(IJNCAA)

ISSN 2220-9085(ONLINE)
ISSN 2412-3587(PRINT)

INTERNATIONAL JOURNAL OF

New Computer Architectures and their Applications

Volume 8, Issue 4,
2018

SDIWC
www.sdiwc.net

International Journal of
NEW COMPUTER ARCHITECTURES AND THEIR APPLICATIONS

Overview

The SDIWC International Journal of New Computer Architectures and Their Applications (IJNCAA) is a refereed online journal designed to address the following topics: new computer architectures, digital resources, and mobile devices, including cell phones. In our opinion, cell phones in their current state are really computers, and the gap between these devices and the capabilities of the computers will soon disappear. Original unpublished manuscripts are solicited in the areas such as computer architectures, parallel and distributed systems, microprocessors and microsystems, storage management, communications management, reliability, and VLSI.

One of the most important aims of this journal is to increase the usage and impact of knowledge as well as increasing the visibility and ease of use of scientific materials, IJNCAA does NOT CHARGE authors for any publication fee for online publishing of their materials in the journal and does NOT CHARGE readers or their institutions for accessing the published materials.

Publisher

The Society of Digital Information and Wireless Communications
20/F, Tower 5, China Hong Kong City, 33 Canton Road, Tsim Sha Tsui, Kowloon, Hong Kong

Further Information

Website: http://sdiwc.net/ijncaa, Email: ijncaa@sdiwc.net,
Tel.: (202)-657-4603 - Inside USA; 001(202)-657-4603 - Outside USA.

The issue date is Dec. 2018.

IJNCAA
ISSN 2220-9085 (Online)
ISSN 2412-3587 (Print)
Volume 8, Issue No. 4
2018

CONTENTS

ORIGINAL ARTICLES

International Journal of
NEW COMPUTER ARCHITECTURES AND THEIR APPLICATIONS

The *International Journal of New Computer Architectures and Their Applications* aims to provide a forum for scientists, engineers, and practitioners to present their latest research results, ideas, developments and applications in the field of computer architectures, information technology, and mobile technologies. The IJNCAA is published four times a year and accepts three types of papers as follows:

1. **Research papers**: that are presenting and discussing the latest, and the most profound research results in the scope of IJNCAA. Papers should describe new contributions in the scope of IJNCAA and support claims of novelty with citations to the relevant literature.
2. **Technical papers**: that are establishing meaningful forum between practitioners and researchers with useful solutions in various fields of digital security and forensics. It includes all kinds of practical applications, which covers principles, projects, missions, techniques, tools, methods, processes etc.
3. **Review papers**: that are critically analyzing past and current research trends in the field.

Manuscripts submitted to IJNCAA **should not be previously published or be under review** by any other publication. Plagiarism is a serious academic offense and will not be tolerated in any sort! Any case of plagiarism would lead to life-time abundance of all authors for publishing in any of our journals or conferences.

Original unpublished manuscripts are solicited in the following areas including but not limited to:

- Computer Architectures
- Parallel and Distributed Systems
- Storage Management
- Microprocessors and Microsystems
- Communications Management
- Reliability
- VLSI

International Journal of New Computer Architectures and their Applications (IJNCAA) 8(4): 168-178
The Society of Digital Information and Wireless Communications, 2018 ISSN 2220-9085 (Online); ISSN 2412-3587 (Print)

Parallel *k*-means Clustering Algorithm on SMP

Athari M. Alrajhi and Soha S. Zaghloul, PhD
College of Computer & Information Sciences, Department of Computer Science, King Saud University, Riyadh
437203971@student.ksu.edu.sa
smekki@ksu.edu.sa

ABSTRACT

The *k*-means clustering algorithm is one of the popular and simplest clustering algorithms. Due to its simplicity, it is widely used in many applications. Although *k*-means has low computational time and space complexity, increasing the dataset size results in increasing the computational time proportionally. One of the most prominent solutions to deal with this problem is the parallel processing. In this paper, we aim to design and implement a parallel *k*-means clustering algorithm on shared memory multiprocessors using parallel java library. The performance of the parallel algorithm is evaluated in terms of speedup, efficiency and scalability. Accuracy and quality of clustering results are also measured. Furthermore, this paper presents analytical results for the parallel program performance metrics.

KEYWORDS

k-means, Clustering, SMP, Parallel Java, Parallel Programming, pj2, Shared Memory Multiprocessors.

1 INTRODUCTION

Data mining clustering techniques are unsupervised learning because they don't use predefined class labels. The clustering goal is to obtain meaningful groupings of objects based on a measure of similarity such that all objects in one group are similar to each other and different from the objects in other groups. Cluster analysis has been widely used in data recovery, web and text mining, image segmentation and pattern recognition. Therefore, several clustering algorithms have been developed. *k*-means is one of the popular partial clustering algorithms [1]. The idea of *k*-means is based on dividing datasets into k number of groups (clusters) such that the squared error between the mean of a cluster and the data points in the cluster is minimized. The mean of a cluster is called *centroid*. The initial centroids are chosen randomly one for each cluster. Then, each point or object belongs to the cluster which has the nearest centroid by computing the Euclidian distance between the point and each centroid. These centroids are updated based on means of each cluster which assign as a new centroid. The assignments and updates are repeated until each centroid remains the same (convergence criterion) [2].

Although *k*-means is capable of dividing the problem domain into smaller parts, it suffers an increase in computation time as the size of the dataset becomes very large. Therefore, an additional technique like parallel processing, to accelerate the computation process is required. Parallel programming can divide the program tasks into smaller independent parts with the aim of running them on multiple processors simultaneously [3]. So, finding those independent parts to reduce the computational time is a challenging issue.

In this paper, we aim to study the parallel *k*-means algorithm and examine its performance on one of the parallel computer architectures called shared memory multiprocessor (SMP).

The rest of this paper is structured as follows; Section 2 discusses the most related work to the problem in-hand. In section 3, a detailed design of sequential and parallel *k*-means clustering algorithm is described. Section 4 presents the results of implementing both sequential and parallel versions. The analytical results are

introduced in section 5. Finally, section 6 contains the conclusion and potential future work.

2 RELATED WORK

Parallel processing of *k*-means clustering algorithm has been able to attract the attention of many researchers around the world. They used different parallel programming models and various techniques in order to achieve a high performance and less computational time. One of the recent studies on the *k*-means algorithm is presented by Kucukyilmaz [4]. In this study, a parallel *k*-means algorithm is implemented on shared memory multiprocessors with 8 cores. Extensive experiments are conducted with varying number of instances, clusters and attributes to illustrate the impact of them on the performance. The results show that the previous parameters hold almost equal importance. These results are obtained by comparing the theoretical results with experimental results. Although this work shows a detailed implementation of the algorithm and a good analysis of the results, no evaluation metrics for the parallel program are used.

In another study, message passing interface is used for parallelizing *k*-means on distributed memory paradigm in [5]. In this work, Kantabutra and Couch proposed a technique to improve the performance in terms of time complexity. Using the evaluation measures, the experimental results show that their technique achieves 50% efficiency of time complexity. In the context of message passing, Ramesh et al. [6] implemented parallel *k*-means for cluster large agricultural dataset. Using a varying number of data size and clusters, the results prove that the parallel algorithm achieves more efficiency and time complexity than the sequential algorithm.

In [7], Farivar et al. proposed an algorithm to implement *k*-means clustering on an NVIDIA GPU using CUDA. The dataset consists of 1 million instances, and the number of clusters is 4000. For an objective comparison, different platforms are used, and consequently, different speed improvement is achieved. The results suggest that the speed performance is increased up to 13x and 68x for each platform compared to the PC implementation. CUDA architecture is also used in [8]. In this work, Wu and Hong presented an efficient CUDA-based *k*-means with load balancing using the triangle inequality. Through extensive experiments, the algorithm achieves better efficiency as compared to CPU-based *k*-means algorithms. As a result, improved performance in terms of speed and scalability is achieved. In the same way, Kumar et al. [8] used the triangle inequality to decrease the unnecessary distance calculations. In addition, they solve the problem of load imbalance which is related to their framework when these computations are avoided.

3 DESIGN AND IMPLEMENTATION

This section represents the heart of this paper where the aspects of *k*-means algorithm design are discussed. Furthermore, the inputs and outputs of the algorithm are illustrated in section 3.1. Detailed design steps are presented in section 3.2. Both sequential and parallel versions of the *k*-means algorithm are designed and discussed in sections 3.3 and 3.4 respectively.

3.1 Inputs and Outputs

The inputs of the *k*-means algorithm are:
- Dataset of n 2-dimensional data points.
- *k* value which indicates the number of clusters.

The output of the *k*-means algorithm is:
- *k* clusters, each one includes a set of points.

3.2 *k*-Means Design

As mentioned in section 1, *k*-means is one of the popular and simplest clustering algorithms that partitions a dataset into *k* groups by minimizing the sum of squared error (SSE) between the mean of a cluster and the data points in the cluster. The algorithm starts with *k* initial

centroids and works iteratively to assign each point to one of the k clusters based on feature similarity until a convergence criterion is met. More formally, given a set of n d-dimensional data points $X = \{x_i, 0 < i < n\}$, a set of k initial centroids $C = \{c_j, 0 < j < k\}$, and a mean of each cluster μ_j, our goal is to minimize SSE as follows [9]:

$$SSE = \sum_{j=1}^{k} \sum_{x_i \in C_j} \|x_i - \mu_j\|^2 \qquad (3.1)$$

The algorithm of k-means is described in the following four steps:

1. Initialization:

This step involves selecting k initial centroids $C = \{c_j, 0 < j < k\}$ from the instance space, where k is the number of clusters [10]. There are many methods proposed for selecting initial centroids. One common way is to randomly either choose k actual data points from the dataset or generate k virtual data points. The actual data point is a point that comes directly from the dataset. In contrast to the actual data point, a virtual data point is a point that not related to any point in the actual dataset [11].

2. Distance Calculation:

This step includes calculation of finding the closest centroid for each data point and computing the distance to it. There are many distance metrics to measure the distance between centroids and data points such as Manhattan, Euclidean distance, cosine similarity, correlation, etc. Euclidean distance is often used as a measure of distance for k-means clustering [12]. The distance between x_i and c_j is given by:

$$d(c_j, x_i) = \sqrt{\sum_{t=1}^{d} (c_{j,t} - x_{i,t})^2} \qquad (3.2)$$

3. Centroid Recalculation:

After assigning each point x_i to the closest cluster c_j, the centroids are re-calculated by compute the average of all points within the cluster as follows [1]:

$$\mu_j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i \qquad (3.3)$$

4. Convergence:

The clusters obtained after the previous steps are actually not optimized. In order to find a minimal SSE, steps 2 and 3 must be repeated until the results become stable. The stability condition is called convergence criterion and can be specified in multiple ways such as the convergence criterion is met after a fixed number of iterations or when centroids remain the same [10].

It is worth noting that the time complexity of the k-means algorithm is $O(n^* k^* i^* d)$, where:
n: number of data points (instances) in the dataset.
k: number of clusters.
i: number of iterations.
d: number of dimensions.

3.3 Sequential k-Means

In this study, sequential model of the k-means clustering algorithm is designed with the aim of calculating the speedup gains of parallel implementation which express the impact of parallelization. As mentioned in section 3.2, the first step of k-means algorithm is selecting k initial centroids randomly. Because the quality of the clustering results highly depends on the quality of this selection, choosing good initial centroids can play an important role in obtaining better results as well as reducing the computational complexity of the algorithm [10]. The results in [11] show that the virtual points perform better than the actual points. Therefore, in this research, we will generate k virtual points

randomly as initial centroids. To guarantee that the generated virtual points values don't exceed the values range of the points in the dataset and consequently get good results, the random virtual points should be scaled using any normalization technique. The sequential algorithm is explained in figure 3.1.

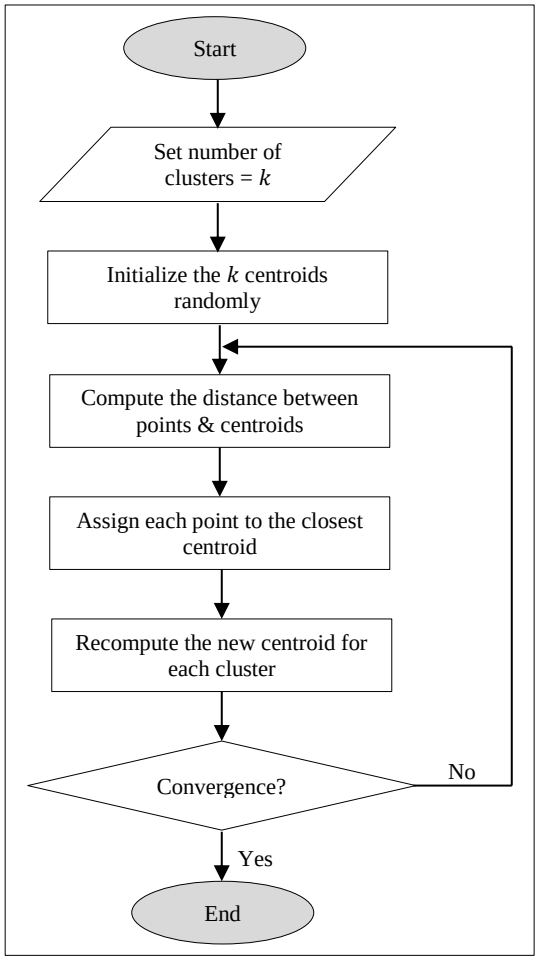


Figure 3.1 Sequential *k*-means Algorithm

3.4 Parallel *k*-Means

Designing a parallel model of the *k*-means on SMP is a big challenge because *k*-means is inherently sequential. Furthermore, the challenge lies not only in the design of a parallel algorithm, but the parallel algorithm must be superior to the serial algorithm in terms of execution time reduction which is the main objective of parallelism. Thus, we have to look for the independent parts of the algorithm that takes a long time of execution, and then execute them in parallel. These parts are often related to computations. When we look at the four steps of *k*-means, we find that the first step (initialization) cannot be parallelized, because it is too simple, and each centroid must be initialized globally. The computational bottleneck of the algorithm is the second step, where the distance between each point and centroid is computed, especially if there is a large number of points. This step can be parallelized by dividing the data points among processors and then, making each thread represents a point. Thus, each point is assigned to one thread to compute the nearest centroid for each point in a parallel manner. After that, each thread stores its result (closest centroid) in its own per-thread variable. At the end of this step, a reduction parallel pattern is used to collect the all threads results according to the closest centroid. The per-thread variables have to be reduced together into one overall variable to be ready to the next step. Figure 3.2 explains this idea. In the third step, new centroids for each cluster are re-computed which can be implemented in parallel, since each thread will represent a centroid. For the last step, the loop cannot be parallelized since an iteration relies on the results of the previous one. Figure 3.3 illustrates the synchronous parallel *k*-means algorithm on SMP. Figure 3.4 shows a flowchart of the parallel *k*-means algorithm.

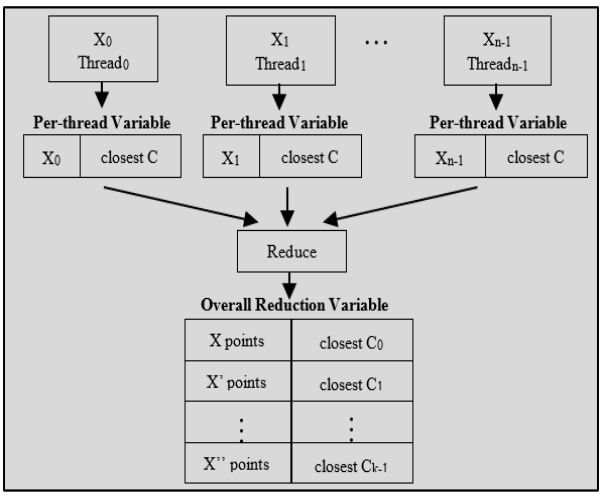


Figure 3.2 Parallel Reduction

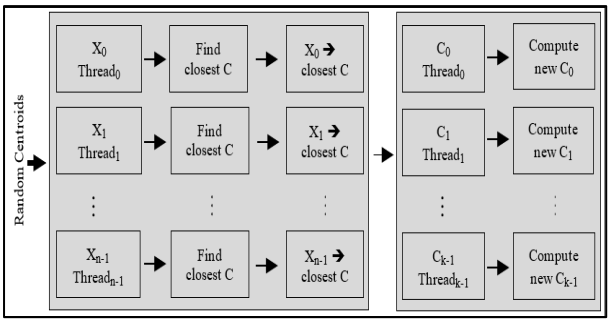


Figure 3.3 Synchronous Parallel *k*-means on SMP

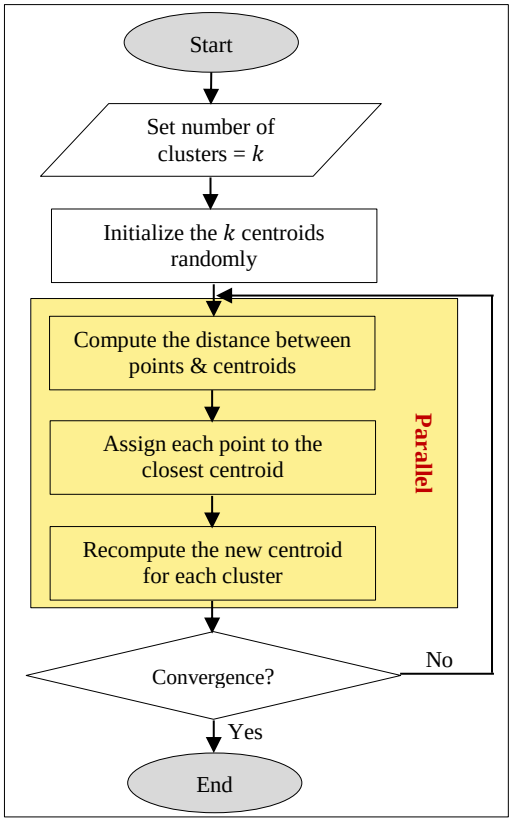


Figure 3.4 Parallel *k*-means Algorithm

4 EXPERIMENTS AND RESULTS

To study the performance of the parallel *k*-means on SMP, some experiments are conducted on both parallel and sequential versions of the algorithm.

In section 4.1, the software and hardware specifications are illustrated. The main performance measurements to evaluate our work are described in section 4.2. Further, the detailed experiments and results are explained and presented numerically and graphically in section 4.3.

4.1 Hardware and Software Environment

Sequential and parallel versions of the *k*-means algorithm are implemented on Intel core i7 with quad-cores running at 2.30 GHz and supports hyperthreading. So, the system operates like it has 8 cores because each core can handle 2 threads. The operating system is Linux Ubuntu 16, and the programming language is Java using NetBeans IDE. Parallel Java 2 (pj2) library is used for threads management. Further, it supports shared memory parallel programming on multicore computers [13].

4.2 Evaluation Measurements

In this section, evaluation metrics are presented and explained to evaluate our work. These measures are divided into two categories. The first category includes the measures that evaluate the quality and the accuracy of *k*-means. The second category involves the measures that related to the performance of parallel programs. The following sections describe these measures.

4.2.1 Quality and Accuracy of Clustering Results

To evaluate the quality of *k*-means clustering results, Sum of Squared Error measure is used to handle this issue. Furthermore, the accuracy of clustering is computed as follows:

1. Sum of Squared Error:

When centroids are initialized randomly, different runs of *k*-means lead to different results in total SSEs because *k*-means algorithm only converges to the local minimum. As mentioned earlier, the quality of the clustering results relies on chosen of the centroids. So, choosing poor initial centroids may cause poor clustering results with higher SSE. One way to address this problem is to perform multiple runs with multiple different initial centroids and choose the one that gives the smallest squared error [9], [1]. Sum of squared error is a common measure to evaluate *k*-means. This evaluation defines a

good measure for the homogeneity of the clustering results. SSE is given by:

$$SSE = \sum_{j=1}^{k} \sum_{x_i \in C_j} \left\| x_i - \mu_j \right\|^2 \qquad (4.1)$$

2. Accuracy of Clustering:

In order to compute the accuracy of points clustering, we have to check that all points used are included in the clusters. Simply, we will sum the number of points in all clusters and compare it with the number of points in the dataset. If the sum is equal to the number of points, then all the points have been clustered. Otherwise, some points are un-clustered (k-means is a crisp (hard) clustering algorithm, it assigns each data point to one cluster exclusively. So, there is no choice for some points to be duplicated). Figure 4.1 shows a flowchart of checking steps. The accuracy of clustering is computed by dividing the number of clustered points by the total number of points and multiplying the answer by 100. The accuracy of clustering is given by the following formula:

$$Acc_{clst} = \frac{cp}{tp} \times 100 \qquad (4.2)$$

Where:

cp: the number of clustered points.

tp: the total number of points.

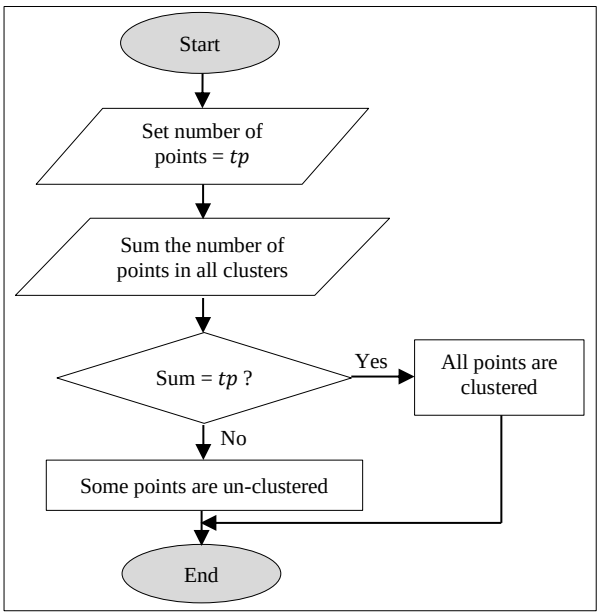


Figure 4.1 Checking the Clustering of Points

4.2.2 Performance Metrics of Parallel Program

It is important to study the parallel programs performance with the aim of determining the best algorithm, evaluating the efficiency of a parallel algorithm, examining the benefits from parallelism, and evaluating parallel hardware platforms. Some main metrics are used to analyze the parallel programs performance as below:

1. Speedup:

To see the benefit of parallelism, it is important to know how much speed gain is achieved by parallelizing a program over a sequential implementation. Therefore, a comparison with the running time of a sequential version of a given application is very important to analyze the parallel version. Speedup can be defined as the ratio of the execution time of the sequential version of a given program running on one processor to the execution time of the parallel version running on P processors with problem size n [14]. Speedup is given by the following formula [15]:

$$Speedup = \frac{S_{seq}(n,1)}{S_{par}(n,p)} \qquad (4.3)$$

Where:

S_{seq}: speed (running time) of the sequential version.

S_{par}: speed (running time) of the parallel version.

p: number of processors.

n: problem size which denotes the number of computations that the program has to perform. As the problem size increases, the speedup increases.

