Princess Sumaya University for Technology
Since 1991 تأسست
PSUT
جامعة الأميرة سمية للتكنولوجيا

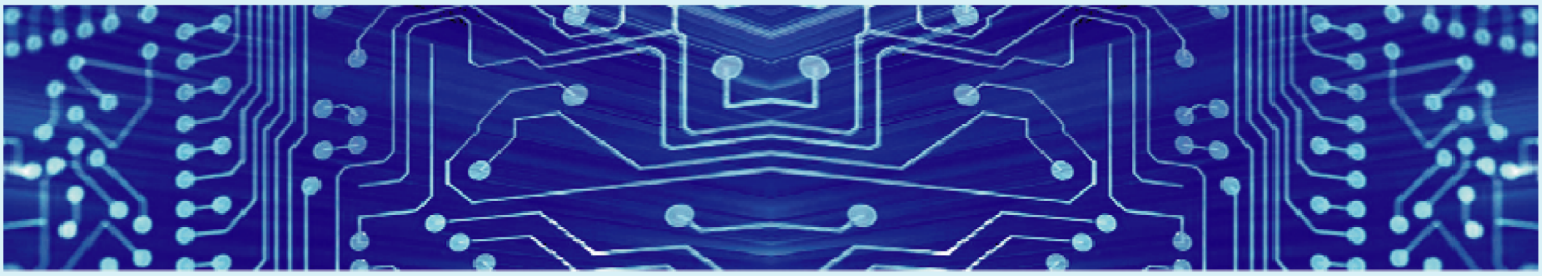
صندوق دعم البحث العلمي
Scientific Research Support Fund

Jordanian Journal of Computers and Information Technology

December 2016 VOLUME 02 NUMBER 03
ISSN 2415 - 1076 (Online)
ISSN 2413 - 9351 (Print)

JJCIT

www.jjcit.org jjcit@psut.edu.jo



An International Peer-Reviewed Scientific Journal
Financed by the Scientific Research Support Fund

Jordanian Journal of Computers and Information Technology (JJCIT)

The Jordanian Journal of Computers and Information Technology (JJCIT) is an international journal that publishes original, high-quality and cutting edge research papers on all aspects and technologies in ICT fields.

JJCIT is hosted by Princess Sumaya University for Technology (PSUT) and supported by the Scientific Research Support Fund in Jordan. Researchers have the right to read, print, distribute, search, download, copy or link to the full text of articles. JJCIT permits reproduction as long as the source is acknowledged.

Aims and Scope

The JJCIT aims to publish the most current developments in the form of original articles and review articles in all areas of Telecommunications, Computer Engineering and Information Technology and make them available to researchers worldwide.

The JJCIT focuses on topics including, but not limited to: Computer Engineering & Communication Networks, Computer Science & Information Systems and Information Technology and Applications.

Indexing

JJCIT is indexed in:

- ScopeMed:
  http://www.scopemed.org
- CrossRef:
  http://search.crossref.org/?q=jjcit
- OCLC WorldCat:
  http://www.worldcat.org/search?qt=worldcat_org_all&q=jjcit
- Scilit:
  http://www.scilit.net/journals/387088
- Impact Factor:
  http://www.scholarimpact.org/recently-indexed-journals/jordanian-journal-of-computers-and-information-technology
- DRJI (Directory of Research Journals Indexing):
  http://drji.org/JournalProfile.aspx?jid=2415-1076

EDITORIAL BOARD

INTERNATIONAL ADVISORY BOARD

"Opinions or views expressed in papers published in this journal are those of the author(s) and do not necessarily reflect those of the Editorial Board, the host university or the policy of the Scientific Research Support Fund".

"مـا ورد فـي هـذه المجلـة يعبـر عـن آراء البـاحثين ولا يعكـس بالضـرورة آراء هيئـة التحريـر أو الجامعـة أو سياسة صندوق دعم البحث العلمي".

