROSPEECH 2003-GENEVA*, Switzerland, Tech. Rep., 2003.

[12] Q. Zhu, N. Ohtsuki, Y. Miyanaga, and N. Yoshida,"Robust speech analysis in noisy environment using running spectrum filtering," *International Symposium on Communications and Information Technologies*, vol. 2, pp. 995-1000, Oct. 2004.

[13] N. Ohtsuki, Qi Zhu and Y. Miyanaga, "The effect of the musical noise suppression in speech noise reduction using RSF," *International Symposium on Communications and Information Technologies*, vol. 2, pp. 663-667, Oct. 2004.

[14] V Tyagi, I. McCowan, H. Misra, and H. Boulard, "Mel-Cepstrum modulation spectrum (MCMS) features for Robust ASR," *in Proc. 2003 IEEE Workshop on Automatic Speech Recognition and Understanding*,St. Thomas, pp. 399-404, 2003.

[15] Dimitrios Dimitriadis,Petros Maragos, and Alexandros Potamianos, "Modulation features for Speech Recognition," *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2002.

[16] Jeih-Weih Hung, Hsin-Ju Hsieh, and Berlin Chen, "Robust Speech Recognition via Enhancing the Complex-Valued Acoustic Spectrum in Modulation Domain," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 24, Issue 2, pp. 236-251, Feb. 2016.

[17] K. Ohnuki, W. Takahashi, S. Yoshizawa, and Y. Miyanaga, "New acoustic modeling for robust recognition and its speech recognition system," *International Conference on Embedded Systems and Intelligent Technology*, 2009.

[18] S. Yoshizawa and Y. Miyanaga, "Robust recognition of noisy speech and its hardware design for real time processing.," *ECTI Trans. Elect., Eng., Electron., and Commun.*, vol.3, no.1, pp. 36-43, Feb. 2005.

[19] K. Ohnuki, W. Takahashi, S. Yoshizawa, and Y. Miyanaga, "Noise Robust speech features for Automatic Continuous Speech Recognition using Running Spectrum Analysis," *in: Proc. of 2008 International Symposium on Communications and Information Technologies (ISCIT)*, pp. 150- 153 (October 2008).

[20] Yiming Sun and Yoshikazu Miyanaga, "A Noise-Robust Continuous Speech Recognition System Using Block-Based Dynamic Range Adjustment," *IEICE Trans. INF. & SYST*, vol.95-D, no.3, March 2012.

[21] T. Chi, Y. Gao, M. C. Guyton, P. Ru and S. Shamma, "Spectro-temporal modulation transfer functions and speech intelligibility," *J. Acoust. Soc. Am.*, 106(5), pp. 2719–2732, 1999.

[22] Naoya Wada and Yoshikazu Miyanaga, "Robust Speech Recognition with MSC/DRA Feature Extraction on Modulation Spectrum Domain," *in Proc. Second International Symposium on Communications, Control and Signal Processing (ISCCSP)*, Marakech, Morocco, Mar. 2006.

[23] M.A Anusuya and S.K. Katti, "Speech Recognitionb by Machine: A Review," *International Journal of Computer Science and Information Security*, (IJCSIS), vol. 6. no. 3, pp. 181-205, 2009

[24] Noboru Kanedera, Takayuki Arai, Hynek Hermansky and Misha Pavel, "On the importance of various modulation frequencies for speech recognition," *Proceedings of EUROSPEECH 97*, Rhodos, Greece, Sep. 1997.

[25] Hynek Hermansky, Eric Wan, and Carlos Avendano, "Speech enhancement based on temporal processing," *IEEE International Conference on Acoustic, Speech and Signal Processing*, Detroit, Michigan, Apr.1995.

[26] Carlos Avendano, Sarel van Vuuren and Hynek Hermansky, "On the properties of temporal processing for speech in adverse environments," *Proceedings of 1997 Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk Mountain House, New Paltz, New York, October 18-22, 1997.

[27] Eslam Mansour mohammed, Mohammed Shraf Sayed, Abdalla Mohammed Moselhy and Abdelaziz Alsayed Abdelnaiem, "LPC and MFCC Performance Evaluation with Artificial Neural Network for Spoken Language Identification," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 6, no. 3, Jun. 2013.

[28] M. H. Moatta and M. M. Homayounpour, "A Simple but Efficient Real-Time Voice Activity Detection Algorith," *17th European Signal Processing Conference (EUSIPCO)*, August 24-28, 2009.

[29] Sherry P. Casall and Robert D. Dryden, "The Effects of Recognition Accuracy and Vocabulary Size Of A Speech Recognition System on Task Performance and User Acceptance," *Industrial Engineering and Operations Research*, 1988.

George Mufungulwa received his B.S. degree in Computer Science from The Copperbelt University, Kitwe, Zambia, in 1999 and his M.Sc. in Distributed Interactive Systems from University of Lancaster, UK, in 2002. He is currently a Ph.D. student at Graduate School of Information Science and Technology, Division of Media and Network Technologies, Hokkaido University, Sapporo, Japan. His research interests include digital signal processing and multimedia systems. He is a member of IEEE and IEICE.

Hiroshi Tsutsui received his B.E. degree in Electrical and Electronic Engineering and his master and Ph.D. degrees in Communications and Computer Engineering from Kyoto University in 2000, 2002, and 2005, respectively. He is currently an associate professor in Division of Media and Network Technologies, Hokkaido University. His research interests include circuits and systems for image processing and VLSI design methodology. He is a member of IEEE, ACM, IPSJ, IEEJ, and IIEEJ.

Yoshikazu Miyanaga received the B.S., M.S., and Dr. Eng. degrees from Hokkaido University, Sapporo, Japan, in 1979, 1981, and 1986, respectively. Since 1983 he has been with Hokkaido University. He is now Professor at Division of Information Communication Systems in Graduate School of Information Science and Technology, Hokkaido University. From 1984 to 1985, he was a visiting researcher at Department of Computer Science, University of Illinois, USA. His research interests are in the areas of speech signal processing, wireless communication signal processing and low-power VLSI system design.

Shinichi Abe developed an automotive electronics products at Pioneer Co. and he is currently a stuff of Vehicle Information and Communication System Center (VICS Center), Business Research division, Tokyo, Japan.