In fact, as the number of processors increases, the speedup increases. However, there is upper bound on speedup due to that there is no program can be completely parallel. Thus, there is a time allocated for the sequential fraction of the program. If f is the sequential fraction of the

program, then its speedup is bounded by $\frac{1}{f}$ regardless of the number of processors. This is known as Amdahl's Law.

In this research, serial and parallel models of the k-means clustering algorithm are designed and implemented in order to calculate the speedup gains of parallel implementation which express the impact of parallelization.

2. Efficiency:

An alternative performance measure of a parallel application is the efficiency. It captures how a program's speedup is close to ideal. In other words, efficiency measures the effectiveness of processors utilization of the parallel program [15]. It can be defined as the ratio of actual speedup to the number of processors [14] and expressed as [15]:

$$Efficiency = \frac{Speedup(n,p)}{p} \qquad (4.4)$$

Where:
n: problem size.
p: number of processors.

In an ideal parallel program, efficiency is equal to one and speedup is equal to p. However, in practice, it cannot achieve the ideal behavior because while running a parallel program, the processors cannot spend 100% of their time on program's computations. Thus, a real parallel program has an efficiency value between zero and one and speedup less than p [16].

3. Scalability:

The scalability of a parallel program is a measure to describe how the program performance changes as the number of processors is increased. As mentioned earlier, a speedup saturation can be observed when the problem size n is fixed, and the number of processors p is increased. However, the attained speedup increases when the problem size n increases for a fixed number of processors p [14]. In this sense, a parallel program is scalable if its performance improve continues as both problem size n and number of processors p are increased. Scalability is given by the following formula [17]:

$$Scalability = \frac{n_{par}}{n_{5eq}} \qquad (4.5)$$

Where:
n_{5eq}: the largest problem size that the sequential program can be handled.
n_{par}: the largest problem size that the parallel program can be handled for a specific number of processors.

4.3 Experimental Results

Extensive experiments on both sequential and parallel models of k-means clustering algorithm are conducted. For a fair comparison between the two versions, the convergence criterion is met after a fixed number of iterations. We used a different number of points and a different number of cores to illustrate the impact of both on running time. Generally, the dataset consists of more than 100,000 points with 2 dimensions. Clustering benchmark dataset called Birch is used [18]. The additional parameters are illustrated in table 4.1.

Table 4.1 Parameters of Experiments

Parameter	Value
No. of Points	40,000 / 80,000 / 120,000 / 160,000
No. of Clusters	20
No. of Iterations	500
No. of Cores / Threads	2 / 4 / 8

1. Accuracy of Clustering and SSE:

For all experiments that aim to evaluate the performance of parallel k-means, the accuracy of clustering and SSE values have been measured. As shown in table 4.2, all experiments result in an accuracy of 100%, implying the efficiency of k-means clustering algorithm despite its simplicity. In order to measure the homogeneity of the clustering results, the program is run

multiple times with multiple different initial centroids and choose the results that give the smallest squared error as well as give the smallest running time as illustrated in the next section. The detailed results of accuracy and SSE are recorded in table 4.2. The large numbers of SSE do not mean large error values, but the reason is that there is a large number of points and each attribute of a point consists of six digits.

Table 4.2 Accuracy of Clustering and SSE

Points	Measure	Sequential	Parallel		
			P=2	P=4	P=8
40,000	Accuracy	100%	100%	100%	100%
	SSE	14382.9	25322.4	71432.6	12756.8
80,000	Accuracy	100%	100%	100%	100%
	SSE	50697.8	45876.6	31665.2	53777.5
120,000	Accuracy	100%	100%	100%	100%
	SSE	22473.2	29548.3	78532.9	43672.1
160,000	Accuracy	100%	100%	100%	100%
	SSE	55821.7	81325.4	28775.1	92341.5

2. Measuring Running Time:

The running time of a sequential program is the time that the program takes from start to end its execution on a computer. The running time of a parallel program is the time that starts with the beginning of parallel computation and ends when the last processor finishes execution [16]. When a parallel program runs multiple times, it hardly yields the same running time, even with identical inputs. One of the potential reasons is that the background processes and other user programs running on the same computer stealing some CPU time away from the parallel program, causing the running time of the parallel program to increase [15], [17]. To address this issue, the parallel program should be run multiple times (e.g., 10 times) and measure the running time for each run, then the smallest running time value would be the best estimate of the true running time of the parallel program [17].

The first set of experiments are conducted to illustrate the impact of both the problem size and the number of cores on running time. The program is run multiple times with a various number of points and cores. The detailed results are recorded in table 4.3 based on the smallest running time. The results are also represented graphically in figure 4.2. As seen in the graph, as the number of points increases, the running time also increases for all cores. Further, the sequential *k*-means has a higher increasing rate in terms of running time as compared to the parallel *k*-means version for all cores. Also, the increasing number of cores results in decreasing the running time as shown in the graph.

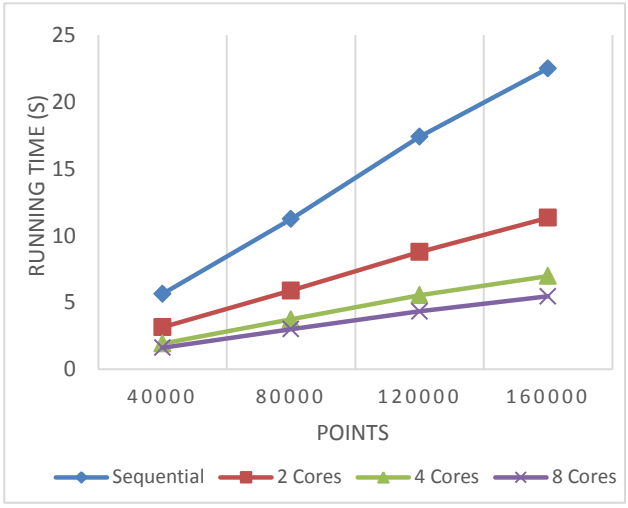


Figure 4.2 The Running Time Against the Number of Cores

3. Speedup Measurement:

In terms of speedup, multiple experiments are conducted to measure the gained speedup of parallelizing the *k*-means algorithm. The speedup is calculated according to formula 4.3 and based on the running time results for a different number of points and a different number of cores. The detailed results are recorded in table 4.3. Furthermore, the speedup results are represented in figure 4.3. As seen in the graph, the increasing number of cores results in increasing the speedup for all number of points. However, the rate of increase begins to

slow as the number of points and cores increases. These results are expected since the scheduling of a large number of threads becomes more complicated. Therefore, the imposed overhead is higher.

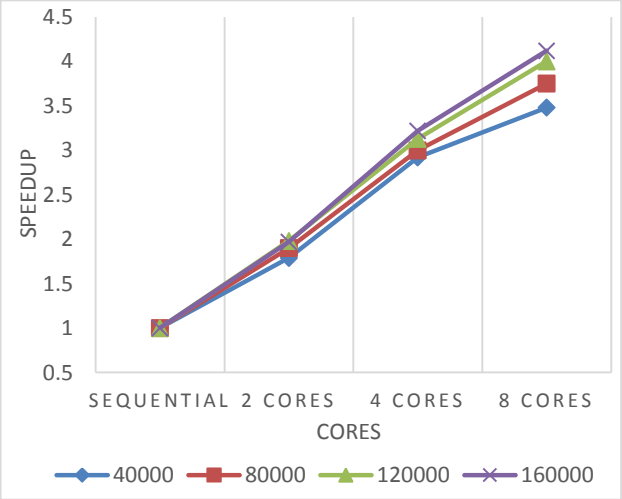


Figure 4.3 The Speedup Against the Number of Cores

4. Efficiency Measurement:

The third set of experiments are conducted with the aim of evaluating the efficiency of using the hardware resources. It is computed according to formula 4.4. The efficiency results shown in table 4.3 are also represented graphically in figure 4.4 for a different number of points and cores. It is noted that as the number of cores increases, the efficiency decreases. For 2 and 4 cores/threads the efficiency is close to 1 because the experiments are conducted on a quad-core, while the efficiency decreases with 8 threads (4 cores with hyperthreading) and may reach to zero with increasing number of cores. These results are expected because the hyperthreading is very expensive in terms of threads scheduling despite its advantages. Therefore, this overhead yield less efficiency.

Although the efficiency of the sequential version is equal to 1, it does not mean that there is a full using of hardware. However, the number of cores and the speedup of the sequential version are both equal to 1.

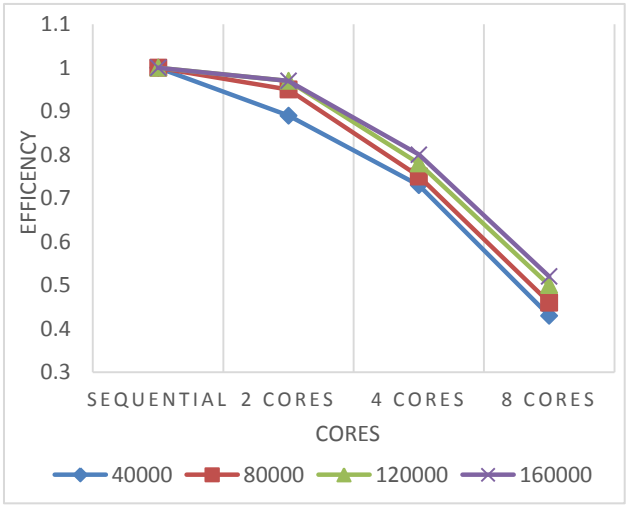


Figure 4.4 The Efficiency Against the Number of Cores

Table 4.3 Running Time, Speedup and Efficiency

Points	Measure	Sequential	Parallel		
			P=2	P=4	P=8
40,000	Time(s)	5.64	3.15	1.93	1.62
	Speedup	1	1.79	2.92	3.48
	Efficiency	1	0.89	0.73	0.43
80,000	Time(s)	11.24	5.89	3.74	3.00
	Speedup	1	1.90	3.00	3.75
	Efficiency	1	0.95	0.75	0.46
120,000	Time(s)	17.39	8.78	5.54	4.34
	Speedup	1	1.98	3.13	4.00
	Efficiency	1	0.97	0.78	0.50
160,000	Time(s)	22.51	11.34	6.98	5.46
	Speedup	1	1.98	3.22	4.12
	Efficiency	1	0.97	0.80	0.52

5. Scalability Measurement:

The scalability is measured according to formula 4.5 by dividing the largest number of points that each core/thread can handle by the largest number of points that the sequential program can handle. The results are recorded in table 4.4. Furthermore, Figure 4.5 shows the relationship between the maximum number of points for each core against the running time graphically. According to the results, the scalability is increased as the number of cores increased also.

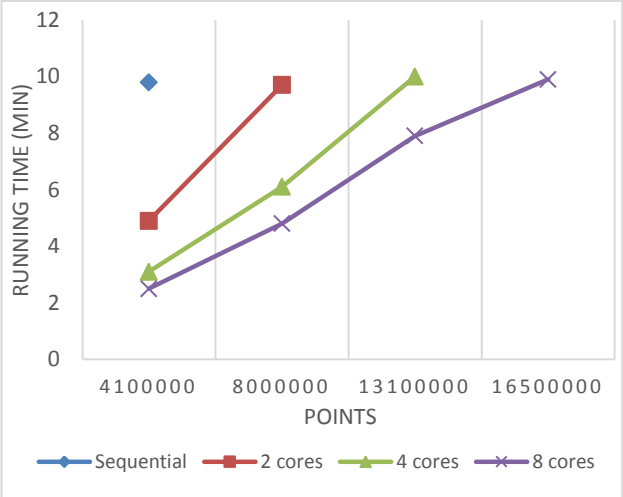


Figure 4.5 The Scalability Against the Running Time

Table 4.4 Scalability of Parallel k-means

Version	Cores	Size up	Scalability
Sequential	1	4100000	-
Parallel	2	8000000	1.9
Parallel	4	13100000	3.2
Parallel	8	16500000	4.0

5 ANALYTICAL RESULTS

For any parallel program, there is a fraction that should be implemented sequentially. In other words, it cannot parallelize the whole program. This is known as Amdahl's Law. The sequential fraction f of the program may consist of:

- Initialization statements.
- Thread creation and synchronization.
- Input and Output.
- Memory deallocation.

Regardless of the number of processors that are available in the parallel architecture, the sequential fraction f uses only one processor. Therefore, the speedup under Amdahl's Law is given by the following formula [15], [17]:

$$Speedup(n,p) \leq \frac{1}{f + \frac{(1-f)}{p}} \qquad (5.1)$$

Where:
$Speedup(n,p)$: the analytical speedup.
f: the sequential fraction.
$1-f$: the parallel fraction.
p: the number of processors.

To get the value of the sequential fraction f, the following formula is used:

$$f = \frac{p * T(n,p) - T(n,1)}{p * T(n,1) - T(n,1)} \qquad (5.2)$$

This formula calculates the sequential fraction f from running time measurements $T(n,p)$ and $T(n,1)$. Thus, Amdahl's Law can calculate the upper bound of the parallel program speedup. The speedup results are recorded in table 5.1. Relatively, the actual and the analytical speedup are almost the same.

Table 5.1 Actual and Analytical Speedup

Points	Measure	Parallel		
		P=2	P=4	P=8
40,000	**Actual Speedup**	1.79	2.92	3.48
	Analytical Speedup	1.79	2.92	3.48
80,000	**Actual Speedup**	1.90	3.00	3.75
	Analytical Speedup	1.91	3.01	3.75
120,000	**Actual Speedup**	1.98	3.13	4.00
	Analytical Speedup	1.98	3.14	4.01
160,000	**Actual Speedup**	1.98	3.22	4.12
	Analytical Speedup	1.98	3.23	4.12

6 CONCLUSION AND FUTURE WORK

This paper targets to accelerate the computation process of k-means clustering algorithm since the time complexity is increased with increasing the problem size. A parallel processing is used as a mechanism to achieve a high performance in terms of running time reduction. Parallel k-means is developed on SMP using java programming language under Linux operating system. We utilize 4 cores with hyperthreading (8 threads) managed by PJ2 library. Accuracy and SSE values are measured to evaluate the quality of clustering results. All results prove the efficiency of k-means with 100% clustering accuracy. In addition, speedup, efficiency and scalability are the metrics used to measure the

performance of the parallel program. The results show that the gained speedup and the scalability increase with increasing the problem size and the number of cores, where the maximum speedup achieved is 4.12 and the maximum scalability reached is 4.0. On the other hand, the efficiency decreases with increasing the number of cores. The analytical results prove that the actual speedup is almost the same as the analytical speedup.

As a future work, we will develop another parallel version of *k*-means to run on cluster parallel programming model. This leads to a valuable comparison between the two models; SMP and clusters.

REFERENCES

[1] E. Karoussi, "Data Mining K-Clustering Problem," 2012.

[2] J. A. Hartigan and M. A. Wong, "A K-Means Clustering Algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979.

[3] T. Rauber and G. Rünger, "Parallel Computer Architecture," in *Parallel Programming for Multicore and Cluster Systems*, 2nd ed., Springer, 2013, pp. 9–104.

[4] T. Kucukyilmaz, "Parallel K-Means Algorithm for Shared Memory Multiprocessors," *Journal of Computer and Communications*, vol. 2, pp. 15–23, 2014.

[5] S. Kantabutra and A. L. Couch, "Parallel K-means Clustering Algorithm on NOWs," *NECTEC Technical journal*, vol. 1, no. 6, pp. 243–248, 2000.

[6] V. Ramesh, K. Ramar, and S. Babu, "Parallel K-Means Algorithm on Agricultural Databases," *International Journal of Computer Sciences*, vol. 10, no. 1, pp. 710–713, 2013.

[7] R. Farivar, D. Rebolledo, and E. Chan, "A parallel implementation of k-means clustering on GPUs," *International Conference on Parallel and Distributed Processing Techniques and Applications*, vol. 13, no. 2, pp. 212–312, 2008.

[8] J. Wu and B. Hong, "An efficient k-means algorithm on CUDA," in *IEEE International Symposium on Parallel and Distributed Processing Workshops and Phd Forum*, 2011, pp. 1740–1749.

[9] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, vol. 31, pp. 651–666, 2010.

[10] P. MacKey and R. R. Lewis, "Parallel k-Means++ for Multiple Shared-Memory Architectures," *Proceedings of the International Conference on Parallel Processing*, vol. 2016–Septe, pp. 93–102, 2016.

[11] A. Apon, F. Robinson, D. Brewer, L. Dowdy, and D. Hoffman, *Initial starting point analysis for K-means clustering: a case study*. 2006.

[12] D. J. Bora and A. K. Gupta, "Effect of Different Distance Measures on the Performance of K-Means Algorithm : An Experimental Study in Matlab," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 2, pp. 2501–2506, 2014.

[13] A. Kaminsky, "The Parallel Java 2 Library," in *The International Conference for high performance computing, networking, storage and analysis.*, 2014.

[14] S. Sahni and V. Thanvantri, "Parallel computing: Performance metrics and models," *IEEE Parallel and Distributed Technology*, vol. 4, no. 1, pp. 43–56, 1996.

[15] A. Kaminsky, "Measuring Speedup," in *Building Parallel Programs: SMPs, Clusters & Java*, 2009, pp. 99–110.

[16] A. Grama, A. Gupta, G. Karypis, and V. Kumar, *Introduction to Parallel Computing*, Second Edi. Addison Wesley, 2003.

[17] A. Kaminsky, *BIG CPU , BIG DATA: Solving the World's Toughest Computational Problems with Parallel Computing*. 2015.

[18] T. Zhang, R. Ramakrishnan, and M. Livny, "BIRCH: A New Data Clustering Algorithm and Its Applications," *Data Mining and Knowledge Discovery*, vol. 1, no. 2, pp. 141–182, 1997.

International Journal of New Computer Architectures and their Applications (IJNCAA) 8(4): 179-185
The Society of Digital Information and Wireless Communications, 2018 ISSN 2220-9085 (Online); ISSN 2412-3587 (Print)

Circuit Analyses and Considerations on Advanced Inverters Constructed by Minimum Circuit Components
- Pursuit for concise PCS of photovoltaic power generations -

Keiju Matsui 1,2, Eiji Oishi 2,
Mikio Yasubayashi 1, Yuuichi Hirate 1, Sudip Adhikari 1, Masaru Hasegawa 1

1; Chubu University Kasugai 487-8501, Japan
2; Minna-denryoku, Inc., Setagaya Monozukuri Gakko, 210 Setagaya 154-0001, Japan

ABSTRACT
Photovoltaic power generations (PVG) have been used generally and applied widely. Various power conditioning systems-PCS used in PVG have been also working on by many researchers. In addition to usual utilization, such PVGs are often applied for the time of disaster. In such a way, the solar panels having limited power are almost installed in limited area such as on top of the roof of the building. Some medical institutions have fairly desire to keep such PVG since they must keep the lives of people. The generating power in such case is fair limited, so the system construction should balance the reduced power. In such reasons, it is important to construct toward simple one. In this paper, in order to give the reply, simple and concise PCS, especially novel inverter is considered. Considering fair reduced power and narrow space of installation, the system constructions should be much compact. The circuit strategy which gratifies their operating characteristic is presented and analytically discussed about circuit construction as advanced converter. The circuit operation is confirmed by using the circuit software constructed by PSpice.

KEYWORDS: Circuit software, PSpice, Power Conditioner, PCS, Solar cell, Minimum construction, Buck-boost inverter, Photovoltaic power generation

1 INTRODUCTION

In modern medical care, the development of the structural function in the operating room is remarkable [1,2]. The endoscopic surgery including surgical robot and the catheter intervention has been applied, so that such remarkable operating technics have been developed with robotic operating room and hybrid operating room. For almost electrical equipment using in such medical facilities, even instantaneous interruption should not be permitted. In general, by means of larger scale interruptible power supply installed by generator and batteries is provided. In such system, however, the system scale becomes large accompanied by high cost [3,4].

The PCS including inverter have been presented in various PVG, so far [3,4]. However, it is necessary to reduce the cost even more. It is said that the system is approaching to an ideal one with respect to efficiency and construction strategy, but that cost would prevent wide spread. In such discussions, there are many subjects to be solved to utilize the PV power in utility interactive power generation. Even more, various safeguard equipment required according to regulations make the cost increase. In fairly reduced power PCS as mentioned above used in limited facilities. In such case of reduced generating power, that is, in such PV power generation systems, there are so many subjects to be resolved [5-7].

The authors have been studied in a series of the small power PV system [7-9]. In this paper, some simple PCS systems, especially the components like inverters and chopper are integrated toward simple construction, which are presented and analyzed by using circuit software, PSpice. In the software, various techniques are originally and skillfully devised and presented. The whole result is analytically discussed.

2 PROPOSAL OF ADVANCED INVERTER

2.1 Circuit configuration

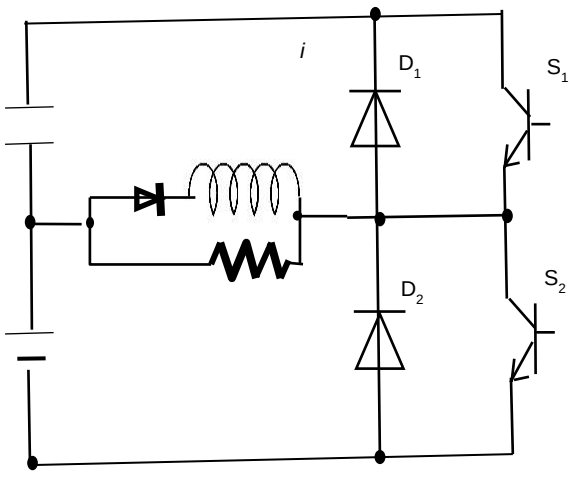


Figure 1. Proposed Advanced Inverter.

Fig.1 shows proposed circuit configuration having resistor load R. In comparison to the conventional high frequency inverter, where in half bridge circuit using boosted chopper, the obtained high frequency is rectified as dc-dc converter [8], the proposed configuration seems to be the same as the number of the circuit component

Fundamental construction is presented in the figure. Various derived circuit construction could be obtained from this circuit. With regard to the efficiency, however, the power conversion process of the proposed one is simple and the number of conversion stage is reduced, so the proposed circuit is prominent in comparison. Those validities are confirmed as follows;

When the switch S_2 is turned-on at the beginning, the power supply provides the load resistor R and inductor L in parallel connection. Secondly, when S_2 is turned-off and S_1 is turned-on, the stored energy in L is provided to R and at the same time the excess energy is delivered to the auxiliary capacitor. During last half period, as the supplied power from L is reduced gradually, a deficiency for supplied power from L is provided from auxiliary capacitor C. Looking at the whole period, it can be seen that the contribution for power transmission through inductor L is predominant. On the other hand, the power transmission through C is a role of auxiliary power transmission. The percentage of power transmission from L to R is about 30% at usual specification, while one from L to R through C is 20%. For your information, the one of the direct transmission E to R is 50%. It is interesting that the auxiliary capacitor C plays a role of like filter.