GEOGRAPHIC GREEDY TRIPLEWISE GOSSIP ALGORITHM FOR WIRELESS SENSOR NETWORKS

Maha I. Raheem

Department of Computer Engineering, Baghdad University, Baghdad, Iraq
maha.i.raheem@ieee.org

(Received: 18-Jan.-2016, Revised: 01-Apr.-2016, Accepted: 17-Apr.-2016)

ABSTRACT

A novel gossip algorithm for distributed averaging with fast convergence and reduced cost of communication over wireless sensor networks (WSNs) is proposed in this paper. This algorithm is proved to improve the behaviour of the standard gossip algorithm (SGA), triplewise gossip algorithms (TGAs) and the geographic gossip algorithm (GGA) by exploiting the geographic information of the network. An analysis of convergence time and cost of communication of the proposed algorithm is performed and a comparison with other existing methods is provided.

KEYWORDS

Wireless Sensor Networks, Distributed Computing, Distributed Processing, Gossip Algorithms, Routing.

1. INTRODUCTION

Agreement/consensus of sensed information is one of issues of distributed signal processing in WSNs. Averaging the initial value of all the nodes in the network is an example of aggregate problem. Distributed averaging methods are widely used to solve agreement problems [1]. Gossip algorithms are widely used in distributed signal processing. Centralized computing, on the other hand, involves collecting data from all network nodes. In centralized computing, computations are performed at a fusion center. Distributed networks consume more power than their centralized counterparts do; the energy consumption depends on the number of radio transmissions and the total number of iterations until convergence. Distributed averaging algorithms have to be designed to avoid unnecessary waste of power and time. Among the advantages gained, gossip algorithms are robust against link failures and a communication bottleneck near the fusion center is avoided [2]. Sums and averages constitute building blocks for many signal processing applications, such as Gram-Schmidt orthogonalization [2]-[3], WSN node localization [4] and linear parameter estimation [5], to name just a few. Gossiping is a modified version of flooding, where the nodes do not broadcast a packet, but send it to a fully or not fully randomly selected neighbour/s. Gossiping avoids the problem of implosion of the network due to collision, but it takes a long time for message propagation throughout the network [1]. Though gossiping has considerably lower overhead than flooding, it does not guarantee that all nodes of the network will receive the message. It relies on the random neighbour selection to eventually propagate the message throughout the network. Gossip algorithms are employed to calculate the average of measurements of a WSN. In a typical pairwise gossip algorithm such as SGA [6]-[7], one node i wakes up at each iteration with probability $P=1/N$, where N is the number of sensor nodes, and performs averaging with one of its neighbors j at random with probability P_{ij}; iterations continue with slow convergence [1], [5]-[8]. SGA has another disadvantage in that it wastes a lot of energy among all gossip algorithms because of significant recalculation of redundant information. This result motivated Dimakis *et* *al.* [8] to modify the SGA by averaging with far away nodes resulting in the introduction of the so-called GGA. The latter algorithm accelerates the averaging process by averaging between any pairwise nodes in the whole network, exploiting the geographic information of activated nodes and their neighbors [1], [5], [8]-[9]. Triplewise gossip algorithms (TGAs), e.g. standard TGA and greedy TGA (G-TGA) accelerate the pair wise averaging methods (GGA and SGA) even further by averaging between three nodes per iteration instead of between only two nodes [10]. In standard TGA, at each iteration, one node wakes up and performs averaging with two of its neighbors at random. G-TGA has been proposed to reduce the time of convergence allowing the activated node to choose two neighbors with different measurements (minimum and maximum) [10]. Averaging/summing aggregate problems show up in distributed sensor networks, while averaging/summing agreement problems do not arise in centralized sensor networks [1]. Distributed consensus algorithms are not confined to WSNs, and they can be applied to distributed processor computing [11], distributed data base management or distributed signal processing on the Internet for example [1]. In distributed manner, every node has a local information. In a cluster-based WSN, it is still not possible to apply gossip algorithms on it. Each cluster needs a fusion centre to connect to the gate way. The fusion centre requires entire information about the cluster, while aggregate values do not need entire information. The latter point makes gossiping more robust against link failure and not possible to apply on cluster-based WSNs.