Pattern Based Motion Estimation using Zero Motion Pre-judgement and Quantum behaved Particle Swarm Optimization

Yogananda Patnaik [1] and **Dipti Patra** [2], Non-members

ABSTRACT

Motion estimation is a fundamental and resource hungry operation in most of the video coding applications. The most popular method used in any video coding application is block matching motion estimation (BMME). The conventional fast motion estimation algorithm adopts monotonic error surface for faster computation. However, these search techniques may trap into local minima resulting in erroneous motion estimation. To overcome this issue, many evolutionary swarm intelligence based algorithms were proposed. In this paper, a pattern based motion estimation using Zero motion prejudgment and Quantum behaved Particle Swarm Optimization (QPSO) algorithm is proposed named as Pattern Based Motion Estimation (PBME) algorithm. The notion of QPSO improves the diversity in the search space which enhances the search efficiency and helps in reduction of the computational burden. At the same time, QPSO needs fewer number of parameters to control. Therefore the proposed algorithm enhances the estimation accuracy. An initial search pattern (Hexagonal Based Search) is being used which speeds up the convergence rate of the algorithm. From the simulation results, it is found that the proposed method outperforms the existing fast block matching (BMA) algorithms of the search point reduction by 40%-75%.

Keywords: PSO; QPSO; Motion Estimation; Motion Vector (MV)

1. INTRODUCTION

The most crucial component of any block based video coding system is motion estimation. This is mainly due to the use of temporal redundancy between progressive frames of video [1]. Consecutive frames of a video sequence have high correlation between them. So high coding efficiency can be achieved in any video coding system, with reduction in temporal redundancy. Exploitation of temporal redundancy between successive frames in a video codec is possible with motion compensation technique by predicting the current frame from the past frame. Motion estimation (ME) has a vital role in inter frame prediction. From the available motion estimation techniques, block-matching algorithm (BMA) is considered to be the most popular one, because of its simplicity and is adopted in most of the video coding standards namely H.264/AVC and H.265/HEVC [2]. In case of BMA frames are broken down to blocks, and individual motion vectors are calculated for each block. Here for each block in the current frame, the motion estimation process is applied to and the motion vector, is the best matching block, in the reference frame. This best matching block becomes the predictor for the current block. The most accurate and computationally expensive method is full search or the exhaustive search because the block- matching is carried out on block by block basis for each and every blocks [3]. Hence this methods seeks for an urgent need for reduction in the computational load while maintaining the quality of the achieved motion information and to find an optimal solution by calculating all possible candidate blocks in the reference frame within its search area.

The application areas of video coding includes fixed and mobile telephony, video conferencing application, DVD and HDTV applications [4]. In case of Block Based Motion Estimation (BBME) frames are divided into non overlapping blocks, and each block of the current frame is matched block by block connected to a search window in the reference frame based on minimum matching criteria. This can be either mean absolute difference (MAD) or sum of absolute difference (SAD), etc. The BBME can be regarded as a problem of optimization. The full search algorithm gives the definite optimal solution because it searches block by block entirely. This method cannot be used in real life scenario because of large computational burden. To overcome this limitation many fast search techniques are available in the literature for various video coding standards like H.26x series, MPEG series, and HEVC. The vital factors to develop ME algorithms are search pattern, choice of initial centre and search strategy. These three factors are used for the calculation of machine performance and the peak signal to noise ratio (PSNR) [1] of the algorithm.

Manuscript received on April 22, 2017 ; revised on May 15, 2017.

Final manuscript received on May 30, 2017.

[1,2] The authors are with DDepartment of Electrical Engineering National Institute of Technology Rourkela, India, E-mail: ypatnaiknitrkl@gmail.com and dpatra@nitrkl.ac.in

2. RELATED WORKS

To fulfil real-time processing needs in the multimedia applications, the speed of ME algorithms needs to be increased. Therefore faster motion estimation search methods became popular and are available in literatures where the research is still thrust provoking to reduce the computational cost of full-search algorithm. The fast motion estimation methods includes three-step search (TSS) [5], four-Step search (FSS), N-step search (NSS) [6], diamond search (DS) [7], cross diamond search methods (CDS) [8], adaptive rood pattern search (ARPS) [9], and hexagonal search (HEXPBS) [10] etc.

By using these search, techniques will be liable to be trapped in the local optima on the error surface. The basic hypothesis behind these search methods coins from a theme that the block distortion measure is reduced monotonously if the search points moves from the utmost point towards the optimum point. Hence these fast search methods get trapped in local best solutions instead of achieving the global best solution. To overcome the problem of local minima, various approaches are available in the literature for the use of a global optimization method for solving the difficulty of motion estimation. Among available advanced methods, the genetic algorithm is one of them. These algorithms are more complex and computationally expensive. Simulated annealing is also applied adaptively of the search process by choosing the intense search region. Other algorithms like the ant bee colony [11] and different evolution algorithms are also proposed for motion estimation. However few attempts have been made in the literature for the use of particle swarm optimization (PSO) for solving the problem of motion estimation [3, 12, 13, 14]. But these methods have intense computational complexity and low estimation accuracy as compared to the other traditional methods. To increase the speed of the traditional PSO algorithm presented in [3], the starting point of the particles uses predetermined pattern instead of the random pattern. The initialization of the particles is either square or diamond pattern about the centre.

Modified PSO algorithm was proposed in [1] to fulfil the requirement of low computational load while preserving high motion estimation accuracy. To speed up and also to increase the motion estimation accuracy various schemes are implemented and applied to the motion estimation process. The modified PSO with certain strategies produce remarkable improvement regarding both estimation accuracy and computational efficiency over the state-of-art methods. But this can be further increased by using more accurate block matching process which can still lower the computational complexity. One of the improved and advanced PSO method for higher accuracy and low computational complexity is QPSO. This paper deals with pattern based motion estimation technique using QPSO and Zero motion prejudgment. QPSO is used here because a particle remains in the bound state, which can appear at any point in the entire search space with definite probability, even if the position is far away from the learning inclination point.

This QPSO algorithm outperforms the conventional PSO Algorithm due to global convergence and the quantum system has more states than a linear system. In the proposed method initially, a video sequence is given as input followed by extraction of frames. Each frame is divided into non overlapping blocks. Before going for motion estimation, zero motion prejudgment is being performed as a preprocessing step to check that whether the blocks are static or in motion. If the blocks are static, then there is no necessity for finding out motion estimation. This step reduces the computation and saves memory. Then a predefined or fixed search pattern namely hexagonal search pattern is used for distribution of the search in the search window and QPSO algorithm is applied for a faster search in motion estimation process. Here P_{best} and G_{best} are calculated. The proposed method saves a lot of memory and at the same time speeds up the process substantially.

The rest of this paper is organized as follows: Section 2 renders materials and methods which talk about background of the basic motion estimation technique, evolutionary algorithms like PSO and QPSO. It also reviews some of the state-of-arts on the evolutionary approach towards motion estimation. Section-3 focuses on the proposed method for motion estimation using QPSO and zero motion prejudgement. The proposed technique is validated and verified using various performance measures and metrics like Peak Signal to Noise Ratio (PSNR), Mean Structural Similarity Index (MSSIM), Feature Similarity Index (FSIM) and Universal Quality Index (UQI) which are discussed in Section-4. Finally, the concluding remarks are presented in the Section-5.