In Fig.2, the power transmission chart is represented, where the power line is plotted by heavy solid line. That transmitted power is fairly large, while thin line represents small power transmission. Looking at the whole period, it can be seen that the contribution for power transmission consists of direct transmission from E to R. In addition, the route through inductor L is also predominant. It is remarkable strategy that the power transmission is constructed by the direct transmission. Furthermore, as another route, one is via-point of L and the other is via-points of L and C.

Fig.3 shows the power flow chart for the conventional high frequency inverter reported as BHB-Boost Half Bridge [8], in which the circuit mechanism is compared and discussed with our strategy. The input power supply voltage is boosted in the half bridge inverter. After this operation, the high frequency ac power is obtained. After the energy of input power supply E is stored into L, the capacitor C is charged from power from L at switch turned-off. By means of this capacitor charge, the load R is supplied. The figure shows the flowchart of such operating mechanism, where there are three step power transmissions. On the other hand, for the proposed high frequency conversion, main transmission is performed by direct conversion from E to load, the other is a simple transmission way via single L or C. The number of conversion stage is much reduced. Consequently, the transmission efficiency can be expected as much improvement.

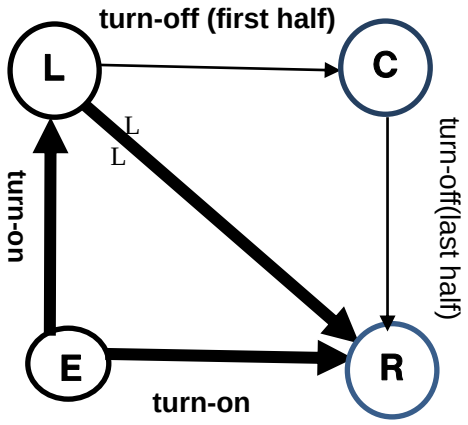


Figure 2. Power Flowchart of Novel Inverter.

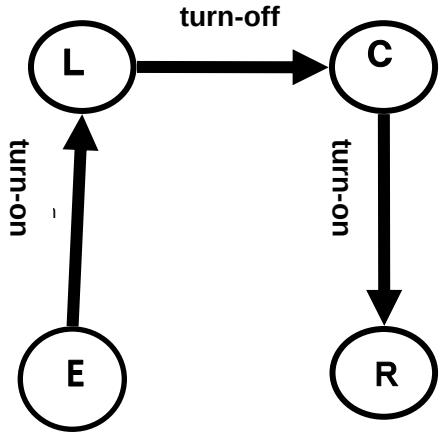


Figure 3. Power Flowchart of Half Bridge.

3 FURTHER DEVELOPMENT TOWARDS BUCK INVERTER

3.1 Operation and characteristic of buck inverter

For further development with minimum circuit component, another method can be derived as shown in Fig.4. By means of alternating switching of S_1 and S_2, the power is delivered to the load R, where inductor L could be connected in series to R. The sinusoidal wave could be obtained, if desired. In this paper, however, the rectangular wave is supposed. Unbalanced charge of C might be concerned, but it can be seen that balanced charge characteristic can be realized by following discussion. In steady state condition, when S_1 is turned on, the relationship is

$$E - vc(0) = Ri^{+}. \cdots\cdots\cdots\cdots (1)$$

When S_1 is turned off,

$$vc(0) = Ri^{-}$$

where $vc(0)$ is charged voltage of C,
i^{+} ; is charging current,
and i^{-} is discharging current.

$$\cdots\cdots\cdots\cdots (2)$$

・In steady state condition,

$i^{+} = i^{-}$, $vc(0) = E/2$

C is charged to half the supply voltage

$$\cdots\cdots\cdots\cdots (3)$$

・In transient condition,

when $S_1$is turned on,

$$E - vc(t) = RC \times dvc(t)/dt.$$

$$\cdots\cdots\cdots\cdots (4)$$

When S_1 is turned off

$$vc(t) = - RC \times dvc(t)/dt$$

$$\cdots\cdots\cdots\cdots (5)$$

When the gradient of V_C, $dvc(t)/dt$ becomes equal, relationship of $E = vc(t)/2$ establishes and C is charged by $E/2$.

Fig.5 shows power flowchart. When S_1 is turned-on and S_2 is turned-off, the load R and C are supplied simultaneously. When S_1 is turned off and S_2 turned on, the R is supplied from power of C. In such a way, the half power from E is directly supplied to load R at S_1 turned on. This operation contributes to efficiency improvement that is fair advantageous. Fig.6 shows the flowchart of conventional BHB inverter for comparison [8]. For half bridge, C is connected in series with supplied power, image of sequential power transmission may be occurred. But actually, the circuit is changed by two steps. Thus, efficiency consideration should be taken into account a little.

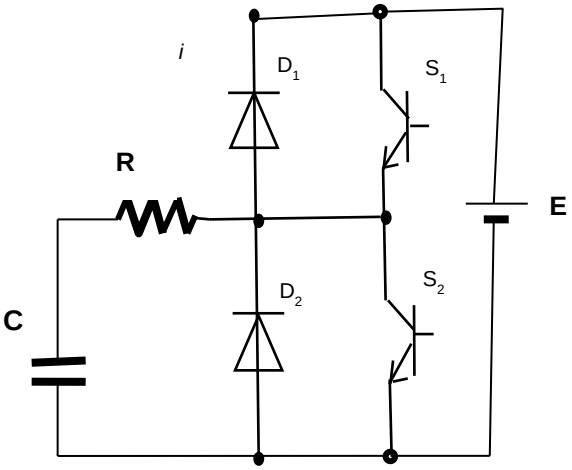


Figure 4. Circuit Configuration of Buck Inverter.

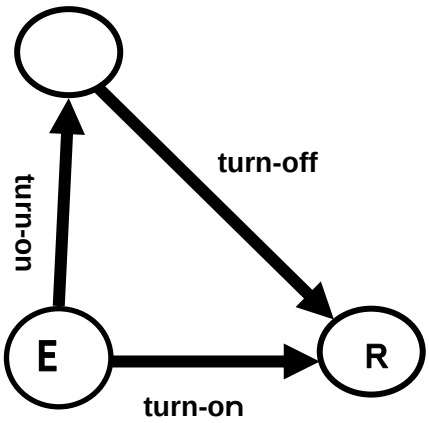


Figure 5. Power Flow of Buck Inv.

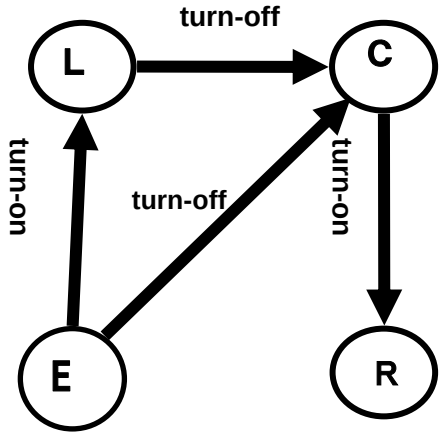


Figure 6. Flowchart of Conventional BHB Inverter. [10].

4 COMPARISON OF ESTIMATED TRANSMISSION EFFICIENCY

4.1 Local transmission efficiency

A. Proposed Buck Boost Inverter

According to the value of inductor, the by-passed load current to the capacitor is varied and the efficiency becomes also varied. The influence due to inductor ripple is considered and the operation is analyzed.

・ *L is sufficiently large* **;** The by-passed current through C becomes much reduced, in which conversion efficiency is a little improved. In order to get hold of conception of efficiency value, it is assumed that efficiency between single stage is, for example $\eta = 90\%$. Thus,

$$\mathrm{E}\rightarrow\mathrm{R}\,(\eta\times\tau_{on}=45\%),\ \mathrm{E}\rightarrow\mathrm{L}(\eta\times\tau_{on}=45\%),\ \mathrm{L}\rightarrow\mathrm{R}\,(\eta\times\tau_{off}\times\eta=40.5\%) \quad \cdots\cdots\cdots\cdots(6)$$

Under consideration of power flowchart in above Fig.5, the total efficiency can be obtained as follows,

$$(\eta\times\tau_{on}=45\%)+(\eta\times\tau_{off}\times\eta=40.5\%) = \underline{85.5} \quad \cdots\cdots\cdots\cdots\cdots\cdots(7)$$

In equation, the total efficiency can be obtained by summation of each value during τ_{on} and τ_{off} period.

・*L is not sufficiently large.***;** The power through C is dependent on the current ripple relative to average load current I_o. For example of assumption, the relative value will be given by $R/(L4f_o) = 1/2$. Thus, the conversion efficiency can be calculated. The power transmission through C is assumed to be constant, $\alpha_C = 1/8$. The direct transmission to R is the remainder, $\alpha_R = 7/8$. In such way, the whole transmission is performed as follows;

$$\mathrm{E}\rightarrow\mathrm{R}(\eta\times\tau_{on}=45\%),\ \mathrm{E}\rightarrow\mathrm{L}(\eta\times\tau_{on}=45\%),\ \mathrm{L}\rightarrow\mathrm{R}\,(\eta\times\tau_{off}\times\alpha_R\times\eta=35.4\%),\ \mathrm{L}\rightarrow\mathrm{C}(\eta\times\tau_{off}\times\alpha_C\times\eta=5.1\%),\ \mathrm{C}\rightarrow\mathrm{R}\,(\eta\times\tau_{off}\times\alpha_C\times\eta\times\eta=4.6\%) \quad \cdots\cdots\cdots\cdots(8)$$

To be put in order, the following total efficiency can be obtained, that is.

$$(\eta\times\tau_{on}=45\%)+(\eta\times\tau_{off}\times\alpha_R\times\eta=35.4\%)+(\eta\times\tau_{off}\times\alpha_C\times\eta\times\eta=4.6\%)=\underline{84.98\%} \quad \cdots\cdots\cdots\cdots\cdots(9)$$

B. Boost Half Bridge-BHB type [8];

・When L is not large ;

the compensated capacitor current includes the ripple, whose value is assumed to be the same as the above example. The power transmission ratio through the upper capacitor is assumed to be constant $\alpha_C = 1/8$. The direct transmission root through L, C_2 to R is the remainder, $\alpha_R = 7/8$. Thus,

$$\mathrm{E}\rightarrow\mathrm{L}(\eta\times\tau_{on}=45\%),\ \mathrm{L}\rightarrow\mathrm{C}_2\rightarrow\mathrm{R}\ (\eta\times\tau_{off\text{-}on}\times\alpha_R\times\eta\times\eta=31.89\%),\ \mathrm{L}\rightarrow\mathrm{C}_1\rightarrow\mathrm{R}(\eta\times\tau_{off}\times\alpha_C\times\eta\times\eta=4.55\%) \quad \cdots\cdots\cdots(10)$$

where $\tau_{off\text{-}on}$ means relative turn-on period and turn-off period represented by $\tau_{off\text{-}on} = 1/2$.

・*When L is sufficiently large* **;** The deviated current through upper C is almost zero. The efficiency is calculated as follows;

$$\mathrm{E}\rightarrow\mathrm{L}(\eta\times\tau_{on}=45\%),\ \mathrm{L}\rightarrow\mathrm{C}(\eta\times\tau_{off}\times\eta=40.5\%),\ \mathrm{C}\rightarrow\mathrm{R}(\eta\times\tau_{on}\times\eta\times\eta=36.45\%),\mathrm{E}\rightarrow\mathrm{R}(\eta\times\tau_{off}=45\%) \quad \cdots\cdots\cdots\cdots(11)$$

The total efficiency is given by

$$(\eta\times\tau_{on}\times\eta\times\eta=36.45\%)+(\eta\times\tau_{off}=45\%) = \underline{81.5\%} \quad \cdots\cdots\cdots\cdots\cdots\cdots(12)$$

where $\tau_{off\text{-}on}$ means relative turn-on period and turn-off period represented by $\tau_{off\text{-}on} = 1/2$.

To be put in order, the following total efficiency can be obtained.

$$(\eta\times\tau_{on}=45\%)+(\eta\times\tau_{off\text{-}on}\times\alpha_R\times\eta\times\eta=31.8\%)+\eta\times\tau_{off}\times\alpha_C\times\eta\times\eta=4.5\%)=81.3\% \quad \cdots\cdots(13)$$

C. Buck inverter type ; Each transmission efficiency is given by the similar manner as before.

$$\mathrm{E}\rightarrow\mathrm{R}(\eta\times\tau_{on}=45\%),\ \mathrm{E}\rightarrow\mathrm{C}(\eta\times\tau_{on}=45\%),\ \mathrm{C}\rightarrow\mathrm{R}(\eta\times\tau_{off}\times\eta=40.5\%). \quad \cdots\cdots\cdots\cdots(14)$$

The total efficiency is $\underline{85.5\ \%}$

D. Conventional Half bridge inverter; Firstly, switch is turned-on, the power is transmitted as follows;

E→R($\eta\times\tau_{on}$=45%), E →C_1($\eta\times\tau_{on}$=45%), C_2→ R($\eta\times\tau_{off}\times\eta$ = 40.5%).. ·····(15)

By the next switching, the operation of transmission is repeated.

E → R ($\eta\times\tau_{on}$=45%), E →C_2($\eta\times\tau_{on}$=45%), C_2 →R($\eta\times\tau_{off}\times\eta$ = 40.5%). ·····(16)

The total efficiency becomes 85.5%.

E. Comparison of Characteristics among Variable Converter

Table 1 Characteristics of Variable Converters.

	Eff %	Num.	Load V	Max.curren
Buck-Boost Inv	85.5	8	*E*	*I*
(with ripple 1/2)	(85.0)	(8)	(*E*)	(*I*)
Boost-Half-Bridge	81.5	9	*E*	*I*
Buck Inverter	85.5	7	*E*/2	*I** 2
Half Bridge Inv.	85.5	8	*E*/2	*I** 2

Table 1 shows characteristics mainly for conversion efficiency among various converters. Two converters on the upper two stages and on the lower two stages are analogous characteristics, respectively.

Firstly, as comparing two converters, Buck-boost Inverter and BHB [8] on upper stage, both inverters are having favorable feature of high load voltage. The maximum device current for BHB becomes double that is fairly disadvantage. The number of devices of the proposed Buck-boost Inv. is decreased from nine to eight, which brings the improved efficiency by a few percent. Secondly, on the bottom stage by two inverters, the disadvantage is that the supplied voltage *E* is given by 1/2, while the current is double . This feature brings shortage of efficiency. For withstand current of device, however, the increased current over load current does not flow. For buck inverter, as compared to conventional Half Bridge Inverter, the number of devices is decreased by unity.

5. DETAILED OPERATIONAL DISCUSSION FOR PROPOSED BUCK INVERTER

In advance of confirmation in the experiment by circuit simulation, the circuit performance is analyzed in detail by using operating circuits. The operation starts from S_1 turning-on in Fig.7.

・Period I

When S_1 is turned-on, conduction to the load starts. At the same time, the charge toward the capacitor begins. The supplied power from E is $E\times i$. Consequently, the circuit differential equation is given by

$$E = L\,\mathrm{d}i/\mathrm{d}t + Ri + \int i/C\,\mathrm{d}t \quad \cdots\cdots(17)$$

In the right side equations, the front two terms are the power which burdens the load and the last term is for the capacitor C charging. When the switch S_1 is turn-off, this period comes to an end.

・Period II

When S_1 is turned-off. Due to the stored energy of inductive load, the current is continued to flow where current flow is supported through D_2. That is, the current is commutating to this path and continues to flow into the capacitor. This energy is supplied from the load of inductive component.

・Period III

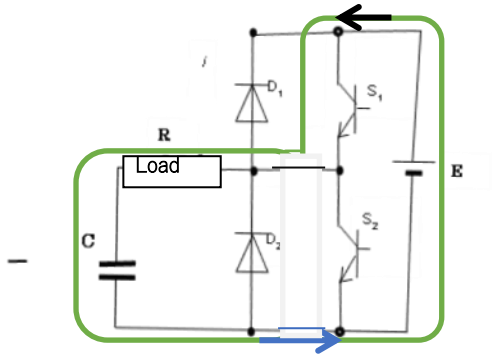
After the energy of L is discharged, the current becomes zero. According to the polarity of capacitor voltage, the current begins to flow in that direction. That is, at the end of period I, S_1 is turned-off and S_2 is turned-on, according to the polarity of capacitor voltage, the current flowing starts and the load is supplied. The circuit equation is as follows;

・Period IV

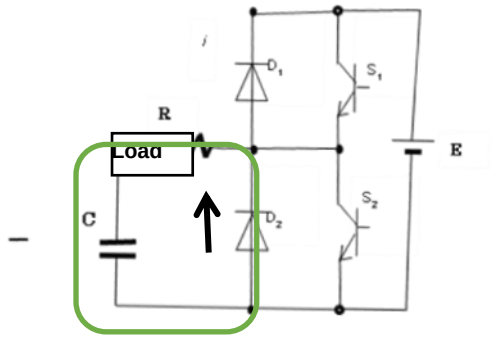
When S_2 is turned-off and S_1 is turned-on, flow is reversed ones during this period when the D_1 is conducted. Namely, due to the stored energy of the inductive load the operation is established. According to the load of inductive component, the energy is returned to the power supply and this period comes to an end. The power regenerative operation is only this period. In this operating circuit, even if there is only one capacitor, that is, even if there is no another capacitor, usual operation can be performed successfully. As a result, the circuit component is reduced by unity. That is this feature is remarkable characteristic. The stored energy of L and C is regenerated towards E and supplied to R as follows.

$$E_C - E = L\,\mathrm{d}i/\mathrm{d}t + \mathrm{R}i \quad \cdots\cdots(18)$$

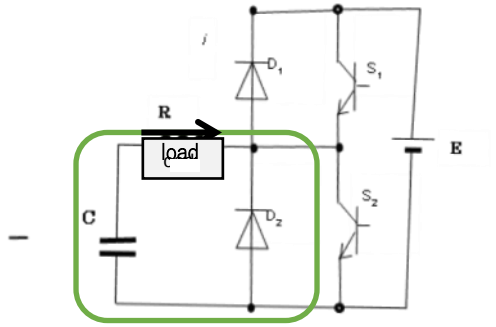
In usual half bridge, the regenerating operation towards supply does not occurs, so this operation characteristic is disadvantageous.



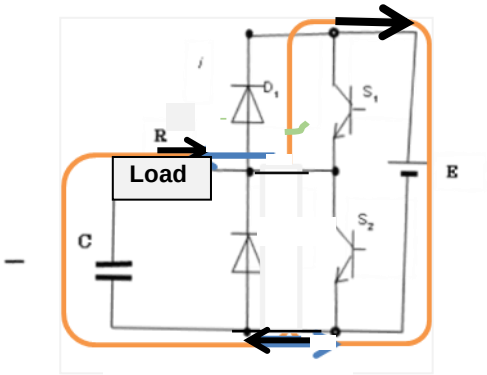
(a) Period I



(b) Period II



(c) Period III



(d) Period IV

Figure 7. Operational Circuit of Buck Inverter.

6. EXPERIMENTAL CONFIRMATION BY USING CIRCUIT SOFTWARE, PSPICE

Fig.7 shows schematic circuit configuration, constructed by circuit software by PSpice. The corresponding circuit is shown as proposed inverter for the first mentioned in Fig,1. The transistor switches are represented by devised pulse generators, operating switching period is shifted half cycle each other. When the switching is turned on, the generator voltage becomes zero. When the switch is turned off, the generator becomes high voltage. At that time, the current is blocked to flow. The dammy resistors play the role to keep the execution of the calculation of differential equations in the software. Fig.8 shows the simulated result by using simulator of Fig.7. Each result shows anticipated ones.

6 CONCLUSIONS

The paper is presented and confirmed by circuit software, PSpice, whose idea is obtained from unified inverter circuit constructed by chopper and inverter circuit. The number of the conventional corresponding circuit configuration is totally ten, while for the proposed construction, the number is eight. This result is the reason why the proposed inverter is called as minimum circuit construction. In parallel load method, which can be developed to dc to dc converter, as the number of conversion stages is reduced satisfactorily, an improved efficiency can be achieved. Finally, as looking at the whole view compared to the usual PVG having large scale inverter. The characteristic may be, however, a little deteriorated. Because of providing towards limited area power supply, quality of electric power is fairly permitted. The authors have been researching about superb photovoltaic power generation system for medical facilities. In such place, employing electrical equipment is strictly required in the standard based on regulation according to like “Japan Industry Standard” [9,10], where requirement for basic safety and

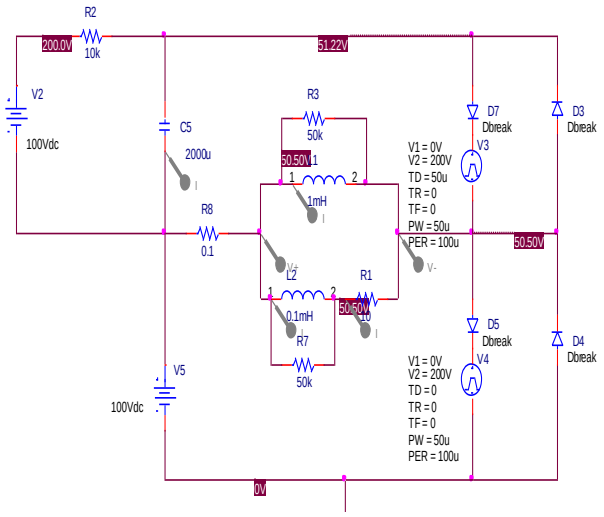


Figure 7. An Example of Schematic Circuit in Simulator, PSpice.

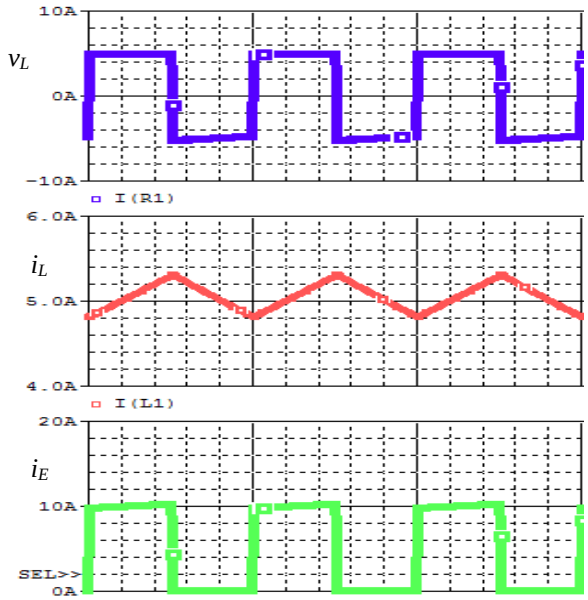


Figure 8. Simulated Waveforms of Proposed Jnverter.

essential performance for the medical equipment, is dictated. The important degree of emergency power supply is varied according to its treatment for medical electrical equipment like life sustainable system, operating room light maintenance, etc. [11]. For example, in artificial respiration equipment or patient monitor, if no battery installation system is adapted, such power supply system is not permitted. In optimum design specification of PV power generation system, adaptive installation may be more efficient.