The proposed algorithm, named geographic greedy triplewise gossip algorithm (GGTGA), exploits the good points in both G-TGA and GGA to improve both convergence time and cost. The rest of the paper is organized as follows. The problem formulation including distributed averaging, network model and time model is presented in Section 2, then our algorithm (GGTGA) is proposed and analyzed in Section 3. Simulation results and conclusion are presented in Sections 4 and 5, respectively.

2. PROBLEM FORMULATION

2.1 Distributed Averaging

In WSN with N nodes, the i^{th} node has an initial scalar measurement, $x_i\ (0),$ in some modality of interest (e.g., temperature, pressure, light, …etc.). The aim of the averaging algorithms is reaching the global average $x_{ave} = \frac{1}{N}\sum_{i=1}^{N} x_i(0)$ from the local measurements [1], [5]-[8] and [10]. We are interested in the number of iterations or rounds required for convergence and the number of radio transmissions passing through the network during the averaging process. At each round $t = 1{:}\,T_{ave},$ a set of nodes updates their estimations [1], [5]-[10], where T_{ave} represents the total practical time of convergence of true global averaging. The gossip algorithms converge to the almost surely true average if $P\ [\lim_{t\to\infty} \varepsilon(t) = 0] = 1,$ where $\varepsilon(t) = ||\boldsymbol{X}(t) - x_{ave}\mathbf{1}||_2$ is the estimation error, $\boldsymbol{X}$(t) is the $N*1$ vector of measurements, and **1** is the $N*1$ unit vector [12]-[13]. The gossip algorithms operate as follows: At each round in a set of nodes, at least two nodes are averaging and updating their estimations per round. Let $S(t)$ represent a set of nodes at time t and $x_i(t)$ the estimation value for $node\ i \in S(t)$. The nodes update their estimations according to Equation (1):

$$x_i(t) = \frac{1}{|S(t)|} * \sum_{i\in S(t)} x_i\ (t-1)\ . \tag{1}$$

The rest of the nodes remain unchanged in this round [11]:

$$x_k(t) = x_k(t-1)\ . \tag{2}$$

Table 1. Key term's definition.

Item	**Definition**
Time of convergence:	The time of convergence is the total time accounted until $\boldsymbol{X}$(t) = x_{ave} **1** is reached.
Communication cost:	The number of the entire messages spent until reaching the exact convergence.

2.2 Network Model

Sensor nodes deployment strategies play an important part in the performance of networks. Many topologies can be found in the wireless model such as ring, grid and random geometric graph (RGG) …etc. The random geometric graph G (N,r) is an irregular model and suitable topology for WSN. RGG is formulated by choosing N nodes uniformly and independently in the unit square $[0,1]^2$, [1], [5], [8] and [10]. The transmission range of a node is $r = \sqrt{\frac{c*\log(N)}{N}}$ in order to maintain connectivity and prevent interference [10]. The radio transmission range r plays an important role in convergence; small radio transmission ranges result in slow convergence, even for fast averaging algorithms [6]. Therefore, r must be set carefully. The constant c will be assigned the value c=2, which is the suitable value for the TGA algorithms [10] in order to test all the considered algorithms under the same conditions to compare their behavior.

2.3 Time Model

We use an asynchronous time model, which is a more suitable time model for distributed networks. In the asynchronous time model, each node has an independent clock, which ticks at the random time rate λ following a Poisson process. The inter-tick times between each two iterations are independently and identically distributed (i.i.d) and are inversely proportional to $N\lambda$. If λ is small enough, then there is only one iteration at a time with high probability and each communication has greater chance to succeed. If λ is too large, then there is a high chance that a node becomes activated while another node is still operating. In this case, and if the network has a huge number of nodes, the network nodes are prone to failure in updating their estimates [1], [6]-[13].