The major contributions of this paper are:

- Motion estimation with evolutionary approach
- Zero motion prejudgement to find out the blocks which are in motion and non-motion.
- Use of QPSO algorithm with Hexagonal based pattern for finding motion vector in motion field with fewer search points

3. MATERIALS AND METHODS

3.1 Motion Estimation

Motion estimation is applied in most of the video applications scenario namely video compression, segmentation and video tracking. The most popular and widely accepted algorithm for motion estimation is the block-matching algorithm (BMA) [15]. In case of block matching a frame is broken down into various rectangular blocks and the motion vector for all the blocks are estimated on block by block basis over a

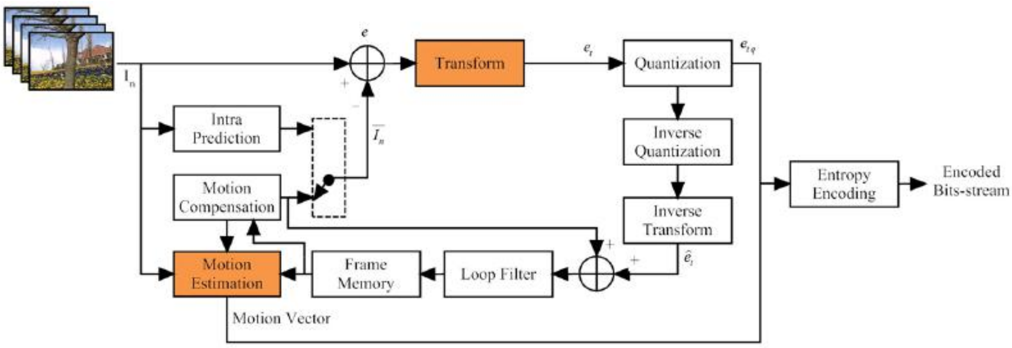


Fig.1: *Basic H.264 Codec structure.*

search range in the reference frame by calculating the nearest match block of pixels as per the defined matching criteria mainly sum of square difference (SSD) or sum of absolute difference (SAD). Motion estimation can achieve substantial amount of compression by exploiting temporal redundancy which exists in most of the video sequences.

Various motion estimation methods were discussed by many researchers for seeking interest in reduction of computational complexity namely block-matching algorithm, parametric based models [16], optical ow models [17] and pel recursive techniques [18]. Block-matching algorithm seems to be the most popular one among all the available models. This is because of the effectiveness and simplicity of the algorithms for both hardware and software applications. The predictor block is the best matching block where the displacement is given by the translational motion vector (MV). Hence block-matching can be treated as an optimization problem to find the best matching block in a predefined search space. The full search algorithm [19] provides an optimal solution on minima matching error because this algorithm takes for all the candidate blocks at a time. But the computational burden is increased substantially which restricts its use in real-time video application. The sum of absolute difference can be calculated as

Gray value of $(i, j)^{th}$ pixel in the current frame, and $f_{t-1}(i, j)$ shows the gray value of $(i, j)th$ pixel in the previous frame. The displacement vector u, v of the candidate block with the minimum $SAD(u, v)$ gives the motion vector. The size of block is $N \times N$ and u, v are in the range $[-ww]$. The basic H.264 codec structure is shown in Fig. 1. The basic motion estimation process using block matching algorithm is illustrated in Fig. 2.

3.2 Overview of QPSO

In a PSO system, the search area cannot cover the entire viable region, and global convergence is not assured. This is the important cause of the early termination in PSO. To conquer this issue of PSO, quantum-behaved particle swarm optimization

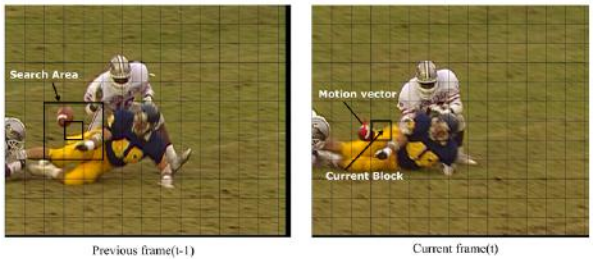


Fig.2: *Block Matching Process.*

(QPSO) is proposed [20]. In the QPSO algorithm, the positioning of a particle is portrayed by its local attractor and by probability density function. By adopting this approach, there is no limit on the trajectory. Further, there is only a single parameter in QPSO. In modern years, QPSO is successfully employed to a whole lot of optimization problems, namely constrained optimization, multi-objective optimization,

$$SAD_{(u,v)} = \sum_{i=1}^{N}\sum_{j=1}^{N} |f_t(i, j) - f_{t-1}(i + u, j + v)| \quad (1)$$

where $-w \leq u, v \leq w(u, v)$ is the distance of the candidate block to the current block, $f_t(i, j)$ implies the Engineering design, and many more [22].

QPSO, a probabilistic algorithm, was proposed by Sun, Feng et al. [23]. The detail of the QPSO algorithm is depicted in Algorithm 1. The source of inspiration in QPSO algorithm is mainly quantum mechanics of particles and the trajectory analysis of PSO which was formulated by Clerk and Kennedy in 2002. From the

Trajectory analysis of PSO it can be observed that every particle in PSO algorithm oscillates around and converges to local minima or remain in bound state. But in case of QPSO, the particles are enduring quantum behaviour and be in a bound state. Further, it is also assumed that the particles are attracted towards the quantum potential well which is centred on its

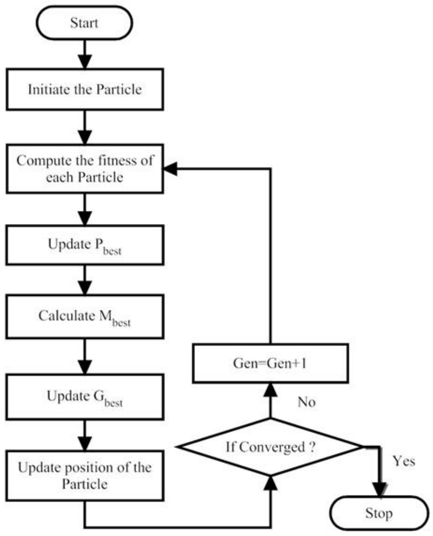


Fig.3: *Flowchart for QPSO Algorithm.*

local point giving rise to a stochastic update equation. Later on, mean best position was introduced to the algorithm to increase the global search capacity of this QPSO [23]. The flowchart of the QPSO algorithm is shown in Fig. 3. Updating the position of particles is done as follows

$$X^l_{k,n+1} = p^l_{k,n} \pm \frac{L^l_{k,n}}{2} \ln(\frac{1}{u^l_{k,n}}) \quad (2)$$

where $X^l_{k,n+1}$ is the current position vector at $(n+1)^{th}$ iteration, $u^l_{k,n}$ is a random number which is uniformly distributed in (0,1) and $p^l_{k,n}$ [24] is a local attractor which ranges from $p^1_{k,n}, p^2_{k,n}, \ldots, p^N_{k,n}$

$$p^l_{k,n} = \varphi^l_{k,n} P^l_{k,n} + (1-\varphi^l_{k,n}) P^l_{g,n} \quad (3)$$

and $p^l_{k,n}$ is the local best previous position which is a vector that ranges from and $p^l_{k,n} = p^1_{k,n}, p^2_{k,n}, \ldots, p^N_{k,n}$.