In the near future, with a spread of ultra-large-capacity and ultra-high-speed internet communication system, remote control surgery like in remote island may be realized, it is said. In such a case, even more reliable power supply system will be important.

ACKNOWLEDGEMENTS

This research is mostly supported by a grant of Research Institute for Life and Health Sciences of Chubu University - Short Term Research Project. We would like to express our appreciation to who it may concern about this project.

REFERENCES

[1] http://hcea.umin.ac.jp/files/pdf/h24files/, "Emergency power supply and medical electrical equipment", 2012.6

[2] New Technology Research Review Committee: Research Institute on Building Cost: "Electrical Installation in Hospital", No.94, pp/74-90, 2016.7

[3] C.-M. Wang; "A novel single stage full bridge buck-boost inverter", *IEEE Trans. on Power Electronics*, vol.19, pp.150-159, 2004

[4] K. Matsui, E. Oishi, Y. Kawata, M. Yasubayashi, M. Umeno, H. Uchida, M. Hasegawa, "Pursuit for Simple Power Conditioner and System Construction of Photovoltaic Power Generation as Veranda Solar", Proceedings of the 3rd International Conference on Industrial Application Eng., pp.531-536, 2015-3

[5] Sinichi Jinno, "Chopper Type Inverter Circuit", Patent publication number -2006-174653, (2006-6)

[6] Fang Zheng Peng: "Z-Source Inverter", IEEE Transactions on Industry Applications, Vol. 39, No. 2, pp.504-510, March/April 2003

[7] K. Matsui, E. Oishi, Y. Kawata, M.Yasubayashi, M. Umeno, M. Hasegawa: "Pursuit of Simple PCS for Photovoltaic Power Generation – Optimum Waveforms", IEEE-Intelec 2015, pp.915-920, 2015-10

[8] Hiroshi Unno, Masanori Hayashi,Yoshiaki Matsuda, "High Efficiency Low Output Voltage DC/DC Power Supply", IEICE, 2000-58, pp.41- 48, 2001-2

[9] http://search.e-gov.go.jp/ : "Japan Industry Standard : JIS T 0601-1", 2017

[10] http://hcea.umin.ac.jp/files/pdf/h24files/, "Emergency power supply and medical electrical equipment", 2012.6

[11] New Technology Research Review Committee-Research Institute on Building Cost,,: "Electrical Installation in Hospital", No.94, pp/74-90, 2016.7

International Journal of New Computer Architectures and their Applications (IJNCAA) 8(4): 186-197
The Society of Digital Information and Wireless Communications, 2018 ISSN 2220-9085 (Online); ISSN 2412-3587 (Print)

A Study for Development of Autonomous Paddy-weeding Robot System –An experimentation for autonomously straight-running based on compass-compensation–

Masashi SUGIMOTO[1*] Yasuhiro INOKI[1] Tomoki SHIRAKAWA[1]
Kanta TAKEUCHI[1] Daiki YOSHIOKA[1] Haruki FUKITA[1]
Toshiyuki YAMAJI[1] Mio ENDO[1] Patchara NAMEEYA[2]
Hiroyuki INOUE[3] Manabu KATO[3] Shiro URUSHIHARA[4]
Kazunori HOSOTANI[3] Hitoshi SORI[3]

[1]Department of Electronic Systems Engineering,
National Institute of Technology, Kagawa College,
551 Kohda, Takuma, Mitoyo C., Kagawa Pref., 769-1192 Japan.
Email: sugimoto-m@es.kagawa-nct.ac.jp (∗ : corresponding author)

[2]Department of Electronic and Telecommunication Engineering,
Rajamangala University of Technology Thanyaburi,
39 Moo 1, Rangsit-Nakhonnayok Rd, Thanyaburi, Khlong Luang, Pathum Thani, 12110 Thailand.
Email: Patchara.Na@hotmail.com

[3]Department of Integrated Science and Technology,
National Institute of Technology, Tsuyama College,
624-1 Numa, Tsuyama C., Okayama Pref., 708-8509 Japan.
Email: {inoue, kato, hosotani, sori}@tsuyama-ct.ac.jp

[4]Department of Electrical and Computer Engineering,
National Institute of Technology, Kagawa College,
355 Chokushi, Takamatsu C., Kagawa Pref., 761-8058, Japan.
Email: urushi@t.kagawa-nct.ac.jp

ABSTRACT

In the agricultural field of Japan, there is two issues are facing; population aging and labor shortage are progressing, rapidly. In order to solve these problems, it is indispensable to create a method that improves productivity and labor-saving technology in agricultural works. A weeding robot is one of an approach for achievement of these problems, it can be considered that labors will be released from tough or dangerous works. In this study, autonomously controlling the weeding robot will be focused on. To achieve this, an algorithm will be proposed that autonomously constructs a state-action space based on various sensor information and can apply it to actually work.

The main contribution of this paper is in the development of vision-based navigation and integrated control system for straight-running or turning behavior to guarantee performance during of the working. The presented system benefits from a magnetic compass, a fixed camera, and a P-control module. An estimation method of boundary of the workspace distance obtains view of in-front-of the robot using a monocular camera to enable detect edges of the workspace. Moreover, an obstacle detection method obtains in-front-of view of the robot using the camera. In this method is applied a deep-learning module to detect and cognition an obstacle. A rotation controller is developed to counter-rotation turn the robot such that at the target boundary point, to avoid the robot reach out of the workspace, or avoid an obstacle. In addition, the P-control module provides that command for straight-run to prevent the robot get stuck or undesirable change a course in the mud of paddy-field. In verification experiment, a state-action space, including the position of the obstacle was constructed and can avoid an obstacle, in each work. Further, we have been confirmed that the robot can counter-

rotation turn 90 degrees (spin-turn) at the boundary point of workspace.

KEYWORDS

Agricultural robot, Crawler, Autonomous navigation, Revise command generation, Work saving technology, Weed management, Edge detection

1 INTRODUCTION

In recent years, aging and labor shortage are rapidly progressing also in the agricultural sector of Japan. Because of this reason, it is indispensable for a technology concerning labor-saving of agricultural work and improvement of productivity. Therefore, introduction of a robot in the agricultural field has been pushed forward, aggressively.

Agricultural robots are roughly classified into a vehicle type robot typified by a tractor [1, 2], a manipulator type robot performing a movement similar to a human hand [3, 4], a facility type robot performing a milking operation, a power assist type an assistance of a working of a person [5, 6]. Even among them, our group has been conducting research on weeding robots, especially based on the vehicle type robot [7, 8]. In other words, a paddy-weeding robot is described in fig. 1 has been focused on.

Even when it is said to be a weeding robot

Figure 1. A Paddy-weeding Robot.

in general the range is extensive and complicated; working environments according to purpose are also diverse, such as weeding on the levee or wet rice culture, grass cutting on the slope of the inter-mountain areas.

The concept of the weeding robot system

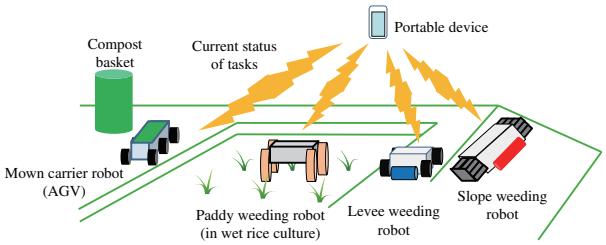


Figure 2. A Covered Area the Weeding Robot Systems Followed.

will be developed is shown in fig. 2. Examples of these element technologies include a traveling section having fine running performance suitable for various working environments, a weeding section for removing weeds, a sensor and control system for autonomous working, and an ICT for providing information to the labor in the remote area. In this paper, the traveling control algorithm for autonomous working will be focused on.

It's necessary to avoid an obstacle autonomously for enable to autonomous control the weeding robot. However, the working environment is often in a non-maintenance environment. For this reason, it's indispensable to develop a flexible behavioral algorithm what to avoid obstacle by utilizing the detection and recognition that. In this paper, we propose a method to enable autonomous traveling of weed control robots.

This paper is organized as follows: In section 2, how difficult to control autonomously in a situation of weeding will be stated. Further, how to estimate a direction and distance within a boundary, will be motivated. Moreover, details about the noise reduction method for using a magnetic compass in a situation of estimate a direction, and the forward monitoring method using edge detection and deep-learning package, will be stated. In section 3, a verification experiment configuration will be described. In section 4, the summary of this work is concluded.

2 APPLICATIVE RANGE OF THE ROBOT SYSTEM AND PROBLEM OF WEEDING

As mentioned above, the working environment is often a non-maintenance environment, has been stated. Furthermore, since the road surface on which the robot travels will be under non-maintenance environment, there is a possibility that the revolutions and speed of the motor won't be coincide with each other. In this case, it is difficult to estimate the speed from only the rotation speed, or estimate the direction. Further, in case of the robot will be operated near the boundary with the outside of the working area, will be considered. There is a possibility that it may deviate out of the area, in the worst case if the turning behavior won't be performed normally. Or perhaps there is a possibility that the robot will leave from management or the robot itself will be damaged. Moreover, even with a non-contact sensor, depending on the weather conditions of the outside world, unexpected information may be mixed with the information obtained from the sensor. In this case, it may the robot act what is not intended by the designer. As a result unfavorable phenomena may occur from the viewpoint of productivity and safety.

In order to solve the problems mentioned above, the crawler-type prototyped robot in this study is shown in fig. 3. This robot is equipped with a GPS module, an ultrasonic sensor, magnetic compass module with gyroscope and accelerometer, and a camera module.

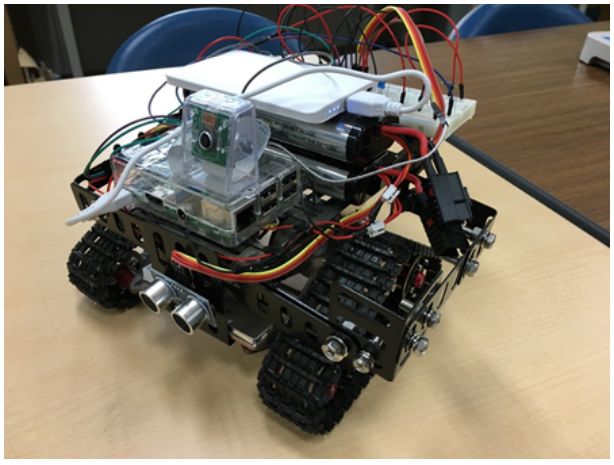


Figure 3. A Crawler-typed Prototyped Robot.

If an obstacle will be detected forward by the ultrasonic sensor, the robot will be stopped temporarily. Next, the front will be photographed by the camera module, and the object will be identified by deep-learning. In this case, when a harmful to human is considered or an object which may possibly fall or be damaged will be detected, the position is recorded by GPS and rapidly retreated. When approaching that point after a certain period of time deceleration. At this time if an object is not detected the point, the position is deleted and normal running will be resumed.

Moreover, in addition, we will try to verify its practicality in order to achieve the following items by using the camera module that installed in the robot system, and the software group that installed in Raspberry Pi 3 Model B+ of the system.

(1) Separation of workspace boundary area for images taken by monocular camera.

(2) Discrimination of out of boundary area or obstacle using deep-learning (Google Inception v3).

(3) Whether or not the continuing ability of working by counter-rotation turn based on the result of (1) or (2).

2.1 A Strategy of Operate of Weeding

In the former study, the proposed method cannot traveling along the wet rice line straightly, and reach the edge of the field, autonomously. Moreover the method of the turning technique that is necessary to turn back and enter the adjacent row of rice lines, but it is a challenge to recognize the edge of the field because of the system had been driven by remote-controlled. Therefore, in this study, the image processing method will be applied for straight run and turn like a spin-turn at the edge of the field as shown in fig. 4. A decision of the turning point based on the photographed result using edge detection is as shown in fig. 5.

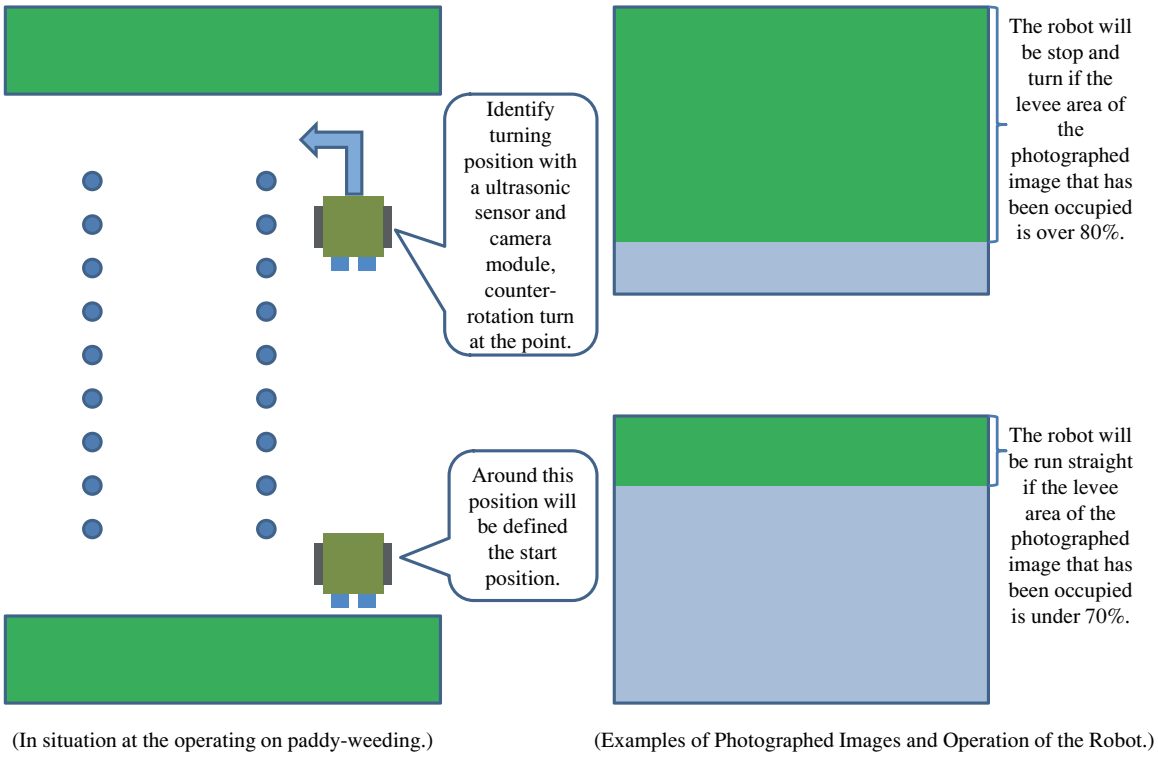


Figure 4. Decision the Turning Point based on the Photographed Result using Edge-detection.

2.2 How to Spin-turn using Magnetic Compass ?

As mentioned earlier, it is difficult to turn right or left at an angle using the difference of the rotation angle of the left and right motor, precisely. Therefore, in this study, the magnetic compass of the 9-axis sensor complex MPU-9250, will be focused on.

Generally, when the magnetic compass is made horizontal to the ground and rotated 360 [deg.], The geomagnetic data in the x-axis and y-axis directions are plotted in a form that draws a circle. Based on this result, if north will be defined as 0 [deg.] and assuming that the magnetic flux density on the x-axis of the sensor is G_x and the magnetic flux density on the y-axis is G_y the azimuth of the robot to which the sensory output will be given as follows[9]:

$$\theta = \tan^{-1} \frac{G_y - G_{y,\text{off}}}{G_x - G_{x,\text{off}}} \tag{1}$$

$G_{x,\text{off}}$ and $G_{y,\text{off}}$ are offset amounts for canceling the magnetic noise emitted from the motor, Raspberry Pi, a GPS module, or other electronic devices.

On the other hand, the magnetic compass of MPU-9250 used in this study is 3-axes and doesn't consider the attitude of the sensor, as pitching. Thus, as the attitude of the sensor changes, the influence of the geomagnetism in the vertical direction affects the geomagnetism of the $x - y$ axis. As a result, it gives an error in the estimated direction. Therefore, to estimate the roll and pitch angle of the sensor from the 3-axis acceleration sensor of MPU-9250, moreover, try to correct the geomagnetism based on that value. By using the information on the 3-axis accelerometer, a_x, a_y, a_z, the roll angle ϕ and the pitch angle ψ can be calculated as follows:

$$\phi = \tan^{-1} \frac{a_y}{a_z} \tag{2}$$

$$\psi = \tan^{-1} \frac{-a_x}{a_y \sin\phi + a_z \cos\phi} \tag{3}$$

Therefore, by using this angle information and the magnetic compass, the direction ϑ is as

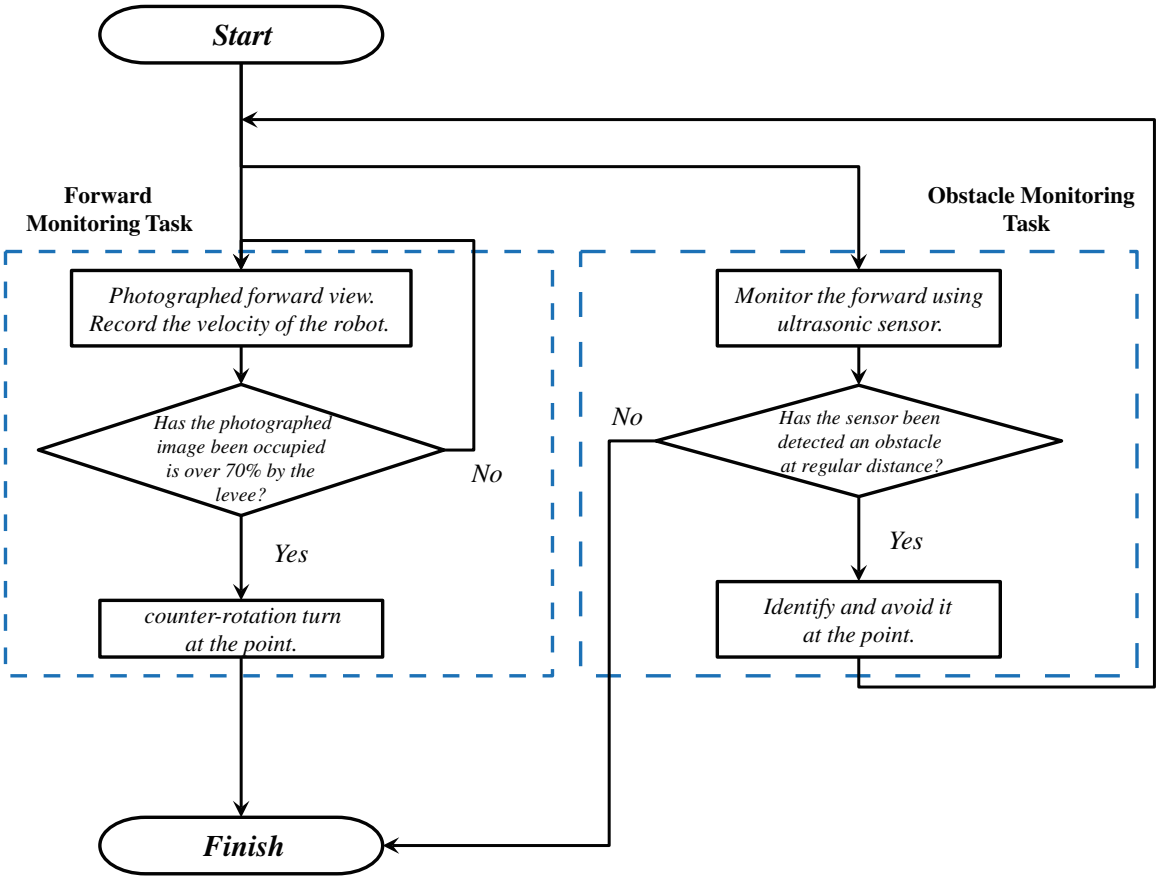


Figure 5. A Flow of Strategy for A Running or Avoiding an Obstacle.

follows:

$$\vartheta = \tan^{-1}\frac{A}{B} \tag{4}$$

$$A = (G_z - G_{z,\text{off}})\sin\phi - (G_y - G_{y,\text{off}})\cos\phi \tag{5}$$

$$B = (G_x - G_{x,\text{off}})\cos\psi + (G_y - G_{y,\text{off}})\sin\psi\sin\phi + (G_z - G_{z,\text{off}})\sin\psi\cos\phi \tag{6}$$

However, in case of detecting of the direction, the result of the magnetic compass may be the mixed with a high frequency noise that make computation of Raspberry Pi or driving noise of motor, or other device during control. Thus, in this study, a median filter and a low-pass filter will be applied to reduce these noises. In this study, we will apply this azimuth estimation algorithm in order to spin-turn. Firstly, the median filter is an effective method that can, to some extent, distinguish out-of-range isolated noise from legitimate sensor features such as edges and lines. Specifically, the median filter replaces a 1-dimensional array by the median, instead of the average, of all values in a neighborhood w of time-series sensory information:

$$y(i,j) = \text{median}\{x(i,j) \in w\} \tag{7}$$

where w represents a neighborhood defined by the user, centered around location $[m,n]$ in the time-series of sensor information.

Moreover, in secondly, a low-pass filter will be applied for reduction of noises. In this study, the low-pass filter is given by the following difference equation:

$$y(nT) = x(nT) + x[(n-1)\cdot T]\; n = 0,1,2,\ldots \tag{8}$$

where $x(n)$ is the filter input amplitude at time n, and $y(n)$ is the output amplitude at sampling time n. Further, where T is the sampling interval in seconds.

2.3 How to Detect Edge of Boundary Area and Obstacle ?

For item (1) of section. 2, the edge detection was performed using the horizontal-direction Prewitt filter is described in eq. (9), and it was decided to cut out an area that could become a boundary line from the result (fig. 6). By carrying out the filtering, it is possible to distinguish the traveling road surface and the slope such as a boundary of workspace with respect to the photographed image.

$$K = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -2 & 1 \\ -1 & -1 & -1 \end{pmatrix} \tag{9}$$

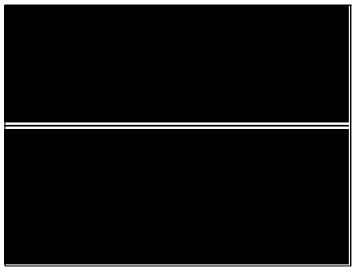
Further, by specifying which part of the photographed image the boundary line obtained by the edge processing exists, the distance of the robot. From this result, the distance to the end of traveling road surface can be specified, moreover, the turning point will be determined from the photographed image, and the turning behavior will be taken.