2.4 Gossip Algorithms

2.4.1 Pairwise or Standard Gossip Algorithm (SGA)

This is also called nearest-neighbour gossip algorithm, the earliest distributed averaging method proposed by [5]-[6]. At each round, the asynchronous averaging algorithm activates one node (s) at random and averages its value with one–hop neighbours (d) at random with probability P_{sd}. Both sensor nodes update their values by replacing their own value with the calculated average. The averaging is done by putting 0.5 in indices (s,s), (s,d), (d,s) and (d,d) of the identity matrix W(t), where W(t) is a random, symmetric, doubly stochastic, independent and identity matrix [13]. Algorithm 1 explains the pairwise gossip strategy systematically.

Practically, time convergence can be defined as the first time when $\|\boldsymbol{X}(t) - x_{ave}\mathbf{1}\|_2$ equals zero [9]. This algorithm converges slowly and wastes energy because of significant recalculation of redundant information. This motivated other researchers to propose other distributed averaging methods. Communication cost can be theoretically and practically calculated. Practically, SGA costs two message transmissions per round and therefore the total number of messages is calculated as:

$$Cost\ (practical)\ =\ total\ time\ practically\ calculated * 2\ \ . \qquad (3)$$

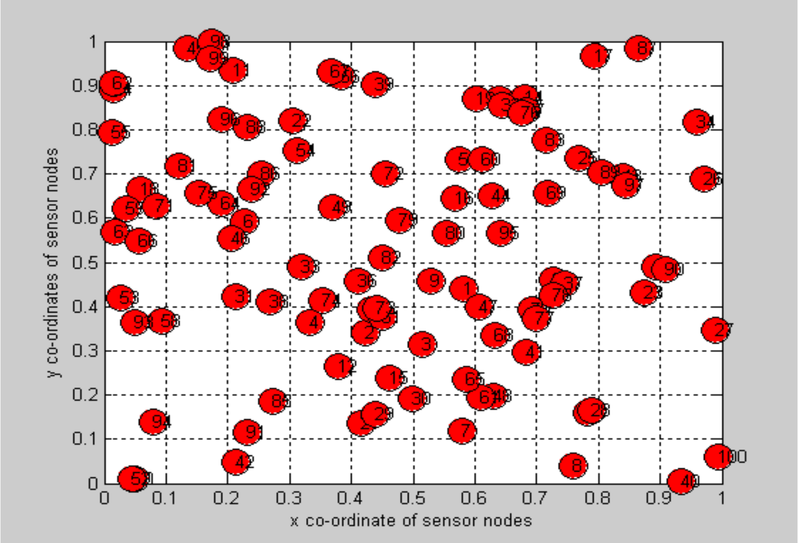


Figure 1. Sensor Nodes Distributed in RGG.

Theoretically, the cost of SGA is computed by $cost = o(\frac{N}{r^2}\log(\epsilon^{-1})\ )$ for $r = \sqrt{2 * \log(N)/N}$. he equation for the cost becomes [8]:

$$Cost\ (theoretical) = O(\frac{N^2}{2*\log(N)}\log(\epsilon^{-1})\ ), \tag{4}$$

where value of ϵ represents averaging accuracy between (0,1).

Algorithm 1: Standard Gossip Algorithm:

```
for t=1: $T_{ave}$.
    s=activated node at random.
    d=neighbor node to node (s) selected at random.
    $x_d(t) = \frac{x_d(t-1)+x_s(t-1)}{2}$, the new estimate is sent back to node(s).
    $x_s(t) = x_d(t)$.
end for.
```

2.4.2 Geographic Gossip Algorithm (GGA)

This algorithm was proposed after a disappointing result of slow mixing time of SGA and waste of energy due to significant recalculation of redundant information. GGA is an asynchronous algorithm that accelerates the pairwise standard averaging by exploiting geographic information of nodes as well as geographic location of their neighbours. Thereby, GGA is able to route a node's estimation value to far away nodes in RGG, by utilizing greedy routing towards the destination [8], [14].