$$L^l_{k,n} = 2.\beta \left| p^l_{k,n} - X^l_{k,n} \right| \quad (4)$$

where $\varphi^l_{k,n}$ is a random number uniformly distributed (0,1) also

$$X^l_{k,n+1} = p^l_{k,n} \pm \beta |p^l_{k,n} - X^l_{k,n}| \ln(\frac{1}{u^l_{k,n}}) \quad (5)$$

where β being the contraction-expansion (CE) coefficient that is used as a tuning factor for convergence speed of the algorithm.

The particle positioning is updated according to the equation

$$X^l_{k,n+1} = p^l_{k,n} \pm \beta |P^l_{k,n} - X^l_{k,n}| \ln(\frac{1}{u^l_{k,n}}) \quad (6)$$

Finally, the QPSO equation is given as

$$X^l_{k,n+1} = p^l_{k,n} \pm \beta |mbest^l_n - X^l_{k,n}| \ln(\frac{1}{u^l_{k,n}}) \quad (7)$$

range of β is set as 1 and then gradually reduces to 0.5 linearly. The criteria for choosing this β value is by trial and error method. In our case, the most suitable value of β is found to be 0.5 for better performance of the algorithm.

4. PROPOSED METHOD

The proposed method deals with a pattern based motion estimation technique using QPSO and Zero motion prejudgment. The flow chart of the proposed algorithm is given in Fig. 5. Here initially a video sequence is given as input followed by extraction of frames. Then these individual frames are partitioned into non overlapping blocks. The images are partitioned into blocks of 8x8 pixels which are called as macroblocks. The macro block partition is hierarchical on the block position, and therefore every macro block consists of four blocks.

Before going for motion estimation, zero motion pre-judgement is being performed as a pre-processing step to check that whether the blocks are static or in motion. If the blocks are static, then there is no necessity for finding out motion estimation. This step reduces the computation and saves memory. Then after these non-static blocks are considered and then to it a predefined search pattern namely hexagonal search pattern is used for distribution of the search in the search window. Then QPSO algorithm is applied for a faster search in motion estimation process. Here P_{best} and G_{best} are calculated.

The block diagram of the proposed method is shown in Fig. 4 and it consists of broadly four parts:

1. Pre-processing
2. Zero Motion prejudgement
3. Choice of Fixed Pattern.
4. Efficient search using QPSO

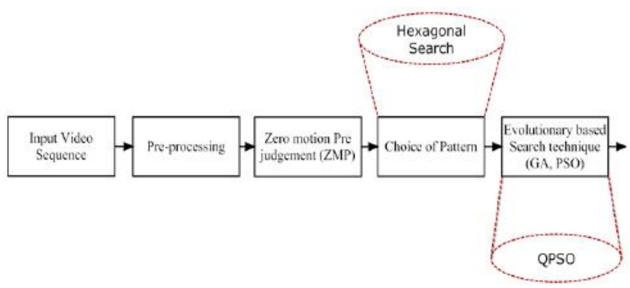


Fig.4: *Proposed method using QPSO.*

Algorithm 1: QPSO Algorithm

begin

Initialization of actual positions and the P_{beat} positions of all particles; Evaluation of their fitness value and global best positions G_0 and Set $n = 0$;

while the end point condition is not satisfied do

choose an appropriate value for α;

for $i = 1$ to M do

Objective function value evaluation $f(X_i, n)$;

Update $P_{i,n}, n$ and G_{ni};

for $j = 1$ to N do

$\sigma_{i,n} = rand1(\cdot)$

$P_{i,n} = \varphi_{i,n} p_{i,n} + (1 - \varphi_{i,n}) g_n$

$u_{i,n+1} = rand2(\cdot)$;

if $rand3(\cdot) < 0.5$ **then**

$X_{i,n+1} = P_{i,n} + \alpha |X_{i,n} - p_{i,n}| \ln(1/u_{i,n+1})$

else

$X_{i,n+1} = p_{i,n} - \alpha |X_{i,n} - p_{i,n}| \ln(1/u_{i,n+1})$

end

end

end

end

Set $n + n + 1$;

end

end

4.1 Pre- Processing

The pre-processing step consists of the following steps:

1. Conversion of input video to frames
2. Conversion of colour image to gray scale
3. Resizing the image frames
4. Block partitioning

The first step is conversion of the input video data to number of frames. These frames are converted to the grayscale images. Then the frames are resized as per necessity of up sampling or down sampling. These resized frames are converted to blocks using block partitioning, which is done by converting the entire frames to 8×8 blocks.

4.2 Zero motion prejudgement

Successive video frames in a video sequence consists of around 70% of the macro blocks that are static, which do not require any further search. So to calculate the static macro blocks the zero motion prejudgement (ZMP) mechanism is adopted before the start of the actual motion estimation process. Due to this a considerable amount of reduction in computation is possible, and the left over search becomes faster and thus saves memory. To check if a block is stationary or not, the block distortion is measured and then compared with a predefined threshold value T. If the distortion value is below the threshold value, then it is called as stationary block and the search process is stopped. Thus this resultant motion vector is (0, 0). The detail process of determining the threshold value is found in the literature [25].