2.4 How to Predict and Compensate the Action Command in Paddy-field ?

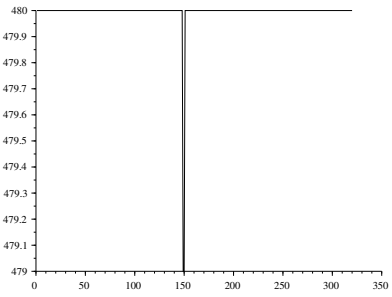
In general, bottom of the paddy field is not necessarily flat. That includes mud, weed, or other things. Sometimes the wheel of robot or other vehicle-typed machine will be affected by these things; it will get stuck or movement course will force to change. At worse, it may occur that damage to rice or perhaps there is a possibility that the robot will leave from management or the robot itself will be damaged, as mentioned in section 2. To avoid these problems, prediction method will be focused on, in this study. The robot will be obtained course of action based on the prediction result if the prediction module will be mounted. In this paper, the time series of compass-direction ϑ will be focused on.

For controlling a robot in a dynamic environment, an action can be chosen based on the current result and former actions and states [10, 11]. In other words, we can obtain the predicted result of the compass-direction and ac-

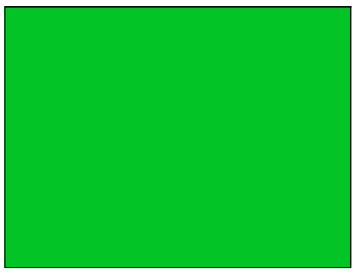
(a) Source Image



(b) Applied Prewitt Filter



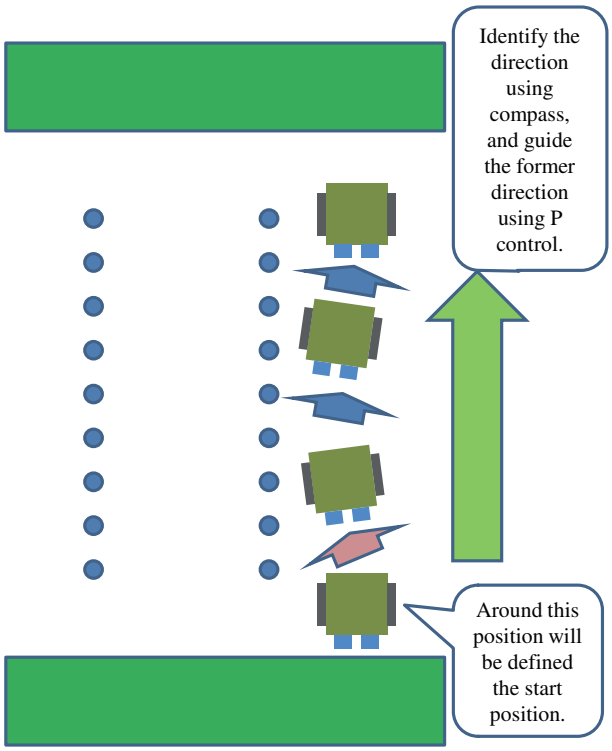
(c) Calculated Feature Quantity



(d) Out of Boundary Area Extraction

Figure 6. An Example of Edge-detection

tion command for DC motors if we can obtain the former compass-direction ϑ series. The mechanism of this method is as follows in fig. 7. To achieve of this mechanism, the follow-



(In situation at the operating on paddy-weeding.)

Figure 7. An Image of Compensate the Action of the Crawler.

ing equations for controlling DC motor based on the compass, will be structured.

$$\mathrm{PWM}_{\mathrm{right}} = \mathrm{PWM} - u \quad (10)$$

$$\mathrm{PWM}_{\mathrm{left}} = \mathrm{PWM} + u \quad (11)$$

$$u = k_{\mathrm{sync}} \cdot (\vartheta\,[t] - \vartheta\,[t-1]) \quad (12)$$

In other words, use the P control for caterpillar synchronization when the crawler runs straightforward because the rotation angle of DC motors is not same even though same PWM is applied (shown in fig. 8).

3 THE VERIFICATION EXPERIMENT

3.1 Outline of the Experiment

In this experiment, the way of two experiments will be carried out; firstly, in order to confirm the usefulness of the proposed method,

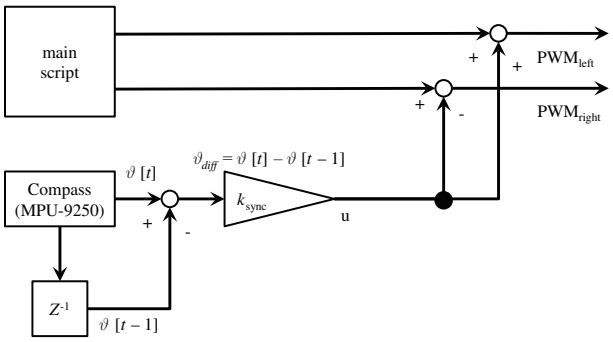


Figure 8. A Block Diagram of P Control for the Caterpillar Synchronization.

the crawler will be able to running straightforward (fig. 9), will be considered. Next, secondly, a ridge of rice fields as the terminal wall will be taken as an example (fig. 10). At this time, the actual taking a picture of forward view and traveling that including running straight or spin-turning, will be repeated. During this process, to verify that it is possible for the robot to turn at a right angle near the ridge.

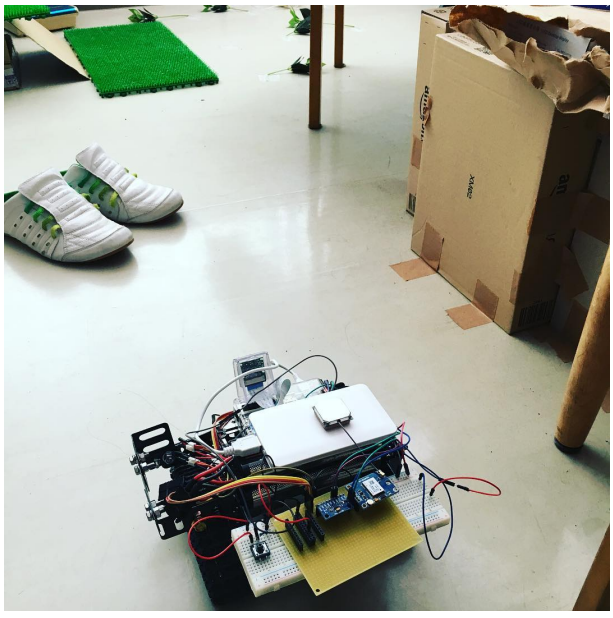


Figure 9. Experimental Environment for Straightrunning.

Figure 10. Experimental Environment for Spin-turning and Straightrunning.

3.2 Experimental Results

3.2.1 An Experimentation of Straightrunning Performance of the Crawler

In this experiment, the crawler had been put on the flat floor (fig. 9). The results of actually applying the proposed method to the straightforward are shown in fig. 11. In this way, it was confirmed that it was possible to running straightforward.

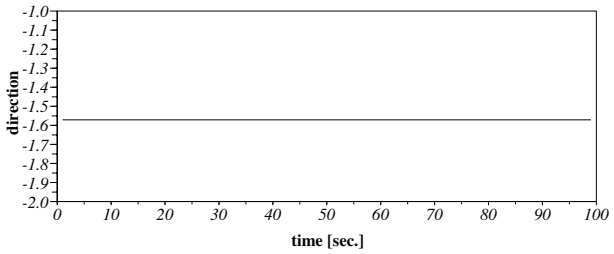


Figure 11. A Tracking Route of the Crawler.

3.2.2 An Experimentation of Running Performance of the Crawler

Here, a wall as an example of levee, will be taken. In this experiment, whether it is possible for the crawler to turn at a right angle, will be verified. During the experiment, the crawler will actually take a picture of a front view for observes the situation, and run or spin-turn, repeatedly. The running record of the crawler will be shown in fig. 13 and 14. From the result, after straight running for a certain period of time, it was confirmed that was turned to the right as seen from the direction of travel. Further, it continued straight running for a certain period of time, has been confirmed.

Figure 12. A Behaviour of Sptin-turning of the Crawler in the Experimental Environment.

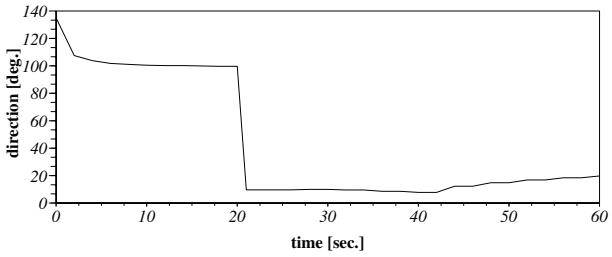


Figure 13. A Process of Migration Pass-way.

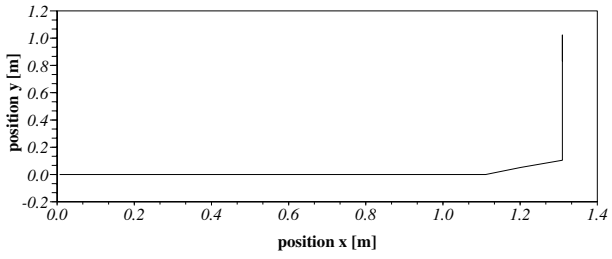


Figure 14. A Trace of Operating the Crawler Movement.

3.3 Discussion on Experimental Results

Now, let's focus on these results. In figure 17(d), the result of actually applying the proposed method to the captured image. In this experiment, the horizontal-direction Prewitt filter had been applied to detect an edge

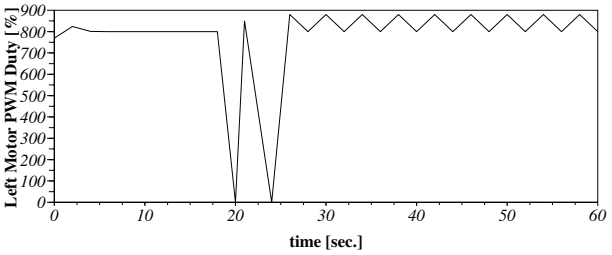


Figure 15. A PWM Duty for Left Motor.

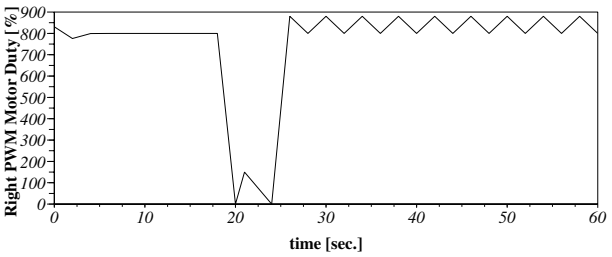


Figure 16. A PWM Duty for Right Motor.

of the boundary area, that has been shown in fig. 17(b). From the result, the sum of the pixel will be calculated to detect the boundary area (fig. 17(c)). In this way, it was confirmed that it was possible to distinguish the water surface and the edge of the wall of the traveling road surface. As similar, in fig 18, the crawler had taken pictures of front to detect the edge of the levee and the surface. In this way, the crawler can detect the edge and obtain the location of spin-turn at the boundary area. Figure 18 had been taken by the crawler to obtain the location of spin-turn. In this figure, an area of the picture had been occupied over 70 percent by the levee. Moreover, it was confirmed that it has robustness to color change due to the surrounding environment, since it is also possible to divide even in the image different from figs. 6 and 17. Next, let's consider about the tracking of movement. Figures 13 and 14 are tracking of movement for the crawler running. In figure 13, the azimuth of the crawler had been hold same direction (almost 100 degrees). However, after 20 seconds, the azimuth had been increased negative-direction and approaching near 10 degrees from initial direction. Finally, it had been converged around 10 degrees. Next, figure 14 will be focused on. Firstly, the crawler had been running straightly. Then, it had been turned and changed movement direction. In this way, it can be confirmed that the crawler had been moved straight or spin-turn based on an edge-detection. Now, let's focus on result of the magnetic compass. The time-waveform of the azimuth obtained by the magnetic compass when during spin-turning (fig. 13) contains many variations that are seen as noise. As this factor, the following issues can be considered:

- Changes in body posture due to unevenness of running road surface.
- A magnetic field changes by the left and right motors at the time of spin-turn.
- A magnetic field changes due to increased power consumption by calculation of Raspberry Pi.

Regarding this point, it can be reduced a certain amount of noise by stopping for a certain period of time at the spin-turning, then measuring the offset amount and correcting it, again. From these viewpoints, we conclude the experimental results are reasonable.

4 CONCLUSION

In this study, we considered a method of self-generating a behavioral strategy by generating a map of the activity-field by itself, while acting in a non-maintenance environment. As a preliminary step, in this paper, we had considered an autonomously-drive algorithm using a self-running mobile crawler-typed robot equipped with the compact monocular camera shown in fig 3. In detail, a turning algorithm that swiveled the spin-turn based on the magnetic compass of MPU-9250 at arbitrary positions by result of the image processing and recognition, had been considered.

In the verification experiments, the effectiveness had been confirmed by implementing the turning algorithm; the crawler had spin-turned in the 9 o'clock direction at the point where the boundary area; end of the workspace had been recognized, and then had been running straight again. As a result of the verification experiment, as the levee had recognized as a location

of the termination as shown in fig. 18 by the crawler, the behavior of the spin-turning had been confirmed, and the effectiveness of the proposed method has been confirmed. As the future work, learning actual objects related to these grasses, assuming the situation of grass cutting and weeding, will be planned. However, there are cases where the weeds are affected by the wind or other things or that will be shaken. To deal these problems, also, an implementation of image sharpening and evaluation of its usefulness is subjected to be studied in the future. In addition, the proposed system will be separated; Raspberry Pi as the sensing and communication function, and Arduino as the driving control function of the motor control or other works. From this plan, load balancing and real-time performance will be improved.

Moreover, the previous study [12] will be applied to behavioral algorithms more efficiency, so that the proposed system can travel and work at constant speed on ramps including uphill roads such as levee will be possible. At the same time, it is necessary to investigate algorithms to introduce multiple robots with different body structures or different roll with individual roles into the workspace, to work cooperatively and efficiently [13].

ACKNOWLEDGMENT

This work was supported by JST A-STEP 18067541.

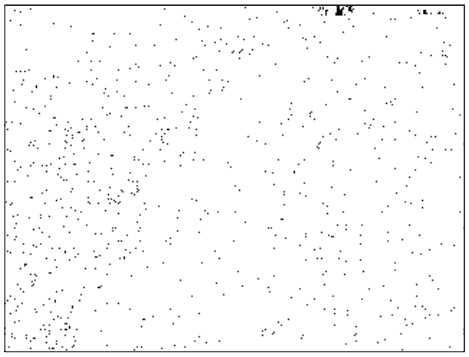
REFERENCES

[1] S. Yaghoubi, N. A. Akbarzadeh, S. S. Bazargani, S. S. Bazargani, M. Bamizan, and M. I. Asl, "Autonomous robots for agricultural tasks and farm assignment and future trends in agro robots," International Journal of Mechanical & Mechatronics Engineering, Vol.13 No. 3, pp.1-6, 2013.

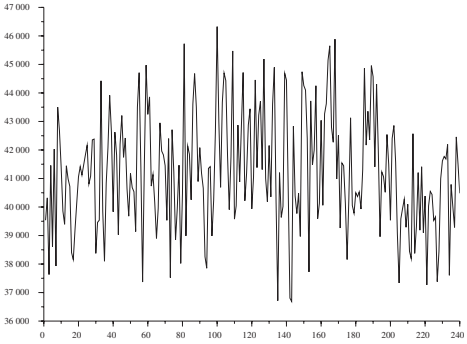
[2] K. Hosotani, Y. Takayama, M. Kato, H. Inoue, H. Sori, S. Urushihara, and M. Sugimoto, "Climbing performance of the flexible spiked crawler track," In Proceedings of the 2018 JSME Conference on Robotics and Mechatronics (ROBOMEC2018), pp.1-4, 2A1-G05, 2018.

[3] S. S. Mehta and T. F. Burks, "Vision-based con-

(a) Source Photo (Experimental Result)



(b) Applied Prewitt Filter (Experimental Result)



(c) Calculated Feature Quantity (Experimental Result)

(d) Out of Boundary Area Extraction (Experimental Result)

Figure 17. An Experimental Result of Edge Detection

Figure 18. A Picture for Obtain the Location of Spin-turn.

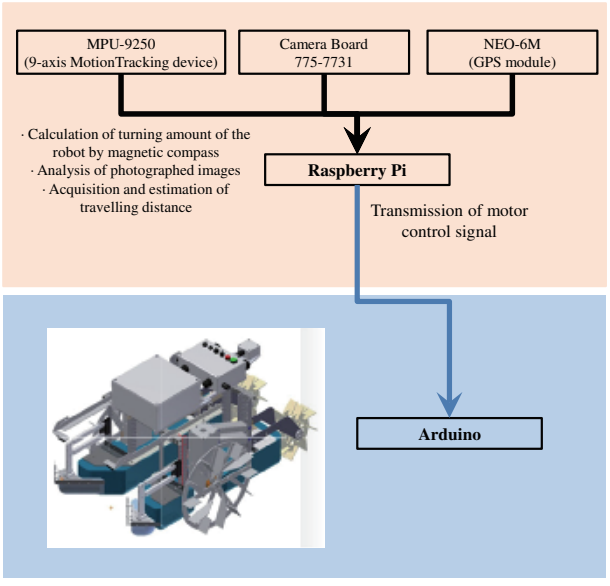


Figure 19. A Description of Future Work.

trol of robotic manipulator for citrus harvesting," Computers and Electronics in Agriculture, Vol.102, pp.146-158, 2014.

[4] L. Bascetta, M. Baur, and G. Gruosso, "ROBI' : A prototype mobile manipulator for agricultural applications," electronics, MDPI, Vol. 6 No.39, pp.1-23, 2017.

[5] H. Inoue and T. Noritsugu, "Development of upper-limb power-assist machine using linkage mechanism - Mechanism and its fundamental motion, " International Journal of Automation Technology, Vol. 8 Issue 2, pp.193-200, 2014.

[6] Y. Imamura, T. Tanaka, and K. Takizawa, "Field test on KEIROKA (Fatigue-reduction) effect and physical fitness change by assistive device for care work, " Journal of Nursing Science and Engineering, Vol. 2 No. 3, pp.142-149, 2015.

[7] Y. Inoki, M. Sugimoto, H. Inoue, M. Kato, H. Sori, S. Urushihara, and K. Hosotani, "A study for construction of autonomous behavior strategy in control of agricultural machinery," In Proceedings of the Japan Society for Precision Engineering 2018 Spring Meeting, pp.835-836, 2018.

[8] M. Sugimoto, Y. Inoki, T. Shirakawa, K. Takeuchi, T. Yamaji, M. Endo, H. Inoue, M. Kato, S. Urushihara, K. Hosotani, and H. Sori, "A Study for Development of Autonomous Paddy-weeding Robot System -An experimentation for autonomously workspace-cognition and counter-rotation turn-," In Proceedings of The Fourth International Conference on Electronics and Software Science, pp.115-124, 2018.

[9] P. Ripka, *Magnetic Sensors and Magnetometers (Artech House Remote Sensing Library)*, Artech House on Demand, 2001.

[10] M. Sugimoto, K. Kurashige, "A Study of Effective Prediction Methods of the State-Action Pair for Robot Control Using Online SVR, " Journal of Robotics and Mechatronics, Vol.27, No. 5, pp.469-479, 2015.

[11] M. Sugimoto, and K. Kurashige, "Future Motion Decisions using State-action Pair Predictions," International Journal of New Computer Architectures and their Applications, Vol. 5 No. 2, pp.79-93, 2015.

[12] M. Sugimoto, N. Iwamoto, R. W. Johnston, K. Kanazawa, Y. Misaki, H. Inoue, M. Kato, H. Sori, S. Urushihara, K. Hosotani, H. Yoshimura, and

K. Kurashige, “Dimensionality reduction for state-action pair prediction based on tendency of state and action, ” International Journal of New Computer Architectures and their Applications, Vol. 7 No. 1, pp.18-28, 2017.

[13] M. Sugimoto, “A study for dynamically adjust-mentation for exploitation rate using evaluation of task achievement, ” International Journal of New Computer Architectures and their Applications, Vol. 8 No. 2, pp.53-60, 2018.

International Journal of New Computer Architectures and their Applications (IJNCAA) 8(4): 198-205
The Society of Digital Information and Wireless Communications, 2018 ISSN 2220-9085 (Online); ISSN 2412-3587 (Print)

Variation Effect of Silicon Film Thickness on Electrical Properties of NANO-MOSFET

A. Tijjani[1], G.S.M. Galadanci[2], G. Babaji[3] and S.M. Gana[4]

Department of Physics, Bayero University, Kano, Nigeria[1,2,3,4]
atijjani.phy@buk.edu.ng (+234 7038277312)

ABSTRACT

Owing to the fact that metal oxide semiconductor field effect transistors (MOSFETs) can be effortlessly assimilated into ICs, they have become the heart of the growing semiconductor industry. The need to procure low power dissipation, high operating speed and small size requires the scaling down of these devices. This fully serves the Moore's Law. But scaling down comes with its own drawbacks which can be substantiated as the Short Channel Effect. The working of the device deteriorates owing to SCE. In this work numerical simulations have been performed to investigate the electronic transport through the Silicon (Si) channel of four terminal Nano-MOS namely; drain, source, top gate and bottom gate. Also, the thickness of Silicon film channel is varied from 1.5 nm, 2.5 nm, 3.5 nm, 4.5 nm and 5.5 nm with other structural dimensions remain unchanged. The simulation is carried out at room temperature (RT) using Nano-MOS simulating software. Three models have been presented such as; ballistic transport using Green's function approach, ballistic transport using semi classical approach, and drift diffusion transport. The electrical properties such as 2D electron density of the sub bands, sub bands energy profile and drain current - gate voltage (I_{DS}-V_{GS}) have been plotted to compare the performance of these three transport models. From the simulation analysis, the drift diffusion transport model shows low performance in comparison with the two other models, maybe due to the electron gas scattering encountered during the transport through Si channel. Meanwhile, Green's function approach and semi classical approach shows almost similar results with high performance.

KEYWORDS
Ballistic transport, Drift Diffusion Transport, MOSFET, Nanoscale, Semi Classical Transport.

1. INTRODUCTION

The numerical modeling of open quantum devices has become an indispensable tool to understand transport physics of semiconductor devices scaled down to nano-meters regime [1]. Non-equilibrium Green's function (NEGF) method is a comprehensive approach to elaborate the quantum transport under external potential bias. Semi-classical approach applies the techniques of Boltzmann kinetic to explain the electron transport. Both of these models are ballistic in nature. Meanwhile, drift diffusion transport has scattering [2].

In 1965 Gordon Moore predicted that the number of transistors per chip would quadruple every three years [3]. The channel length which is an important dimension has been shrinking continuously and will continue to decrease [4]. The reason behind this continuous miniaturizing is to have high speed devices in very large scale integrated circuits. As we are scaling down the size of device, channel length of the device shrinks and this nearness between source and drain reduces the gate electrodes controlling influence on the potential distribution and current flow in the channel which in result deteriorates device performance. The double-gate (DG) transistor is considered one of the most promising devices for extremely scaled CMOS technology generations. Indeed, due to a good electrostatic control of the channel by the two gates, it is expected to provide smaller short-channel effects (SCE), near ideal sub-threshold slopes and higher drive currents when compared to single-gate (SG) transistors [5, 6].