At each round, one node is activated, a location point (*xd*,*yd*) is chosen and greedy route towards the closest node to the chosen location is started. The receiving node calculates pairwise average and utilizes the route in reverse. The activated node will receive the new estimate and update its value. Algorithm 2 shows the behaviour of GGA. Routing is expensive in terms of communication cost. Nevertheless, it is the reason for achieving convergence acceleration by gossiping with random nodes, which are far away in the network [8]. GGA saves a factor $\sqrt{\frac{N}{log\ N}}$ over SGA in terms of communication cost on RGG so the equation will be as follows [8]:

$$Cost\ (theoretical) = O(\frac{N^{1.5} * \log \epsilon^{-1}}{2 * \sqrt{\log(N)}}). \tag{5}$$

The number of messages is not constant at each round, thus the number of messages cannot be calculated practically, since the route is variable in length at each round. In terms of time of convergence, the estimation error equation is used to show the practical time for convergence. The time required for convergence can be computed theoretically depending on Equation (6) in terms of ϵ [8]:

$$T_{ave}(theoretical) = O(N * \log \epsilon^{-1}). \tag{6}$$

Algorithm 2: Geographic Gossip Algorithm:

```
for t=1: $T_{ave}$.
s=activated node at random.
(xd,yd) selected point by node (s) starts greedy routing toward the closest node to the chosen point.
d =is the closest node.
$x_d(t) = \frac{x_d(t-1)+x_s(t-1)}{2}$, the new estimate is sent back to node (s) through the same route.
$x_s(t) = x_d(t)$.
$end\ for$.
```

2.4.3 Standard Triplewise Gossip Algorithm (Standard TGA)

TGA is a recently introduced, asynchronous algorithm that enhances the distributed ability by enlarging the gossip group and thereby reaching a good estimation of the average with fewer rounds and less communication cost [10]. In SGA, the communication complexity is very high on RGG [1], [7].

In standard TGA, at each round, one node wakes up at random, selects two of its neighbours at random and averages their estimations and then all these three nodes update their values to be equal to the new local averaging estimation. There is an exception when the number of neighbours for the activated node is equal to one, then pairwise averaging is performed instead of triplewise averaging [10].

Practically, standard TGA costs four message transmissions per round. One transmission is required from the source node (*s*) to two destination nodes (d_1 & d_2) and two transmissions from (d_1 & d_2) to (*s*). Finally, node (*s*) calculates the averaging of three nodes (*s*, d_1 & d_2) and transmits the new estimation to both nodes (d_1 & d_2) [10]. Algorithm 3 illustrates the averaging method by standard TGA.

$$Cost\ (practical) = total\ practical\ time\ required\ for\ convergence * 4. \tag{7}$$

Algorithm 3: Standard Triplewise Gossip Algorithm:

```
for t=1: $T_{ave}$.
s=activated node at random.
node (s) sends a broadcast message to all its neighbors.
$d_1$ & $d_2$ =two neighbor node to node (s) which was selected at random.
$x_s(t) = \frac{x_{d1}(t-1)+x_s(t-1)+x_{d2}(t-1)}{3}$, the new estimate is sent back to nodes ($d_1$ and $d_2$).
$x_{d1}(t) = x_{d2}(t) = x_s(t)$.
$end\ for$.
```

2.4.4 Greedy–Triplewise Gossip Algorithm (G-TGA)

This is an asynchronous algorithm almost identical to standard TGA, but instead of the activated node dealing with two neighbours at random, G-TGA deals with two neighbours having specific different values. This point improves the time for convergence more than standard TGA and then provides less message transmissions. The activated node chooses two of its neighbours: one having the minimum estimate and the second having the maximum estimate among all neighbours. This algorithm requires six message transmissions per round. Algorithm 4 explains the behaviour of G-TGA systematically. Node (s) is activated, then two neighbour nodes N_s are selected one with the maximum value and the second with the minimum value among all neighbours of the activated node. First, 4-radio transmission is required as in standard TGA; after the activated node changes its value, the two destination nodes will update and broadcast their values [10].