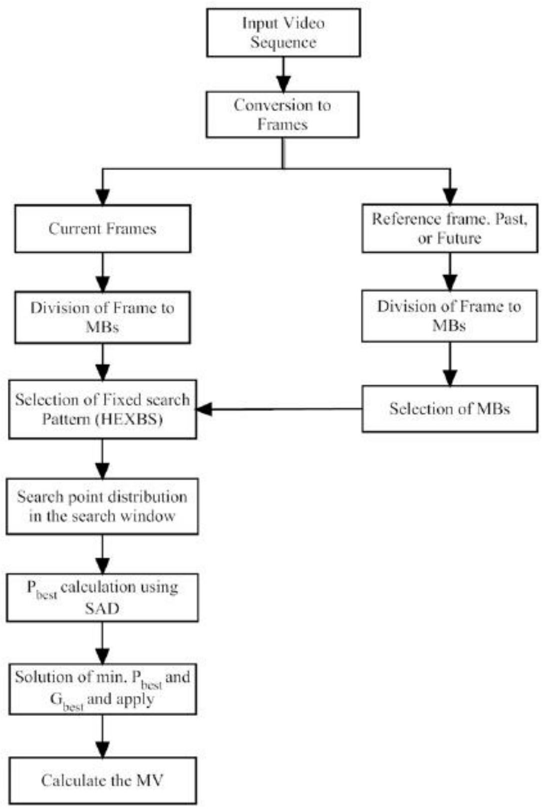


Fig.5: *Flowchart of the proposed algorithm*

Algorithm 2: Proposed PBME using Zero motion Pre-judgement and QPSO

```
begin
    Initialization of actual positions and the P_best positions of all particles;
    for every frame i do
        for every macro block j do
            zmpc ← SAD(I_{i−1}(j), I_i(j)),;
            if zmpc < t then
                the macro block (MB) is static
                MV = [0, 0]
                Continue
                    else if MB_j is in the leftmost column of frame i then
                        if MB_j == 1 then
                            initial particles are in square pattern
                            else if MB_j is in the bottom right column then
                                initial particles are in diamond pattern
                                else
                                  MB_j is in the bottom right corner column initial particles are in cross
                                    diamond pattern
                                end
                                else
                                    initial particles are in Hexagonal pattern
                            end
                        end
                    end
                    start the QPSO search
                    for each iteration time t do
                        for each particle p do
                            Evaluate the SAD using equation 1
                            update P_best and G_best
                            update position
                        end
                    end
        end
    end
end
```

4.3 Choice of Fixed Pattern: Hexagonal Based search

An assumption made by Zhu et al. [10] is that when the distortion within a small neighbourhood increases monotonically, a circular shaped search method is more effective for obtaining the highest search speed uniformly [27]. The triangle or diamond shape cannot be approximated with circle to that of the Hexagon. This pattern consists of seven points being used for the search out of which the centre point is enclosed by other six points of a Hexagon. Among the six points, two horizontal points are at a distance of 2 from the centre and the remaining four points are at a distance of p=5 from the centre. These six end points are uniformly distributed around the centre. Each time the centre point moves to any of the six points then at that point three new points are coming out, and the remaining three points are overlapped. This search process continues till it comes to an edge. The hexagonal search pattern and the motion estimation direction is shown in Fig. 7

Table 1: *Video Data Set Configuration.*

Sequence	Y_RES	UV_RES	Freq.	N F	CF	Class
Silent	720×576	360×576	50	300	4:2:2	A
News	720×486	360×486	60	300	4:2:2	A
Akiyo	720×486	360×486	60	300	4:2:2	A
Container	720×486	360×486	60	300	4:2:2	B
Hall Object	720×486	360×486	60	300	4:2:2	A
Mother Daughter	720×480	360×240	60	300	4:2:0	A
Foreman	720×576	360×576	50	300	4:2:2	B
Silent	720×576	360×576	50	300	4:2:2	A
Coastguard	720×486	360×486	60	300	4:2:2	B

4.4 Efficient search using QPSO

QPSO based approach is an efficient search technique as compared to the conventional fast search techniques because a particle remains in the bound state, which can be present at any point in the entire search space with definite probability, even if the position is far away from the learning inclination point. Thus in a quantum system, principle of superposition holds good, and so this system has far more states than a linear system. Also in a quantum system, a

Fig.6: *Video data set (a) Garden (b) Football (c) Tennis (d) Clare [26].*

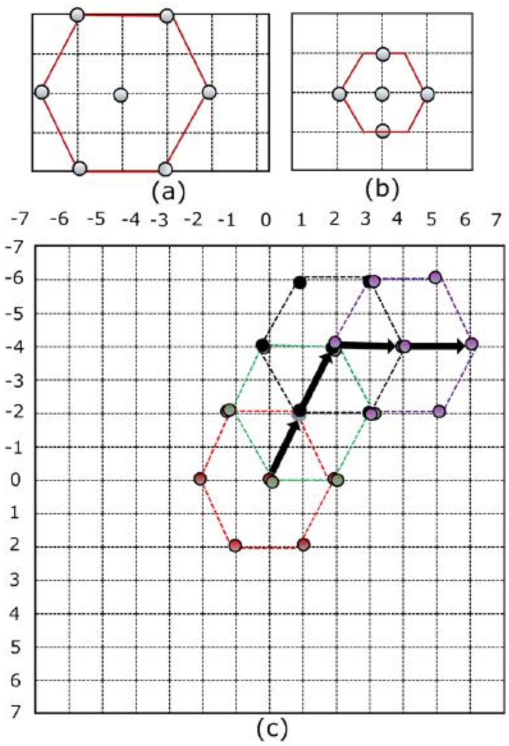


Fig.7: *Hexagonal Pattern: (a) Large Hexagonal pattern; (b) Small hexagonal pattern; and (c)Example of a search path locating the motion vector (+4,-4).*

particle can be present in any point with a certain probability distribution as this has no fixed Trajectory [28]. Hence QPSO algorithm has a global convergence property and hence reduces the computational load resulting in an efficient search method.

5. SIMULATION RESULTS

This section deals with a detail discussion of the various evaluation parameters which are used to quantify the effectiveness of the proposed pattern based QPSO motion estimation algorithm. A detailed comparative performance of the proposed algorithm with the state of the art is presented latter in this section.

5.1 Experimental Environment

This section deals with the simulation results of the proposed QPSO based ME compared to those of the existing conventional algorithms. The simulations are carried out with the standard benchmark datasets as shown in Table 1. These standard videos have different formats degrees and types of motion. They are QCIF: 176×144, CIF: 352×288 and SIF: 352×240. The Clare video is gentle, smooth with low motion variation, which consist of mostly stationary blocks. Whereas the Tennis video is complex and contains medium motion activity. But the sequences like Garden and Football consist of high motion activity, which are dependent on camera panning and complex motion activity.

5.2 Experimental results and discussion

The input individual frames are partitioned into macro blocks of size 8×8 with a maximum displacement within the search range of ±7 pixels in both directions. For the performance comparison of the proposed method, various conventional search algorithms are simulated and implemented. The motion displacement 'p' or the search range has direct impact on both computation complexity and prediction quality of the block matching technique. A small p provides poor compensation for areas with faster motion and thus resulting in poor prediction quality. On the other hand large ± p results in good prediction quality, thereby increasing the computational load since there are $(2p+1)^2$ blocks. A large p will produce longer MV thus increasing the motion overhead. So, in general, the maximum permissible displacement of p=±7 pixels is sufficient for low bit rate environment.