2. THEORITICAL BACHGROUND

2.1. Ballistic and Diffusive Carriers Transport Mechanisms

Before the discovery of ballistic carrier transport, the transport mechanism of the conventional MOSFET is mainly "diffusive", which means that the electron takes a random walk from the source to the drain, traveling in one direction for some length of time before getting scattered into some random direction as sketched in Fig.1.0 below. The mean free path mfp, that an electron travels before getting scattered is typically less than a micrometer (also called a micron = 10^{-3}mm, denoted by μm) in common semiconductors, but it varies widely with temperature and from one material to another [7].

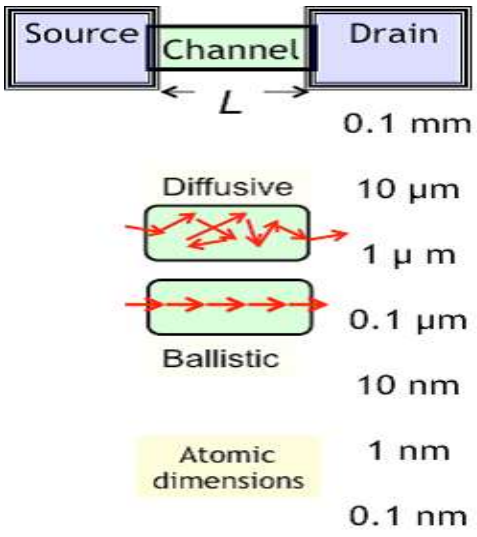


Figure 1.0: Diffusive and Ballistic carriers transport mechanisms

2.2. Transport Equations

Transport equations such as the Boltzmann equation can describe electron dynamic. They determine the dynamics of electron density distribution in response to perturbation such as external electric field and electron density gradient. The electron drift-diffusion equation is given by

$$j_n = qn\mu_n\varepsilon + qD_n\nabla_n = -qn\mu_n\nabla\emptyset + qD_n\nabla_n \qquad (1)$$

Where;

j_n is the electron current density

n is the electron density

μ_n is the electron mobility

D_n is the electron diffusion coefficient

$\emptyset$ is the electric field

ε is the electric potential

2.3. Poisson's equation

Poisson's equation is a fundamental equation describing the spatial relationship between a certain electron density distribution and the corresponding electric field. It holds true no matter which transport equation/model we use, so it is a common routine for all simulation options.

$$\nabla^2\emptyset = -\frac{1}{\epsilon}(p - n + N_D^+ - N_A^-) \qquad (2)$$

Where;

$\emptyset$ is the electrical potential

n is the electron density

p is the hole density

N_D^+ is the donor density

N_A^- is the acceptor density

In Nano-MOS, we assume the absence of holes and only treat electrons. Thus,

Poisson's equation becomes

$$\nabla^2 = -\frac{1}{\epsilon}(-n + N_D^+ - N_A^-) \qquad (3)$$

3. METHODOLOGY

In this paper, online Nano electronic device simulation software, Nano-MOS is used to study the variation effect of silicon film thickness at nano regime, where by five different DG-Nano-MOS with thickness of silicon film channel 1.5, 2.5, 3.5, 4.5 and 5.5nm are simulated respectively with other structural dimensions fixed. The double gate Nano-MOS are simulated using nanoMOS simulation software. Figure 2.0. shows Si double gated MOSFET.

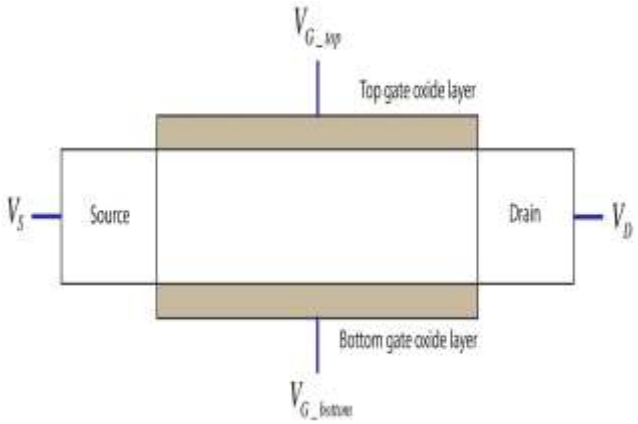


Figure 2.0: Ideal double-gated MOSFET structure

3.1. Nano-MOS Software

The Nano-MOS is a 2D simulator for thin body, fully depleted double gate n-MOSFETs. A choice of three transport models is currently available (drift diffusion, classical ballistic and quantum ballistic). The transport models treat quantum effects in the confinement direction exactly and the names indicate the technique used to account for carrier transport along the channel. Each of these transport models is solved self-consistently with Poisson's equation. Several internal quantities such as sub-band profiles, sub-band areal electron densities, potential profiles and I – V information can be obtained from the source code.

3.2. SIMULATION PROCEDURE:

- The device modelling was done by choosing device type, that is double gate MOSFET.
- Then transport and bias are set by choosing ballistic transport using Green's function approach.
- Then device description such as, source and drain doping concentration, source and drain length, top and bottom gate length etc was also set.
- Then silicon film channel thickness was varied from 1.5nm to 5.5nm.
- The program is then run to obtain results.
- However, steps 2 to 5 was repeated with ballistic transport using semi-classical approach and lastly with drift diffusion transport.

Table 1.0: Input Parameters and their Values

INPUT PARAMETERS	VALUE
Source/drain doping concentration	1e20/cm^3
Source/drain overlap	0
source/drain length	7.5nm
source/drain potential fixed	0.6V
Channel length	10nm
Top/ bottom insulator thickness	1.5nm
Top/bottom gate length	10nm
Temperature	300K
Gate voltage step size	0.05V
Silicon film channel thickness	1.5nm – 5.5nm

4. RESULTS AND DISCUSSION

4.1. Simulations Result of 2D Electron Density of Sub-bands along Channel

4.1.1. SIMULATIONS RESULT AT 1.5nm Tsi

The plot of 2D electron density of sub-bands along channel shown in figure 3.0 shows that the three transport models curve have roughly the same outline, except that there are discrepancies in the magnitude of 2D electron density of the sub-bands in the drain reservoir, channel and source reservoir region. The distribution of electron density was almost the same for ballistic transport using Green's function and semi-classical approach, whereas the drift diffusion model has more electron on the channel region due to scattering mechanism.

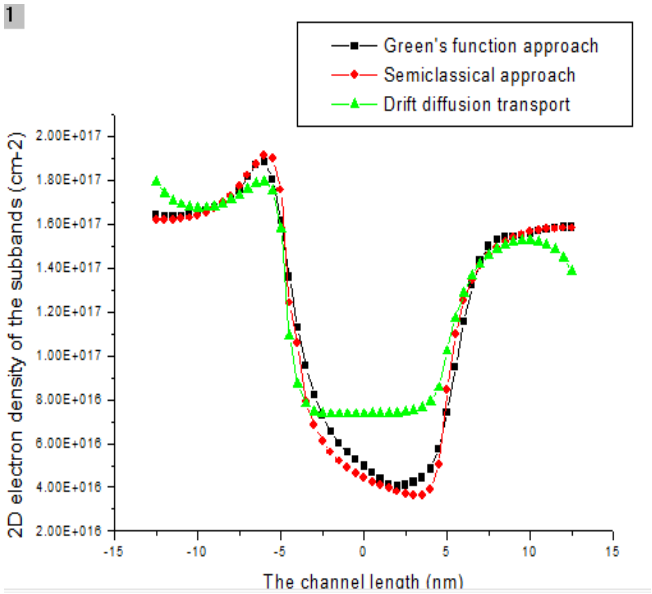


Figure. 3.0: 2D electron density of the sub-bands along channel

4.1.2. SIMULATIONS RESULT AT 2.5nm Tsi

The plot of 2D electron density of sub-bands along channel at 2.5nm Si film thickness shows that the distribution of electron density was almost the same for ballistic transport using Green's function and semi-classical approach while drift diffusion transport has more electron on the channel region and is more than that of 1.5nm.

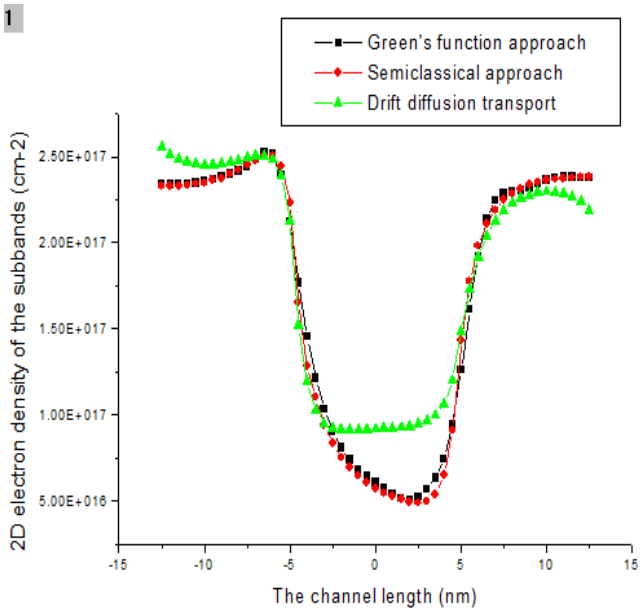


Figure. 4.0 2D electron density of the sub-bands along channel

4.1.3. SIMULATIONS RESULT AT 3.5nm Tsi

The plot of 2D electron density of sub-bands along channel at Si film thickness of 3.5nm shows that the distribution of electron density of Green's function and semi-classical approach are almost the same while drift diffusion has more electron at channel region.

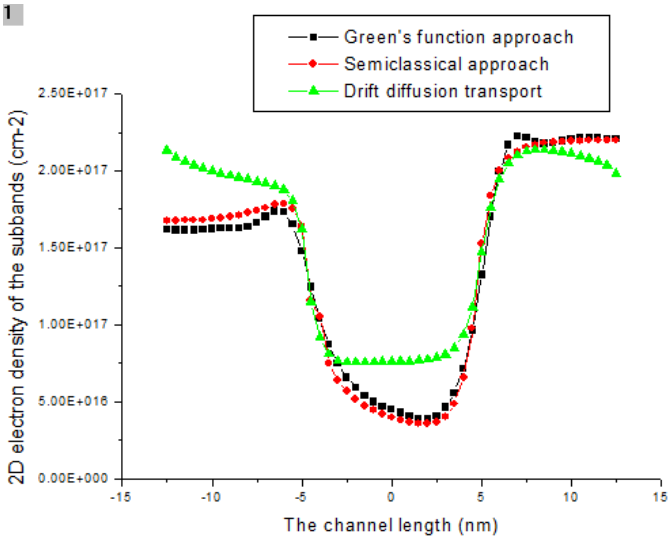


Figure. 5.0: 2D electron density of the sub-bands along channel

4.1.4. SIMULATIONS RESULT AT 4.5nm Tsi

The plot of 2D electron density of subbands along channel at 4.5nm Si film thickness shows that the distribution of electron density of Green's function and semi-classical approach are almost the same for while drift diffusion transport has more electron at channel region

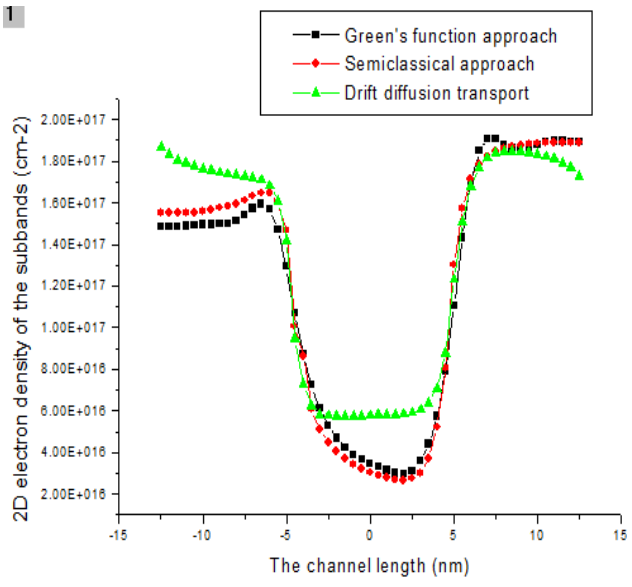


Figure. 6.0: 2D electron density of the sub-bands along channel

4.1.5. SIMULATIONS RESULT AT 5.5nm Tsi

The plot of 2D electron density of sub-bands along channel at Si film thickness of 5.5nm also shows that the distribution of electron density of Green's function and semi-classical approach are almost the same while drift diffusion transport has more electrons at channel region. Generally, the magnitude of 2D electron density of the sub-bands increases as the thickness of Si film increased, since there are more number of electron quantity.

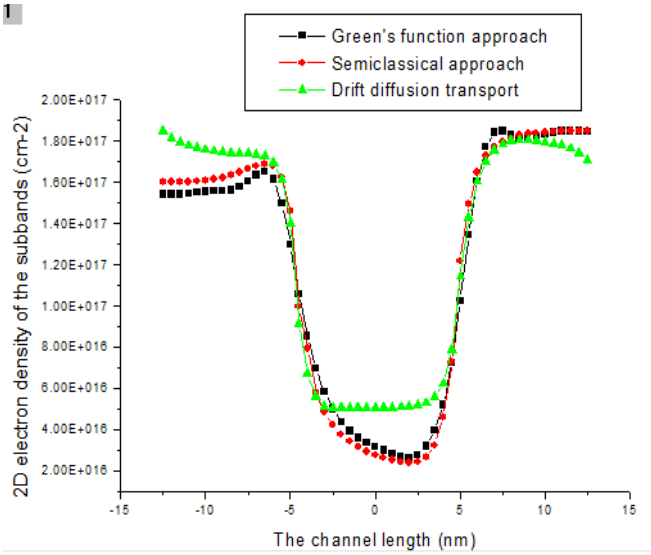


Figure. 7.0: 2D electron density of the sub-bands along channel

4.2. SIMULATIONS RESULT OF SUBBANDS ENERGY PROFILE ALONG CHANNEL

4.2.1. SIMULATIONS RESULT AT 1.5nm Tsi

The plot of sub-bands energy profile along the channel for Si film thickness of 1.5nm shows that Green's function and semi-classical approach have roughly the same potential barrier. On the other hand, drift diffusion transport shows a moderate higher potential barrier.

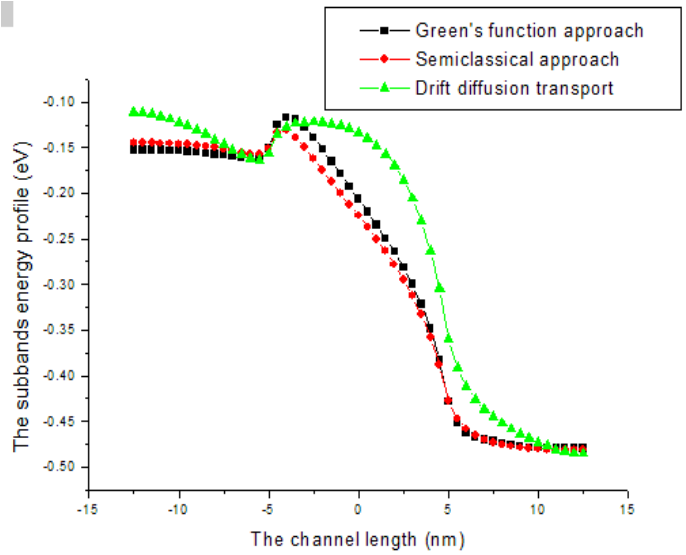


Figure. 8.0: sub-bands energy profile along channel

4.2.2. SIMULATIONS RESULT AT 2.5nm Tsi

The plot of subbands energy profile along the channel for Si film thickness of 2.5nm also shows that Green's function and semi-classical approach have almost the same potential barrier while drift diffusion transport shows a moderate higher potential barrier and is higher than that of 1.5nm

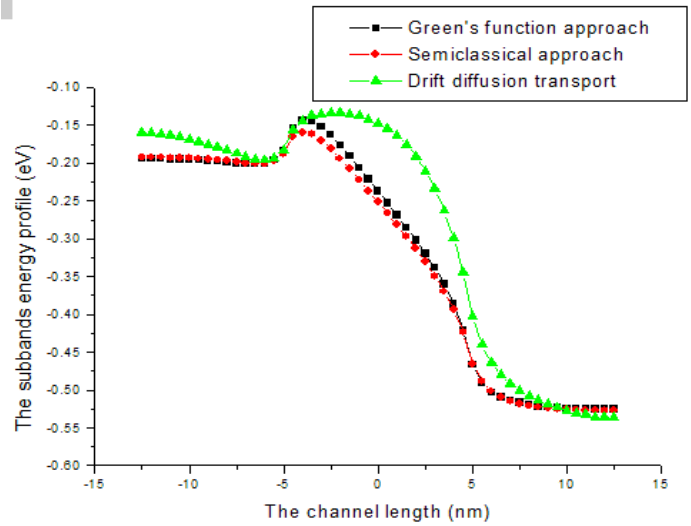


Figure. 9.0: sub-bands energy profile along channel

4.2.3. SIMULATIONS RESULT AT 3.5nm Tsi

The plot at 3.5nm Si film thickness shows that Green's function and semi-classical approach have almost the same potential barrier and is higher than that of 2.5nm. On the other hand, drift diffusion transport shows a moderate higher potential barrier.

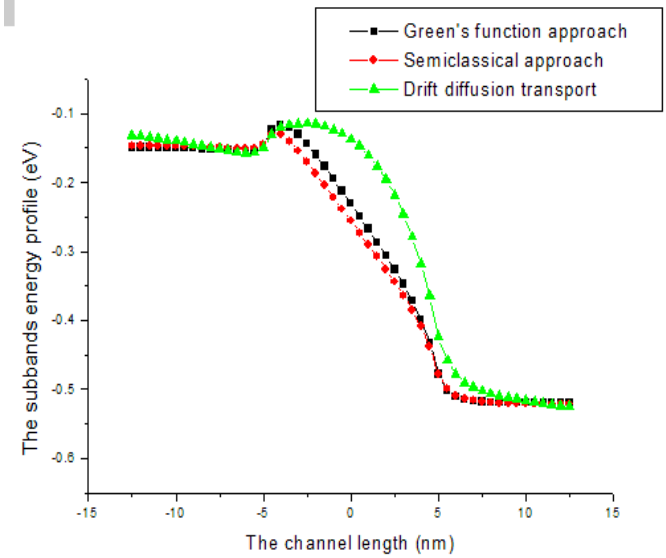


Figure. 10.0: sub-bands energy profile along channel

4.2.4. SIMULATIONS RESULT AT 4.5nm Tsi

The plot at 4.5nm Si film thickness shows that Green's function and semi-classical approach have roughly the same potential barrier and drift diffusion transport shows a moderate higher potential barrier and is more than that of 3.5nm.

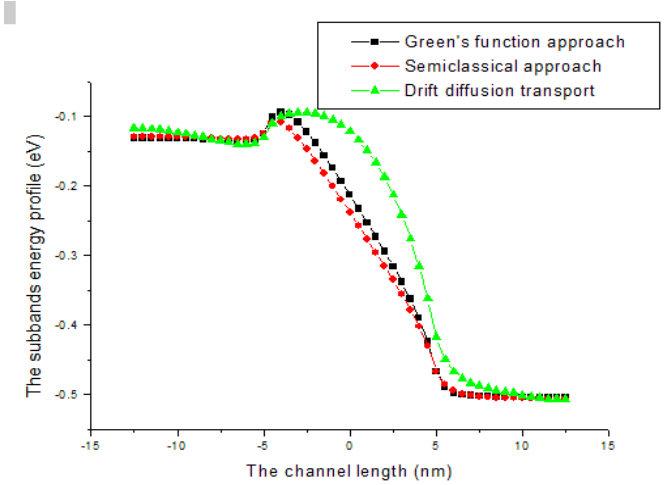


Figure. 11.0: sub-bands energy profile along channel

4.2.5. SIMULATIONS RESULT AT 5.5nm Tsi

The plot at 5.5nm Si film thickness shows that Green's function and semi classical approach have almost the same potential barrier and is higher than that of 4.5nm. On the other hand, drift diffusion transport shows a moderate higher potential barrier and also is higher than that of 4. 5nm.Therefore as thickness of Si film increased, the potential barriers for all 3 transport models increased

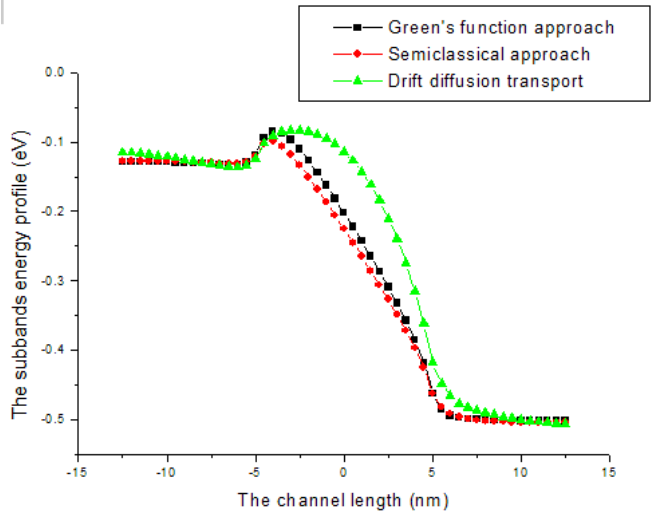


Figure. 12.0: sub-bands energy profile along channel

4.3. SIMULATIONS RESULT OF DRAIN CURRENT AGAINST GATE VOLTAGE

4.3.1. SIMULATIONS RESULT AT 1.5nm Tsi

The plot of drain current versus gate voltage for Si film thickness of 1.5nm shows that Green's function and semi-classical approach have almost the same current at off state as well as on state. On the other hand, drift diffusion transport has the lowest drain current due to scattering mechanism.

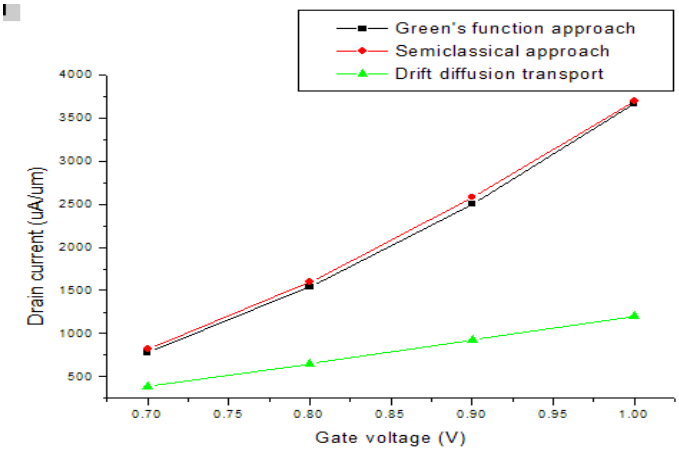


Figure: 13.0: Drain current versus gate voltage

4.3.2. SIMULATIONS RESULT AT 2.5nm Tsi

The plot drain current versus gate voltage for Si film thickness of 2.5nm shows that Green's function and semi-classical approach have almost similar current because are ballistic in nature, while drift diffusion has lowest current due to scattering. However, as thickness of Si film is increased, the peak on state current is also increased for all models.