$$Cost\ (practical) = total\ practical\ time\ required\ to\ convergence * 6. \tag{8}$$

Algorithm 4: Greedy-Triplewise Gossip Algorithm:

```
1. for t=1: $T_{ave}$
2. s=activated node at random
3. node (s) sends a broadcast message to all its neighbors
4. $d_1$ & $d_2$ =two neighbors of node (s) and having different values
5. $d_1$ =node has a minimum value
6. $d_2$ =node has a maximum value
7. $x_s(t) = \frac{x_{d1}(t-1)+x_s(t-1)+x_{d2}(t-1)}{3}$ The new estimate is sent back to nodes ($d_1$ and $d_2$)
8. $x_{d1}(t) = x_{d2}(t) = x_s(t)$
9. $end\ for.$
```

3. Geographic Greedy Triplewise Gossip Algorithm (GGTGA)

Our algorithm is proposed to reduce the number of radio transmissions and of iterations to reach global convergence. Every node has known its location and the geographic location of its neighbours. For each *t=1, 2, …,* T_{ave}, one node is activated and chooses a location point (x_d,y_d) at random. The activated node will use greedy routing toward two nodes within the transmission range of the chosen location, one of them having the minimum value measurement and the other having the maximum value among the other nodes in the transmission range. The activated node (s) will send its activation message to the two chosen nodes (d_1 & d_2) by forwarding the message through the path.

The two destination nodes (d_1 & d_2) will receive the activation message of node (s), then the destination nodes will send their estimated values (the maximum and minimum estimation values) to node (s) using the same route that node (s) followed to send its activation message to (d_1 & d_2). The activated node (s) will compute the average of its value and the values of the two destination nodes according to the following equation:

$$x_s(t) = \frac{x_s(t-1)+x_{d_1}(t-1)+x_{d_2}(t-1)}{3} \tag{9}$$

Then, node (s) uses the same route to forward the new updated value to both nodes (d_1 & d_2). At the end of the iteration, Equation (10) results, while the remaining nodes remain unchanged as in Equation (11):

$$x_{d_1}(t) = x_{d_2}(t) = x_s(t), \tag{10}$$

$$x_k(t) = \ x_k(t-1)\,. \tag{11}$$

If we let $\boldsymbol{X}$(t) indicate the vector of estimated values at the end of the time slot t, then the algorithm execution can be described as a sequence of iterations:

$$\boldsymbol{X}(t) = \boldsymbol{W}(t)\boldsymbol{X}(t-1), \tag{12}$$

where **W** is the weighted averaging matrix [10]. **W** is a random, symmetric, doubly stochastic and semi definite-programming selected averaging matrix. **W** (t) is an i.i.d selected matrix at every time round [1], [5], [9]-[10] and [14]. For any gossip algorithm $\mathbf{W}^2$=**W** at each round, i.e., **W**(t) is a projection matrix since averaging the same set twice no longer changes the vector **X**(t) [1], [4].

Let $\alpha_{i,j} = e_i - e_j$, where $e_i = [0,0 \ldots.,1,\ldots,0]^T$is an *N*1* unit vector with i^{th} element equals 1. With probability $\frac{1}{N} * P_{i,j} * P_{i,k}$, the random symmetric matrix $\boldsymbol{W}(t)$ is:

$$\boldsymbol{W}_{i,j,k} = I - \frac{\alpha_{i,j}\alpha_{i,j}{}^T + \alpha_{i,k}\alpha_{i,k}{}^T + \alpha_{j,k}\alpha_{j,k}{}^T}{3}. \tag{13}$$

Algorithm 5: Geographic Greedy-Triplewise Gossip Algorithm (GGTGA):

```
1. for t=1: $T_{ave}$.
2. s=activated node at random.
3. (xd,yd) selected point at random by node (s) starts greedy routing toward the two nodes within transmission range of the chosen point, one with the maximum value and the other with the minimum value.
4. $d_1$ =node with the minimum value sends its value to node (s).
5. $d_2$ =node with the maximum value sends its value to node (s).
6. $x_s(t) = \frac{x_{d1}(t-1)+x_s(t-1)+x_{d2}(t-1)}{3}$, the new estimate is sent back to nodes ($d_1$ and $d_2$).
7. $x_{d1}(t) = x_{d2}(t) = x_s(t)$.
8. end for.
```

Nodes require memory to save this information. Sensor nodes that need to save nodes location call for additional memory requirements. This is the weakness point of our proposed algorithm GGTGA and GGA.