Table 2: *Average PSNR of different methods with various data sets.*

Test VS	FS	NTSS	DS	ARPS	PSO	QPSO
Tennis	25.6	23.85	23.63	23.62	24.68	25.02
Football	18.7	17.53	17.31	17.41	17.91	18.25
Clare	39.8	38.89	38.89	38.88	39.63	39.71
Garden	20.2	18.83	18.24	18.79	19.26	19.51

The simulations are based on the performance indices like PSNR, SSIM, and number of search points. The simulations were performed on core i7 processor 3.7 GHz with 4GB RAM. To compare the performance of the proposed method with the standard existing methods namely FS, TSS, 4SS, ARPS and DS were implemented. Average value of PSNR and search points are calculated to provide the statistical significance of the proposed method. As this is an iterative process, the permissible total number of iterations $It_{\max}$ has to be limited. The simulation results are carried out with the initial particle position of a hexagonal based pattern. Table 2 illustrates the average PSNR value of video frame obtained using various methods. From this table, it can be seen that the FS has maximum PSNR value for all the data sets. Several experiments were carried out and verified with all the standard datasets, but here only four datasets results are included.

The search efficacy is usually calculated by finding the number of search points used for the estimation of motion vector. Table 3 shows that the proposed QPSO algorithm takes the smallest number of average search points which is in the range of 7.23 to 9.52 as compared to other state of the art. The simulations were tested for all the four data sets as shown in Fig. 6. Less number of search point reduce the computational burden which enhances the efficacy of motion estimation following which video coding efficiency is improved.

It is found that, PSO takes 20 iterations to converge, whereas the proposed motion estimation algorithm converges in only 5 iterations. The comparison of computation time is presented in Table 6.

Table 3: *Average PSNR of different methods with various data sets.*

Algorithms	Garden	Clare	Football	Tennis
FS	190.5614	190.5614	190.5614	190.561
NTSS	22.1077	22.2635	22.1451	22.098
DS	11.7323	12.6424	12.6715	15.386
ARPS	10.6322	11.9778	12.7612	16.349
PSO	8.5291	9.6247	10.8664	11.051
Proposed	7.2315	8.3191	9.3214	9.529

Table 4: *Average Speed Improvement Rate (SIR).*

Average SIR in %	Garden	Clare	Football	Tennis
QPSO Over NTSS	67.29	62.63	57.91	56.88
QPSO Over DS	38.36	34.20	26.44	38.07
QPSO Over ARPS	31.98	30.55	26.96	41.71
QPSO Over PSO	15.21	13.57	14.22	13.78

In the proposed scheme, i.e. QPSO based search algorithm, the state of the particles are described with the help of only position vector instead of position and velocity. Furthermore, from the QPSO algorithm, we can see that, there is only one control parameter which, makes the realization simpler than that of the PSO algorithm. QPSO is performing better because the number of computations are reduced by maintaining the same or improving the quality of the video data. This is mainly due to the mutation operator that provides diversity in the search space and thus increases the global search capability.

This technique provides perfect motion estimation with less computational complexity. This also provides high accuracy as compared to full search and diamond search.

Different search technique

FS: The full search algorithm is very easy to implement, and it give the most accurate result. This is due to the calculation of all possible displacements within the search range using block distortion measure. So, no specific algorithm is necessary, it is merely a 2-D search.

TSS: As three steps are used to find the best matched macroblock within the search window of the reference frame, so the name is three step search. The steps size of the search window is initially set to half of the search area. Nine points including the centre point and eight checking points on the boundary of the search window are selected at each step. Next, the search centre moves forward to the matching point with the minimum SAD of the first step and the step size of the second step is reduced by a factor of two. The stopping criteria is when the step size comes down to a single pixel, and the optimum motion vector with the minimum SAD is obtained.

4SS: The four step search algorithm uses nine points for comparison, and then the points are selected based on the following algorithm: the search step starts with a size of 2. MAE is calculated for all the points, and the point with minimum MAE detected. Now the centre is moved to the detected point. This is carried out till it reaches to the boundary and the step size is reduced to a single pixel point.

ARPS: A rood-shaped search pattern is used, in which the size of the rood arms have a provision for adjustment adaptively while searching. The motion vector of the immediate left neighbouring block is used to predict the motion vector for the current block.

DS: Diamond Search uses two search patterns a large diamond for general-purpose gradient search and a smaller diamond for final stage improvement. The approach is very much again similar to the 2-D Logarithmic search in that the large diamond pattern at stage k is cantered on the point with minimum BDM from stage k-1. When the search remains at the centre of the pattern, the small diamond pattern is invoked to refine the motion vector before termination.

HEXBS: The HEXBS may find motion vector in motion field with fewer search points than the diamond search algorithm.

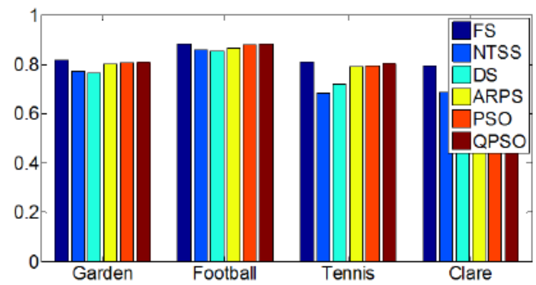


Fig.8: *SSIM plot for all the four video sequence.*

Larger is the motion vector the more significant is the speedup gain for this method. The simulation results verifies the superiority of the proposed method to that of the other fast search methods by utilizing fewer number of search points at the cost of minimum degradation in distortion. HEXBS also uses fewer search points each time thereby saves the cost and hardware size.

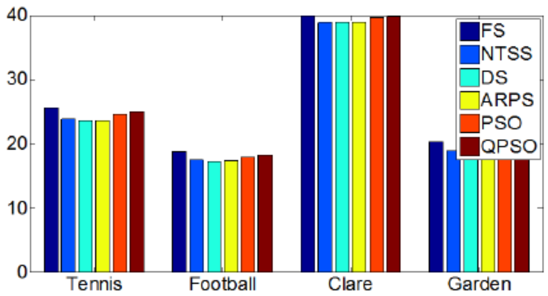


Fig.9: *Average PSNR plot for all the four video sequence.*

5.3 Comparative analysis

The average speed improvement rate (SIR) is summarized in Table 4 for different algorithms. The SIR in SIR % is given by

$$SIR = \left(\frac{N_2 - N_I}{N_1}\right) \times 100\% \tag{8}$$

where N2 is the number of search points used in method 2, while N1 is the number of search points used in method-1. The results in Table 4 demonstrate that the proposed method can reduce the number of search points from 13% to 67% compared to other block matching algorithms. Fig. 10, 11, 12, 13 display the plots for average PSNR for the all four data sets on frame by frame basis. The SSIM plot is depicted in Fig. 8. This plot shows that the reconstruction quality of the frames for QPSO is better as compared to the other state of the art techniques including PSO.