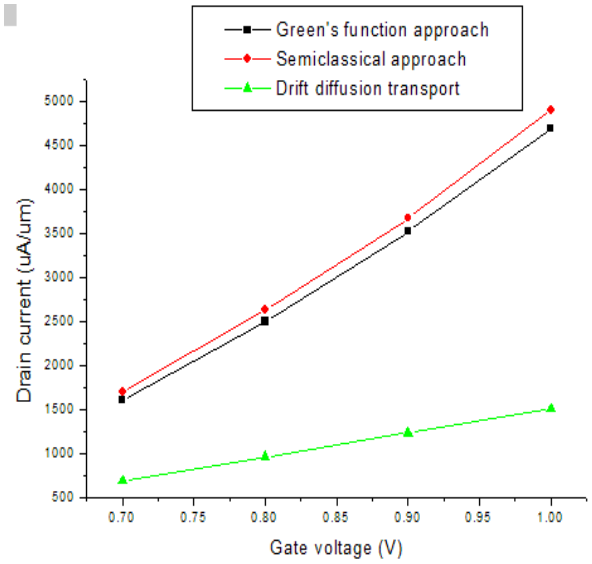


Figure: 14.0: Drain current versus gate voltage

4.3.3. SIMULATIONS RESULT AT 3.5nm Tsi

The plot of drain current versus gate voltage for Si film thickness of 3.5nm shows that Green's function and semi-classical approach have the highest drain current and almost the same, because are ballistic in nature and drift diffusion transport has the lowest drain current because of scattering. However, as thickness of Si film increased, the peak on state current also increased as shown from the plot.

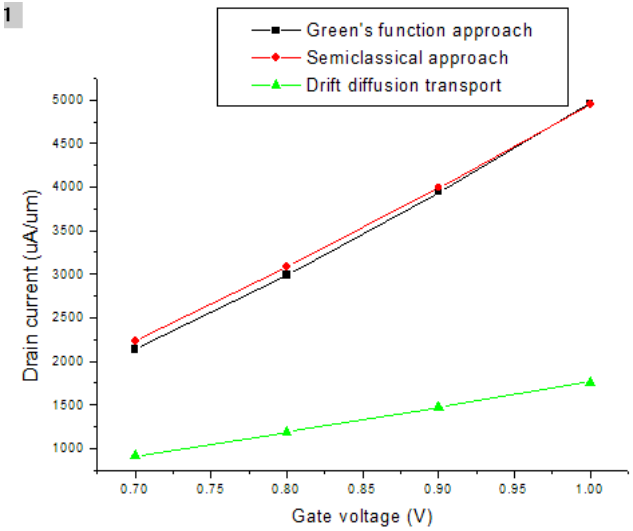


Figure: 15.0: Drain current versus gate voltage

4.3.4. SIMULATIONS RESULT AT 4.5nm Tsi

The plot of drain current versus gate voltage for Si film thickness of 4.5nm shows that Green's function and semi-classical approach have highest drain current and is higher than that of 3.5nm, also drift diffusion transport has lowest drain current and is also higher than that of 3.5nm. Therefore, as thickness of Si film increased, drain current and peak on state current increased.

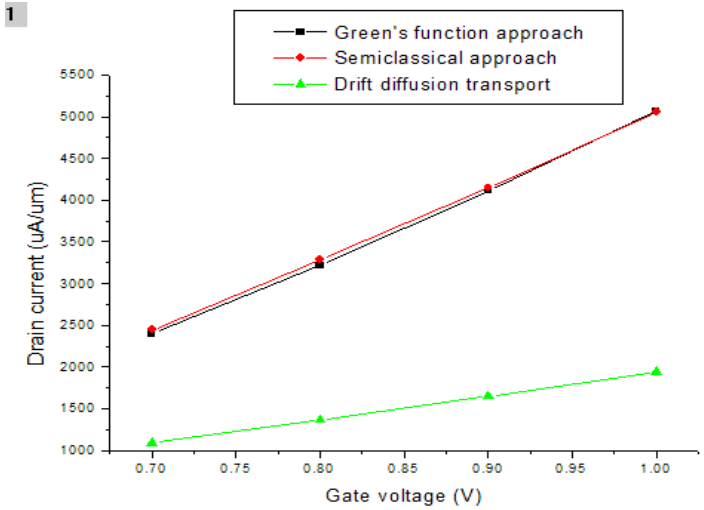


Figure: 16.0: Drain current versus gate voltage

4.3.5. SIMULATIONS RESULT AT 5.5nm Tsi

The plot of drain current versus gate voltage for Si film thickness of 5.5nm shows that Green's function and semi-classical approach have highest drain current compared to 1.5nm, 2.5nm, 3.5nm and 4.5nm. Also, drift diffusion transport has highest drain current compared to 1.5nm, 2.5nm, 3.5nm and 4.5nm. However, the peak on state current is higher than that of 1.5nm, 2.5nm, 3.5nm and 4.5nm.

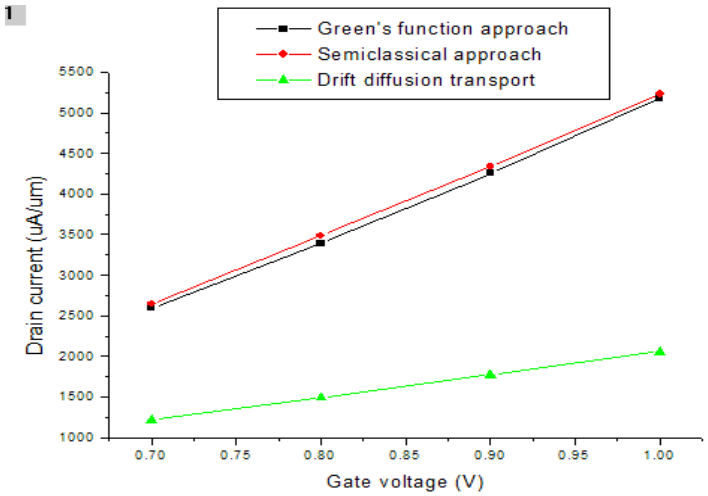


Figure: 17.0: Drain current versus gate voltage

5. CONCLUSION

In this work three transport model studied, from which it was observed that, Green's function and semi-classical approach produced roughly the

same characteristics, since both of them deal with ballistic transport. However, drift diffusion transport has the lowest performance due to existence of scattering. When all other structural design of nanomos are fixed, increment in Si film thickness result in more number of electron in the channel region. Thus, it is observed that thick-body nanomos can perform better than thin-body nanomos.

REFERENCES

1. Vinod, K. K.: Physics of carrier-transport mechanisms and ultra-small scale phenomena for theoretical modeling of nanometer MOS transistors from diffusive to ballistic regimes of operation. Physics Reports, vol.398, pp.67-131. (2004).
2. Haiyan, J., Sihong, S., Wei, C., Pingwen, Z.: Boundary treatments in non-equilibrium Green's function (NEGF) methods for quantum transport in nano-MOSFETs. Journal of Computational Physics, vol.227, pp.6553-6573. (2008).
3. Moore, G.: Cramming more components onto integrated circuits. Electronics. Vol.38, pp.114. (1965).
4. Harsh, V.: Effects of Gate Length and oxide Thickness on DG MOSFET. International Research Journal of Engineering and Technology, Volume: 2, Issue:2. pp.71-78. (2015).
5. Galadanci, G.S.M., Tijjani, A., Gana S. M.: Performance analysis of electrical characteristic of single gate and double gate nano-MOSFET devices. American Journal of Engineering Research (AJER), vol. 7, no.6., 248 - 259. (2018).
6. James and Savas Kaya.: Accurate treatment of interface roughness in nanoscale DG MOSFETs using non-equilibrium Green's functions", Solid-State Electronics. Vol.48 pp.1843-1847. (2004)
7. Datta, S.: Lesson from nanoelectronics world scientific Pardue University U.S.A. (2012)
8. N. Ben Abdallah, M. Mouis and C. Negulescu,)."An accelerated algorithm for 2D simulations of the quantum ballistic transport in nanoscale MOSFETs", Journal of Computational Physics. vol.225 pp.74-99. (2007)

International Journal of New Computer Architectures and their Applications (IJNCAA) 8(4): 206-213
The Society of Digital Information and Wireless Communications, 2018 ISSN 2220-9085 (Online); ISSN 2412-3587 (Print)

Forecasting of Traffic Load for 3G Networks Using Conventional Technique

G.S.M. Galadanci, S.B. Abdullahi and Z.A. Bature
Department of Physics, Bayero University Kano (BUK)
PMB 3011, Kano 700241, Kano-Nigeria
ggaladanci.phy@buk.edu.ng, sba_elt.bch@outlook.com

ABSTRACT
The degrading performance of network coverage, resource allocation, and utilization is due to the rapidly increasing number of cellular subscribers, which is immensely difficult to predict the traffic load in Nigeria. The available developed algorithms and models did not well consider the behavior of the traffic load using the adopted input variables of this research. This paper arguably constructs an Artificial Neural Network (ANN) as Conventional technique to forecast the instantaneous traffic load per cell or eNodeB of 3G networks in Kano Metropolis. Four Active 3G networks data was extracted and recorded with the aid of W995 TEMS Pocket Phone over thirty five cells. The forecasted models when tested apparently tracked the measured traffic load with RMSE of 0.148365%, 0.21878%, 0.3327% and 1.32220%, thus achieved MAPE of 0.00394%, 0.00696%, 0.00109% and 0.03978% for A, B, C and D networks respectively. These validated that the Conventional technique can be a valuable tool in forecasting traffic load in Nigeria and could also be adopted in forecasting of large-scale metropolis cellular networks.

KEYWORDS: Artificial Neural Network (ANN), traffic load, RSCP, Path loss, SIR.

1 INTRODUCTION

One of the most significant revolutions of mobile communications is that subscribers can use their mobile phones to access the subscribed services as long as they are in the operator's coverage area [1]. However, this revolution results in an exponential increase of service demands. The multifarious service demands brought up research development in areas of power control, handover procedures, frequency hopping, discontinuous transmission, spectrum efficiency, multiple access technology, cluster size and frequency re-use [2]. These techniques were developed aimed to increase the number of subscribers that can access the limited transmission channels in the mobile communication networks [3]. Therefore leads to development of different cellular standards from the second generation (2G) to third generation (3G) and now beyond, but still the subscribers' voice and data traffic demands are rapidly increasing. In order to reduce the pressure of circuit switched networks (2G), high switched packet networks (3G) are introduced to provide higher quality signal with balance traffic load to subscribers. Universal Mobile Telecommunication System (known as UMTS) is a third generation (3G) telecommunications technology. The most common form of UMTS makes use of WCDMA (Wideband Code Division Multiple Access, which is an air interface standard and compulsory feature of any mobile telecommunications device of the 3G network) and this is the type of 3G technology employed in Nigeria. The transmitting (Tx) and receiving (Rx) Frequencies, measures in Mega Hertz (MHz) of the four active 3G mobile networks in Kano as well as Nigeria in general were given in table 1 below.

Table 1: 3G Band Plan [4]

Networks	Tx Frequency (MHz)	Rx Frequency (MHz)
A	2140-2150	1950-1960
B	2110-2120	1920-1930
C	2120-2130	1930-1940
D	2130-2140	1940-1950

The four active 3G networks in Kano have voice connection of more than 7, 811, 290 subscribers. Therefore the subscribers of these networks are very aggressive with the abysmal services receiving from the operators. One of the core concepts to solve the conflictions and inherent problems between multifarious service demands and limited radio resource is to increase the

number of eNodeBs in order to get higher total capacity. But the major issues of employing more eNodeBs into the networks were among others expensive and maintenance/installation cost, thus necessitated the need of having a very good tool in forecasting the behavior of traffic load. The major challenge of forecasting traffic load in cellular network is the non-linearity behavior of the network parameters. However, as bandwidth is cheap forecasting in networking often relies on simple techniques and/or over provisioning. A precise forecast of traffic loads in the 3G network is an essential task for a network carrier, as the upgrade path for network equipment needs two to four month time to be implemented [5]. This work established that traffic load in Kano is much lower than reported from the Nigeria Communication Commission (NCC), likely due to more expensive 3G infrastructure deployment plans in Nigeria. Traffic load is moderately correlated between the nodes belonging to the eNodeB. Finally the work also contributed in the construction of ANN linguistic models and when tested apparently tracked the measured traffic load, which could be adopted as a valuable tool in forecasting traffic load in 3G cellular network.

In recent times several research works were conducted on traffic load some based on statistical approaches, mathematical modellings, adaptive neuro fuzzy interference system (ANFIS), algorithm adoption, and fuzzy logic, as in [6] presented statistical methods, the random and the "best carrier" were analyzed and the results showed significant similarities in certain conditions subscribers per cell of SINR. Traffic cell analysis was carried out, traffic characteristic was determined and also traffic performance parameters were depicted in [2]. Similarly, [3] presents different methods to predict the load increase in a 3G live network, the DHR delivers a better MAE (Mean Absolute Error) performance. [1] Statistically modeled the downlink throughput per cell distributions over time and over different cells of an existing 3G network based on real network throughput data. Likewise [7] collected data from measurements on live 3G networks and compared the measured cell loads and number of users with the values estimated by the models. [8] Presented the design and implementation of a Fuzzy Logic multi-criteria handoff algorithm based on signal strength, path-loss and traffic load of base stations and the received signal to interference ratio as to balance traffic in all the neighboring sites at any time. The limitations of intelligent methods include the appropriate selection of membership function and choice of structure among others [9]. Therefore, to the best of our knowledge, up to now, this is the first paper dealing with the forecast of traffic load in a 3G cellular network using ANN.

The paper is organized in the following: Section 1 presents the nature of the problem, purpose, contributions and the previous literature. The methodology, measurement setup and the forecasting method formed section 2. The results and discussion is covered in section 3 and also validation of the models is also cover in this section. Finally section 4 formed the conclusion.

2 MATERIALS AND METHODS

2.1 Typical 3G network

Typical 3G network comprised of a Radio Access Network (RAN) and a core network (CN). Radio Network Controllers (RNCs) control multiple Base Stations (NodeBs), which connects with multiple User Equipments (UEs). Each NodeB is basically configured with multiple sectors (commonly 3 and up to 6 sectors) in different directions. The core network consists of Serving GPRS Support Nodes (SGSN) and Gateway GPRS Support Nodes (GGSN) to perform data access and charging functionality. The combination of HSDPA and HSUPA is called High Speed Packet Access (HSPA) [10], the network structure is depicted in Figure 1.

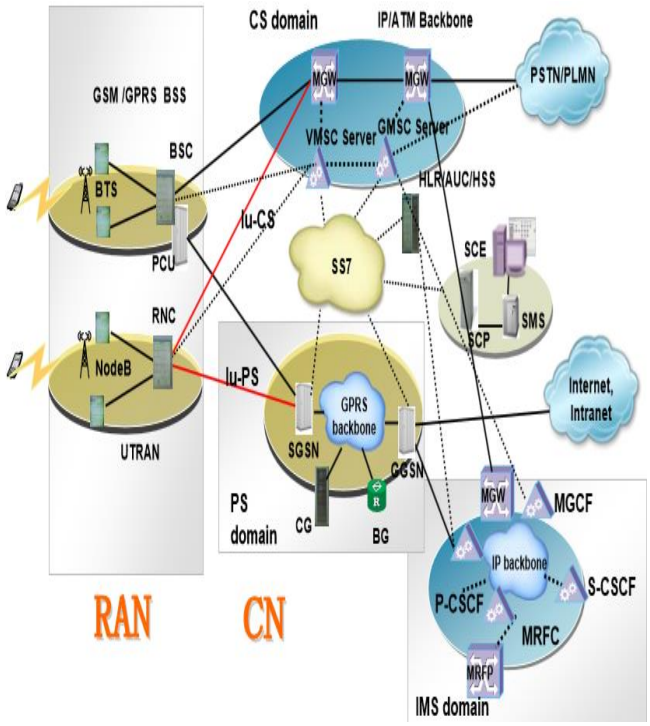


Figure 1: The 3G WCDMA network architecture [11]

2.2 Definitions of Important Parameters

a) RSCP (Received Signal Code Power): This is the downlink power/signal strength received by the UE on the pilot channel and measured in dBm. A higher RSCP indicates a better channel [12].

b) Transmitting Power (Tx power): This is the performance metric used to measure the transmitting ability of a cell/nodeB [12].

c) Path Loss: The path loss is the difference (dB) between the transmitted power and the received power. It represents signal level attenuation caused by free space propagation, reflection, diffraction and scattering [13]. Total path loss increases only substantially and appreciably with an increase in path-length, foliage distance, and reduction in transmitted frequency. Also occurs even when there are no obstacles between the transmitting and receiving antenna [14].

d) Signal-to-Interference Ratio (SIR): This is defined as the power of a certain signal of interest divided by the sum of the interference power (from all the other interfering signals) and the power of some background noise. Received Total Wideband Power (RTWP) measures the total level of noise within the 3G frequency band of any cell and captures uplink interference. A lower RTWP indicates a better channel and unloaded network [10].

e) Traffic Load: Traffic is the minutes of calls in Erlang [14]. Therefore Traffic is determined from the number or volume of calls intensity and service time (mean holding time, MHT) [2]. As in [2] traffic can be categorized into offered traffic, carried traffic and block traffic. Thus [2],

Offered traffic = carried traffic + block traffic (1)

The following formula could be adopted by the RF engineers to calculate the number of calls or Erlangs [13]:

$$ErlangB = \frac{Established\ callsX\ MHT}{3600} \quad (2)$$

2.3 Experimental Setup

The experimental setup is represented using flow chart as shown in figure 2. Step 1: The four carriers in consideration were examined using the same technique and at the same locality of thirty five cells/NodeBs with the aid of TEMS Pocket. TEMS Pocket is an advanced cellular network diagnostics tool for both indoor and outdoor measurement built into a Sony Ericsson W995 phone. TEMS Pocket is suitable for day-to-day verification, maintenance and troubleshooting of cellular networks but is also handy for many cell planning tasks. Some further key features of these devices are: GSM/GPRS quad-band, WCDMA/HSPA , 8.1 megapixel camera, Integrated GPS and in-built scanner [15]. Step 2: The data was collected for a period of 16 weeks (4 months), various times a day at an interval time of different hours. Informations such as RSCP, Tx power, pathloss, SIR, traffic load, etc were obtained and recorded for both intra and inter calls on a research's designed form. Step 3: the artificial neural network was constructed with the above informations. Step 4: here, the voice traffic load was predicted and finally Step 5: the models were evaluated using error criteria.

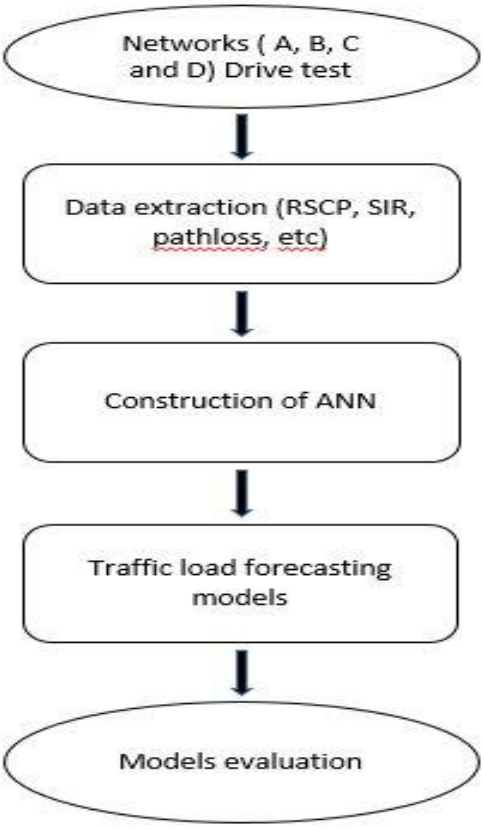


Figure.1: The experimental Setup flow chart

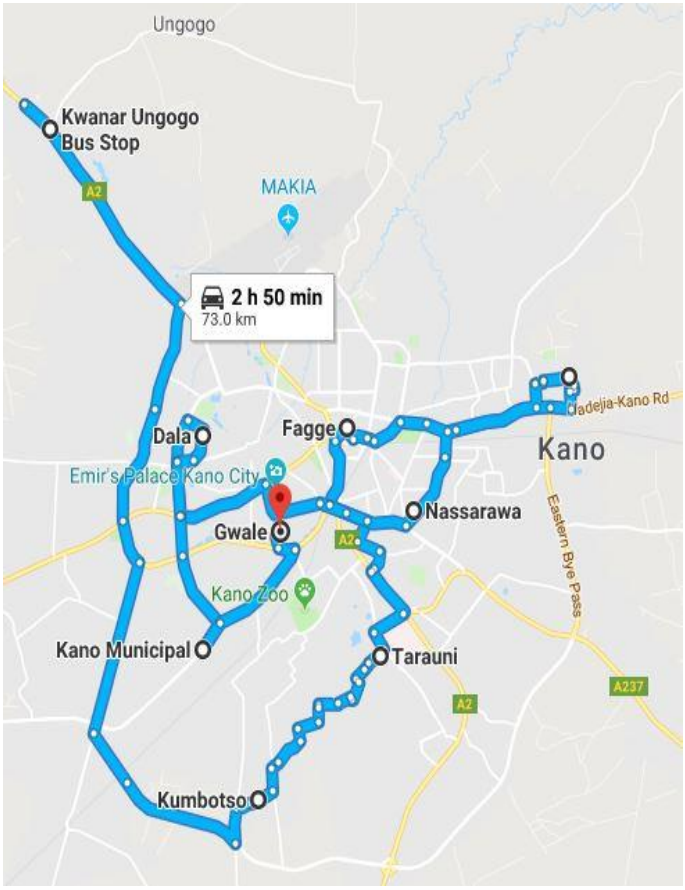


Figure 3: Research location

2.3 Study Location

Kano metropolis, Kano Nigeria is situated between latitudes 11° 25'N to 12° 47'N and longitude 8° 22' E to 8° 39' E east and 472m above sea level. Kano metropolis is boarded by Madobi and Tofa Local Government Areas (LGAs) to the southwest, Gezawa LGA to the east, Dawakin Kudu LGA to the southeast, and Minjibir LGA on the northeast. The study area of this research is made up of eight local government areas (LGAs): Dala, Fagge, Gwale, Kano Municipal, Nassarawa, Tarauni and parts of Kumbotso and Ungogo Local Governments. This location was chosen based on the fact that the population of the people in the area continue to grow substantially due to education, marketing and trading has been the dominant economic activity of the populace of the metropolitan Kano that is why it is referred as the Centre of Commerce in the country due to long flourished marketing activities. Kano metropolis is the third largest town in Nigeria after Lagos and Ibadan. It has a population of 2,826307 people [15].