Table 5: *SSIM comparison of each video sequence.*

Methods for ME	Garden	Football	Tennis	Clare
FS	0.8178	0.8835	0.811	0.7965
NTSS	0.7707	0.8606	0.6838	0.6875
DS	0.7664	0.8533	0.7195	0.7264
ARPS	0.802	0.865	0.7912	0.7632
PSO	0.8095	0.8793	0.7964	0.7813
Proposed	**0.8102**	**0.8825**	**0.8034**	**0.7892**

Fig. 9 shows the bar plot for the average PSNR for all the four data sets. This plot shows that the average PSNR value for the proposed method is higher as compared to the FS, DS, ARPS and the PSO based search method. Table 5 shows the SSIM comparison, where the proposed method is having good reconstruction as compared to the other methods.. It is obvious that the SSIM index for FS will be higher, this is because all blocks are involved in the block matching process as compared to the other but the proposed method is having high SSIM index from all the other methods except full search.

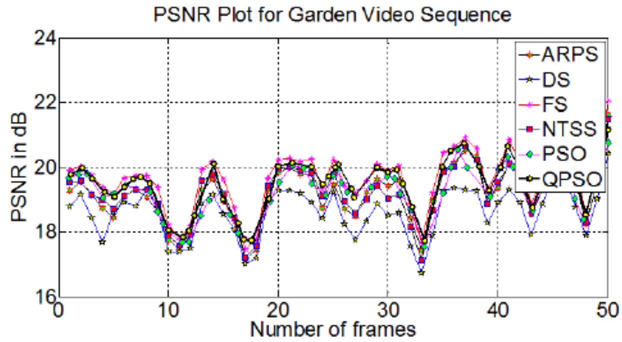


Fig.10: *PSNR plot for Garden sequence.*

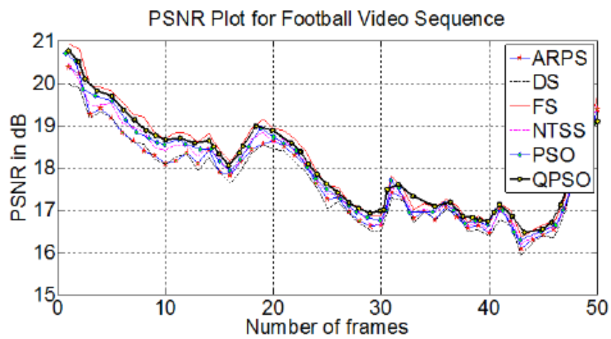


Fig.11: *PSNR plot for Football sequence.*

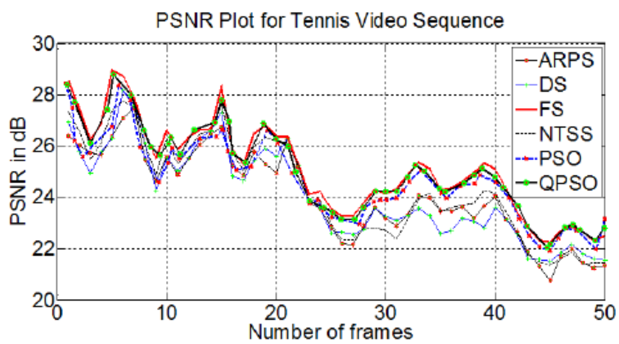


Fig.12: *PSNR plot for Tennis sequence.*

5.4 Complexity Analysis

The computation complexity depends on the time complexity and the number of operations being carried out. The computation time is discussed in Table 6 and the computational complexity is given in Table 7. From Table 6, it can be seen that the proposed method takes less time for the computation of motion vector for all four datasets. The complexity of

Table 6: *Computation time (second) of various motion estimation algorithms.*

Video Data set	FS	NTSS	DS	ARPS	PSO	**Proposed**
Football	3081.26	111.49	64.64	61.96	56.39	**38.17**
Tennis	3099.05	110.45	53.41	46.5	49.75	**32.41**
Garden	3062.48	106.27	49.09	40.9	44.87	**28.76**
Clare	3057.64	104.47	48.41	39.58	42.65	**26.73**

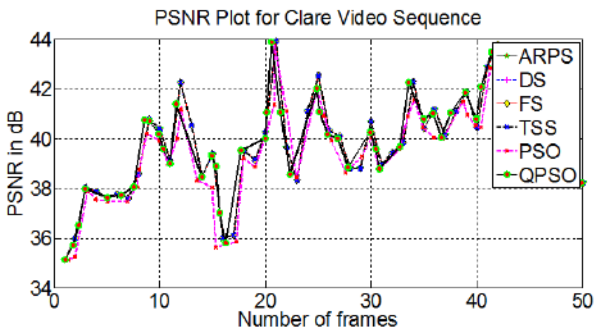


Fig.13: *PSNR plot for Clare sequence.*

the proposed method is only 1+ p log p operations which is very less than the other state of art methods.

Table 7: *Computational complexity of various search methods for motion estimation.*

Search Method	Complexity
ES	$2*(2p+1)^2$
TSS	[1+8log2(p+1)]
NTSS	[1+8log2(p+1)]+8
4SS	[18 log2(p+1)/4+9]
ARPS	2*(2*(p+1))
PSO	p^2
Proposed	1+ p log p

6. CONCLUSIONS

A novel pattern based motion estimation approach, utilizing QPSO has been proposed in this paper to provide a lower computational burden with enhanced accuracy. In this scheme a pattern is predefined (HEXBS) initially followed by evolutionary based search techniques using QPSO algorithm. QPSO is being used here to overcome the problem of local optima as well as faster computation. The algorithm evades the problem faced by the conventional PSO. The validation of the proposed algorithm has been tested with various experiments and conclusions were inferred. From the experimental results it can be validated that the proposed approach out performs the existing conventional methods in terms of PSNR, SSIM, SIR %, and average number of search points etc. Various graphical representations are made for the justification of the proposed method to be superior as compared to the other reference techniques.

This work can be further extended by going beyond this QPSO and using some other variants of PSO, namely CLPSO, DMS-PSO FIPS, ELPSO, CPPSO, and GOPSO with an aim to achieve better block matching with low computational complexity.

References

[1] J. Cai and W. D. Pan, “On fast and accurate block based motion estimation algorithms using particle swarm optimization,” *Information Sciences*, vol. 197, pp. 53-64, 2012.

[2] Jalloul, M.K., and M. A. Al-Alaoui. “A novel Cooperative Motion Estimation Algorithm based on Particle Swarm Optimization and its multicore implementation,” *Signal Processing: Image Communication*, 39, pp. 121-140, 2015.