2.4 Methodology

An artificial neural network (ANN) is a computational model based on the structure and functions of natural neurons [16]. The technique of adjusting the weights is called training or learning. ANN is considered nonlinear statistical data modeling tools where the complex relationships between independents and dependents are mapped or modeled. The ANN considered here is used for engineering purposes, such as forecasting, data compression and pattern recognition. ANN has several advantages but one of the most recognized of these is the fact that it can actually learn from observing data sets. This gives it universal acceptability as a random function approximation tool. ANNs have three layers that are interconnected as shown in Figure 3. The first layer consists of input neurons; $x_1, x_2, \ldots x_n$. Those neurons send data on to the second layer (weights); $y_1, y_2, \ldots y_n$. Which in turn sends the output neurons to the third layer; z [16]. Training an artificial neural network involves choosing from allowed models for which there are several associated algorithms [17].

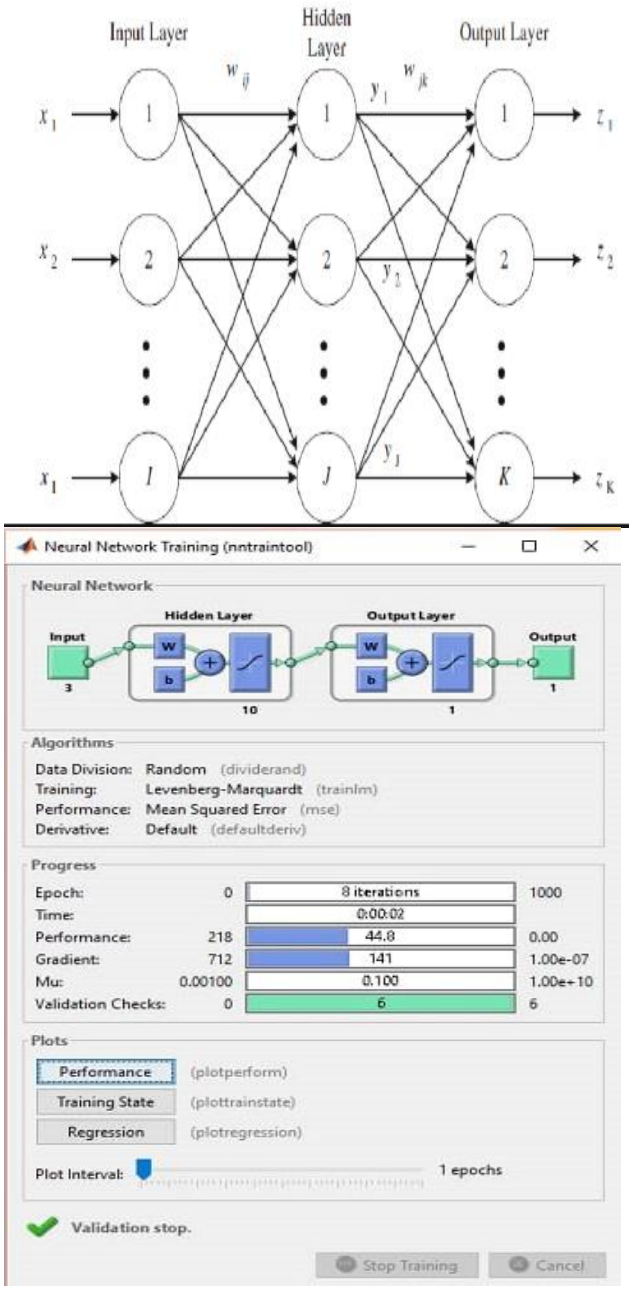


Figure 4: ANN Structure and it workspace

The data was extracted with the aid of W995 Transmission Environmental Monitoring System (TEMS) phone which recorded the RSCP, SIR, Pathloss, MHT, Established Calls, Hand over and traffic load files respectively. These data were refined and converted into an Excel format due to the likely unavoidable errors, outliers and missing values. The data was embedded into the ANN which is considered here as a conventional technique. The conventional technique model was developed using MATLAB (R2013a) with RSCP, SIR and pathloss as input data and the traffic load at enodeBs as the target output. Each data set consisted of three input columns and one output column. The construction was set as depicted in fig. 4 using 1 input layer with 3 inputs data, two hidden layers and one output layer. The first hidden layer consisted of 10 neurons, second consisted of 20 and the output consisted of 1 neuron layer, the output of the target is the eNodeBs' traffic load of the networks. The sampling interval is 1, completion time is 16.25sec and 10 maximum number of iterations were reached (epoch is 10). Thus each model is trained with one data set. The input data is normalized from the value of -1 to 1, these data set is further divided into two sets: 90% of these data are fed into the training process at random, while 10% was selected for testing phase. The output values were obtained when the input values were inserted into the first hidden layer, thus the output of the first hidden layer was fed into the second hidden layer, and the output of second hidden layer was fed into the input of the output layer. The value obtained at the output layer is the value of the network output. The models training were controlled by the following conditions: maximum epoch, minimum error and early stopping criteria.

The developed models were statistically analyzed using Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) as equated in (3) and (4). These are to compare the performance of the forecasting techniques and to consider the effect of the magnitude of the actual values so as to judge the models.

$$RMSE = \sqrt{\frac{\sum(x_i - y_i)^2}{N}} \qquad (3)$$

$$MAPE = \frac{100\%}{N} * \sum_{i=1}^{N} \left| \frac{x_i - y_i}{x_i} \right| \qquad (4)$$

Where x_i is the measured value, y_i is the forecast value and N is the number of samples.

3 RESULTS AND DISCUSSION

Figure 5, 6, 7 and 8 demonstrated the performance of the models during training phase of the three input variables of RSCP, Pathloss and SIR, for the four active 3G networks respectively. While figure 9, 10, 11 and 12 depicted the performance of the four 3G networks input variables during testing phase. The performance of the traffic data when evaluated using Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) during training and testing phases of the A, B, C and D Networks respectively are presented in Table 2.

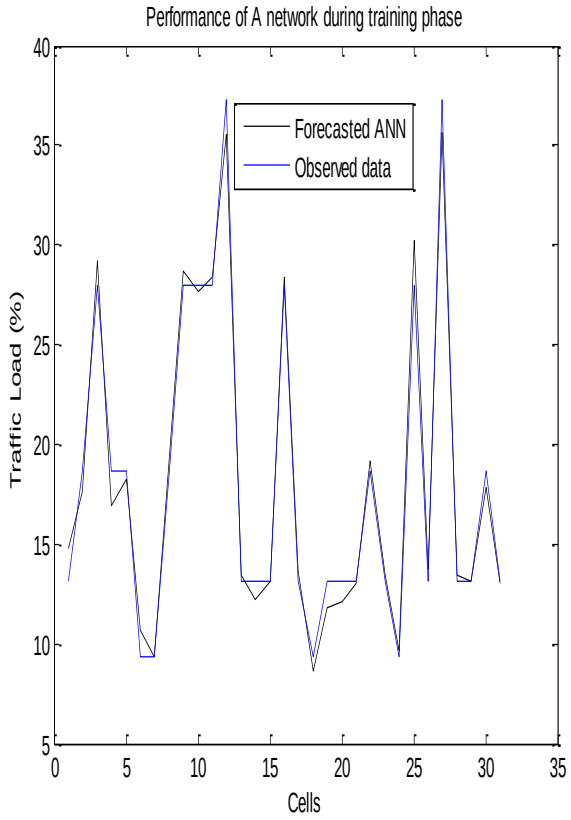


Figure 5: Performance of A Networks during Training

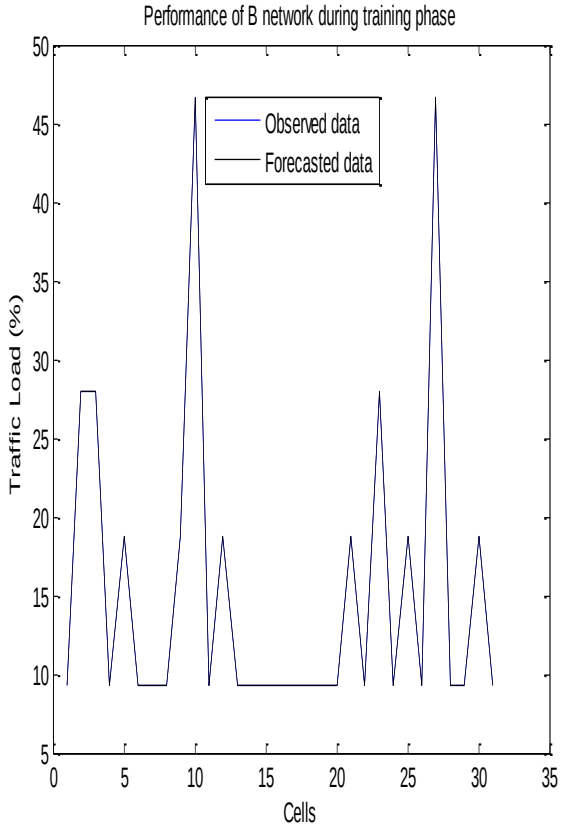


Figure 6: Performance of B network during training

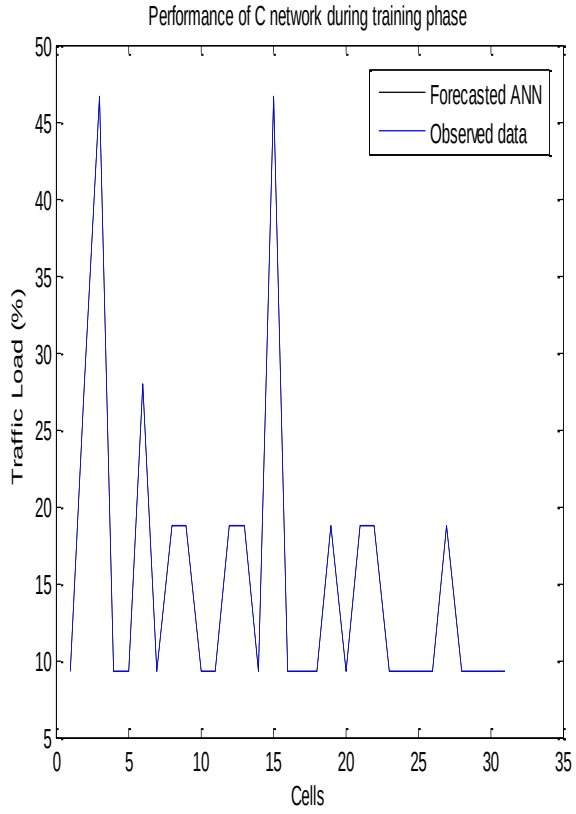


Figure 7: Performance of C network during training

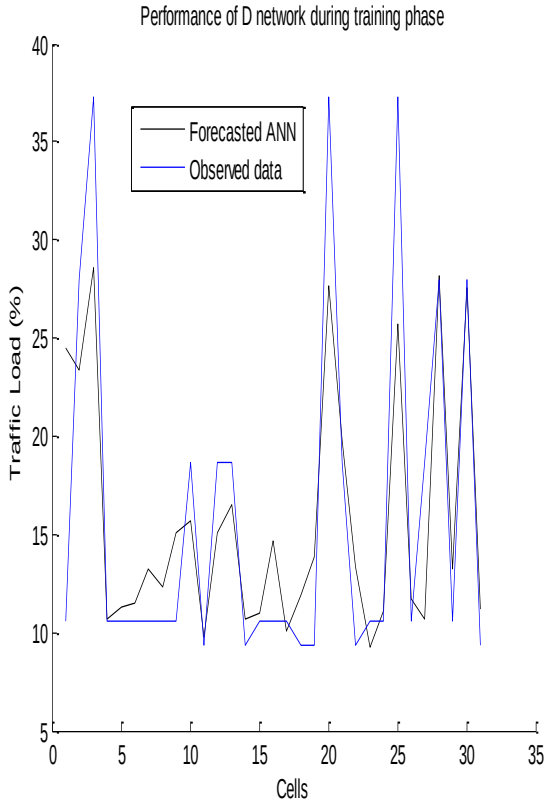


Figure 8: Performance of D network during training

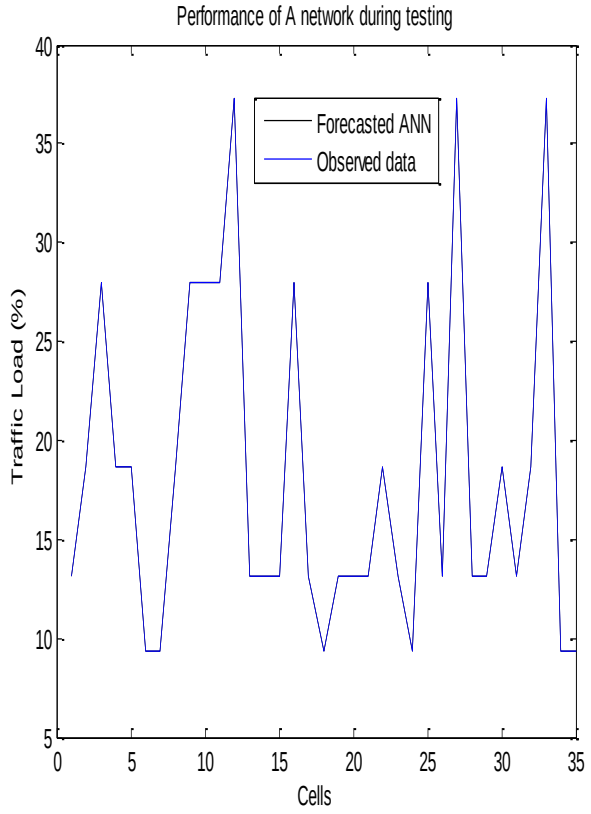


Figure 9: Performance of A network during testing phase

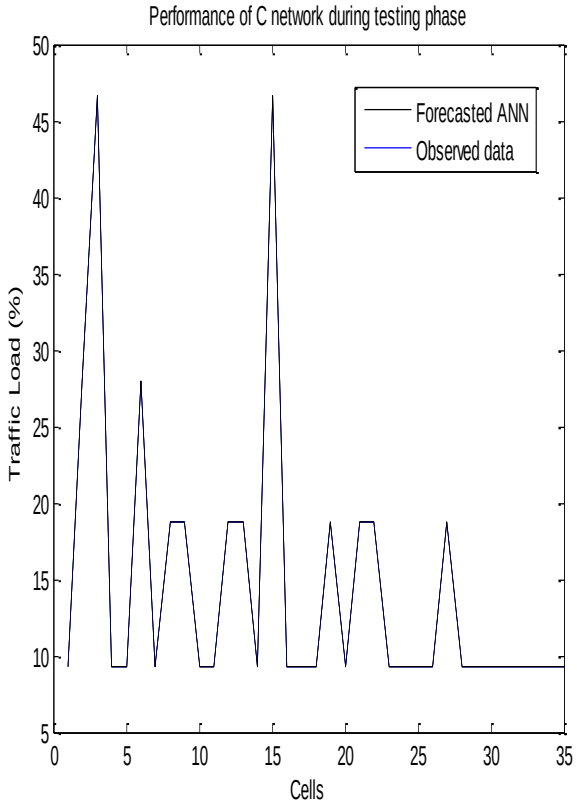


Figure 11: Performance of C network during testing phase

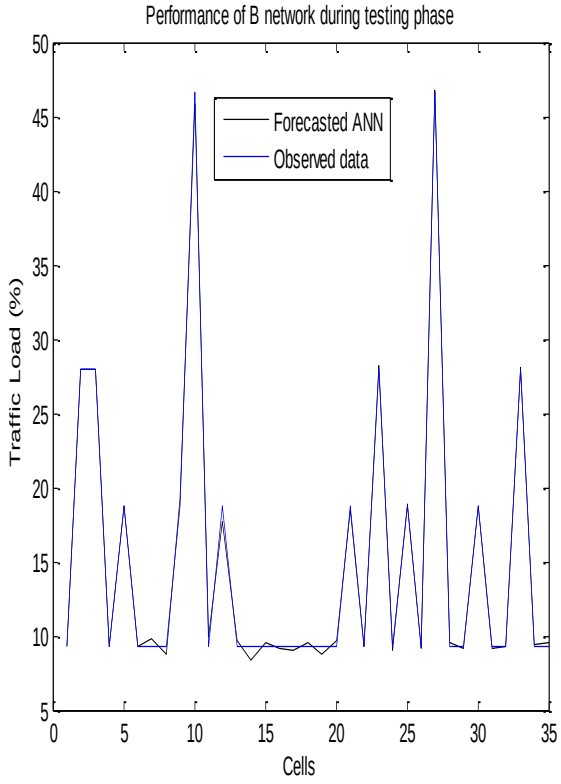


Figure 10: Performance of B network during testing phase

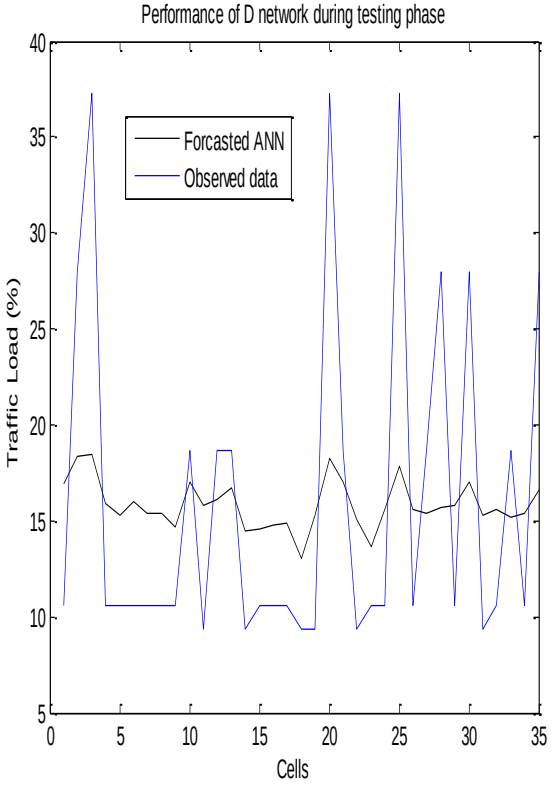


Figure 12: Performance of D network during testing phase

3.1 Validation of the Models

It is important to demonstrate how representative the models are for the simulations of networks with a given average traffic load and the models' ability to make representative of user estimations for different individual cell/eNodeB loads. Thus the forecasted models arguably demonstrated good fit to the traffic load when tested using equations (3) and (4) respectively. The result of the test is tabulated below in Table 2.

Table 2: Performance Evaluation of the Models

Networks	MAPE Training (%)	RMSE training (%)	MAPE Testing (%)	RMSE testing (%)
A	0.00437	0.13666	0.00394	0.14836
B	0.00132	0.03507	0.00696	0.21878
C	0.00065	0.01711	0.00109	0.03327
D	0.03240	0.89103	0.03978	1.32220

4. CONCLUSION

This paper predicted the traffic behaviors of four active 3G networks in Kano metropolis using Artificial Neural Networks as Conventional technique. The prediction error estimate is consistent at some certain levels. Whereby the traffic data for A, B and C networks when predicted were significantly accurate in tracking the observed data while D network did not track the observed data exactly but still is consistent with less than 2% error for both MAPE and RSME which is strongly recommended. The results proved that the conventional technique models could be served as valuable tools in forecasting voice traffic load in 3G networks at the case study location or any environment with similar network deployment.

REFERENCES

1. Nan, E., Chu, X., Guo, W. and Zhang, J.: User Data Traffic Analysis for 3G Cellular Networks. In: International Conference on Communications and Networking in China (CHINACOM), Beijing, pp. 468--472. Beijing, China (2013).
2. Osahenvemwen, O.A. and Emagbatere, J.: Determination of Traffic Load and Traffic Performance Parameters in Mobile Communication in Nigeria, Research Journal of Applied Sciences, Engineering and Technology, 4(11), pp. 1432—1437 (2012).
3. Chihani, B., Bertini, E., Jeanne, F. and Crespi, N..: HEP: Context-aware Communication System. International Journal of New Computer Architecture and Applications, 1(1), pp. 15--24 (2011).
4. NCC. *2100MHz—3G Band Plan*. Retrieved from www.ncc.gov.ng
5. Svoboda, P., Buerger, M. and Rupp, M.: Forecasting of Traffic Load in a Live 3G Packet Switched Core Networks. In: Proc. CSNDSD08 Conference, pp. 433--437, Vienna, Austria (2015).
6. Curpen, R., Illiescu, M. and Sandu, F.: An Analysis of 3G-4G Traffic Guidance Methods, Review of the Airforce Academy, 2(32), pp. 112—116 (2016).
7. Wikell, A.: Evaluation of Statistical Models in 3G and 4G Networks, Degree Project Research, KTH Datavuevenakap och Kommunikation, Stockholm, Sweden (2013).
8. Girma, S.T., Konditt, D.B.O., and Ndungu, K.E.N.: Fuzzy Logic Based Traffic Balancing in a GSM Networks, Journal of Research in Engineering, 1(2), pp. 63—74 (2014).
9. Muhammad, A., Gaya, M.S., Aliyu, R., Abdulkadir, R.A., Umar, I.D., Yusuf, L.A., Ali, M.U. and Khairi, M.T.M.: Forecasting of Global Solar Radiation using Anfis and Armax Techniques. In: Proc. of ICFMM2017, IOP Conference Series: Material Sceince and Engineering 303, pp. 1—6, England (2017).
10. Hu, Z., Chen. Y., Qiu, L., Xue, G. and Zhu, H.: An Indepth Analysis of 3G Traffic and Performance. In: Proc. of All Things Cellular '15, ACM, London, pp. 1-6, England (2015).
11. Huawei: WCDMA Handover Principle and Relevant Parameters," pp. 1–104, (2006).
12. L. A. Akinyemi, N. T. Makanjuola, O. O. Shoewu, and F. O. Edeko, "Evaluation and Analysis of 3G Network in Lagos Metropolis , Nigeria," vol. 2, no. 3, pp. 81–87, 2014
13. Chinedu, A.C. and Nkwachukwu, C.: Development of a Pathloss Model for 3G Networks at 1.857GHz in Port Harcourt Nigeria, International Journal Engineering Sciences and Research Technology, 6(5), pp. 423–433 (2017).
14. Boulmalf, M., Abrache, J., Aoum, T. and Haroud, H.: Traffic Analysis for GSM Networks. In: Proc. of the 9th International Conference on Computer Systems and Applications, IEEE/ACS, pp. 498—503, USA (2012).
15. Galadanci, G.S.M. and Abdullahi, S.B.: Performance Analysis of GSM Networks in Kano Metropolis of Nigeria, American Journal of Engineering Research, 7(5), pp. 69–79 (2018).
16. Ishak, W.H., Ku-Mahamud, K.R. and Norwawi, M.N.: Neural Network Applications in Reservoir Water Level Forecasting and Release Decision, International Journal of New Computer Architecture and Applications, 1(2), pp. 265—274 (2011).
17. Gershenson, C.: Artificial Neural Networks for Beginners, ResearchGate, (2003).