[3] S. Immanuel Alex Pandian, G. Josemin Bala and J. Anitha, “A pattern based PSO approach for block matching in motion estimation,” *Engineering Applications of Artificial Intelligence*, vol.26, Issue 8, pp. 1811-1817, 2013.

[4] Z. Ping, W. Ping and Y. Hongyang, “A novel block matching algorithm based on particle swarm optimization with mutation operator and simplex method,” *WSEAS Transactions on Systems and Control*, vol.6, Issue 6, pp.207-216, 2011.

[5] R. Li, B. Zeng and M. L. Liou, “A New Three step Search Algorithm for Block Motion Estimation”, *IEEE Transactions on Circuits and Systems for Video Technology*, vol.4, no.4, pp. 438-442, Aug. 1994.

[6] J. Lu and M. L. Liou, “A simple and efficient search algorithm for block-matching motion estimation,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol.7, no.2, pp. 429-433, Apr. 1997.

[7] S. Zhu and K.K. Ma, “A New Diamond Search

Algorithm for Fast Block-Matching Motion Estimation," *IEEE Transactions on Image Processing*, vol.9, no.2, pp. 287-290, Feb. 2000.

[8] C.H. Cheung and L.M. Po, "A novel cross-diamond search algorithm for fast block motion estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.12, no.12, pp. 1168-1177, Dec. 2002.

[9] Y. Nie and K.K. Ma, "Adaptive rood pattern search for fast block-matching motion estimation," *IEEE Transactions on Image Processing*, vol.11, no.12, pp. 1442-1449, 2002.

[10] C. Zhu, X. Lin and L.P. Chau, "Hexagon-based search pattern for fast block motion estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.12, no.5, pp. 349-355, May 2002.

[11] E. Cuevas, D. Zaldvar, M. P. Cisneros, H. Sossa and V. Osuna, "Block matching algorithm for motion estimation based on Artificial Bee Colony (ABC)," *Applied Soft Computing*, vol.13, Issue 6, pp. 3047-3059, 2013.

[12] Du Gy, Huang Ts, Song Lx, Zhao Bj, "A Novel Fast Motion Estimation Method Based on Particle Swarm Optimization," In proceedings of Fourth International Conference on Machine Learning and Cybernetics (August) , pp. 5038-5042, 2005.

[13] Bakwad KM, Pattnaik S, et al. "Small Population Based Modified Parallel Particle Swarm Optimization for Motion Estimation," *In proceedings of 16th International Conference on Advanced Computing and Communications*, pp. 367-373, 2008.

[14] Bakwad KM, Pattnaik SS, Sohi BS, Devi S, Gollapudi SV, Sagar CV, Patra PK, "Fast motion estimation using small population-based modified parallel particle swarm optimisation," *International Journal of Parallel, Emergent and Distributed Systems*, 26(6) , pp. 457-476, 2011.

[15] Huang YW, Chen CY, Tsai CH, Shen CF, Chen LG, "Survey on Block Matching Motion Estimation Algorithms and Architectures with New Results," *Journal of VLSI signal processing systems for signal, image and video technology*, vol.42, Issuse 3, pp. 297-320, 2006.

[16] D. Tzovaras, I. Kompatsiaris and M. G. Strintzis, "3D object articulation and motion estimation in model based stereoscopic videoconference image sequence analysis and coding," *Signal Processing: Image Communication*, vol.14, Issue 10, pp. 817-840, 1999.

[17] Barron JL, Fleet DJ, Beauchemin SS, "Performance of optical ow techniques," *International Journal of Computer Vision*, 12(1), pp. 43-77, 1994.

[18] Skowronski J, "Pel recursive motion estimation and compensation in sub bands," Signal Processing: Image Communication, 14(5), pp. 389-396, 1999.

[19] J. Jain and A. Jain, "Displacement Measurement and Its Application in Inter frame Image Coding," *IEEE transactions on Communications*, vol.29, no.12, pp. 1799-1808, Dec. 1981.

[20] Zhou J A and Love, P E D and Wang, X and Teo, K L and Irani, Z, "A review of methods and Algorithms for optimizing construction scheduling," Journal of the Operational Research Society, 64, pp. 1091-1105, 2013.

[21] Sun J, Lai CH, Wu XJ, "Particle swarm optimisation: classical and quantum perspectives," *CRC Press*, 2011.

[22] Eberhart, Russell C and Shi, Yuhui, "Particle Swarm Optimization: Developments, Applications and Resources," *Proceedings of the 2001 Congress on Evolutionary Computation (IEEE)*, pp. 81-86, 1995.

[23] Sun, J., Xu W, & Feng, B. "A global search strategy of quantum-behaved particle swarm optimization," *Proceedings of IEEE Conference on Cybernetics and Intelligent Systems*, vol. 1, pp. 111-116, 2004.

[24] Sun J, Fang W, Wu XJ, Palade V, Xu WB, "Quantum-Behaved Particle Swarm Optimization: Analysis of Individual Particle Behaviour and Parameter Selection." *Evolutionary Computation*, 20(3), pp. 349-393, 2012.

[25] Y. Ismail, J. B. McNeely, M. Shaaban, H. Mahmoud and M. A. Bayoumi, "Fast motion estimation system using dynamic models for H.264/AVC video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.22, no.1 , pp. 28-42, Jan. 2012.

[26] Y.Cho CIPR Data set for YUV sequence http://www.cipr.rpi.edu/resource/sequences/sif.html, 2005.

[27] Karthik, R., and R. Menaka, "Statistical characterization of ischemic stroke lesions from MRI using discrete wavelet transformations," *ECTI Transactions on Electrical Engineering, Electronics, and Communications*, 14.2, pp. 57-64, 2016.

[28] Salhi Meriem,et al. "Mobility-Assisted and QoS-Aware Resource Allocation for Video Streaming over LTE Femtocell Networks," *ECTI Transactions on Electrical Engineering, Electronics, and Communications 13.1*, pp. 42-53, 2015.

[29] Smita Pradhan and Dipti Patra, "RMI based non-rigid image registration using BF-QPSO optimization and P-spline," *AEU -International Journal of Electronics and Communications*, 69(3):609-621, 2015.

Yogananda Patnaik has received his M.tech. from Electrical Engineering Department, National Institute of Technology, Rourkela India in 2006. Currently he is pursuing Ph.D. at National Institute of Technology, Rourkela. His areas of interest are Signal / Image Processing / Video Processing, Computer vision.

Dipti Patra has received Ph.D. from National Institute of Technology, Rourkela India in 2006. Currently, she is working as Associate Professor in the Department of Electrical Engg., N.I.T., Rourkela, India. Her areas of interest are Signal / Image / Video Processing, Computer Vision and Visual Servo Control. She is a Senior Member of IEEE, Fellow of IETE and IEI, India.