3.1 Time of Convergence

The convergence time of the proposed algorithm will improve over that of G-TGA, standard TGA, GGA and SGA, as we will show in Section 4 in the simulation results. For GGA, the convergence time T_{ave} is theoretically given by Equation (6), where ϵ is the averaging accuracy with values between 0 and 1 [8]. The value of ϵ is very important for the calculation of the communication overhead for GGA in Sub-section 3.2 below, since the number of messages per round is not constant. It gives the level of accuracy and takes the same value that was taken in Equation (6). Practically, all gossip algorithms depend on the estimation error function to see the speed up of distributed averaging algorithm. Many methods can be used to accelerate convergence, such as adding a shift register [15], but additional hardware causes the node to consume power and increases its size. Other algorithms use the broadcast property for communication, such as broadcast gossip algorithm (BGA) [16]. BGA has a lot of disadvantages; it needs to optimize certain parametric values well like (γ) [16] and does not converge to the correct average of initial value of all nodes [10]. Deterministic averaging algorithms are faster than randomized averaging algorithms (gossip algorithms), putting conditions for the selection of destination nodes which helps accelerate convergence, while full randomization increases redundancy and hence slows convergence. If the selection of nodes is not fully random, the convergence time will be enhanced.

3.2 Communication Cost

Now, we need to see how our algorithm reduces the number of radio transmissions as well. It is worth noting that the number of messages in GGA and the proposed GGTGA is not constant per iteration, since it depends on the path it takes in each iteration. So if we find the average cost per iteration in GGA, we can estimate the communication cost for the proposed GGTGA by proportionally taking into account the convergence time T_{ave} obtained practically for GGTGA. The rationale behind this assumption is that GGTGA also employs geographical routing as in GGA, and therefore, the average communication cost in a path should be comparable. The result is then multiplied by 2 to account for the two routes of GGTGA. Although this communication cost would only be an estimate, it is at least guaranteed to be of the same order of magnitude of the exact value.

4. SIMULATION

We use Matlab to simulate and consider a static, time invariant [13] connected network consisting of 100 nodes that are uniformly and independently distributed in the unit square $[0,1]^2$. We will first consider the convergence times for G-TGA, GGA and SGA and compare them with that of the proposed GGTGA. Figure 2 shows the convergence time that results in an estimation error $\varepsilon(t) = ||\boldsymbol{X}(t) - x_{ave}\mathbf{1}||_2$ equal to zero [9], [13].

As is clear from Figure 2, GGTGA needs only 323 iterations. It remarkably accelerates the time for convergence. The slowest algorithm, which has a slow mixing time, is SGA. It needs 7213 iterations for convergence. The convergence times for the different algorithms are shown in Table 2. Figure 3 shows the logarithmic representation of the estimation error given by Equation (14) below [9], [13]:

$$\varepsilon(t) = log(||\boldsymbol{X}(t) - x_{ave}\mathbf{1}||_2). \tag{14}$$

We now turn to the calculation of the number of radio transmissions. Substituting the practical value of T_{ave} of Table 2 for GGA in Equation (6) that represents the theoretical convergence time for GGA, and assuming exact equality, we compute ϵ for GGA in order to substitute it in the corresponding equation to obtain the number of radio transmissions. As for the proposed GGTGA, the communication cost is found as outlined in Section 3.2.

Figure (4) and Figure (5) represent simulation for the number of message-passings for each presented gossip algorithm. Table 2 shows the number of radio transmissions for the various algorithms as well as the required ϵ if needed. GGA needs ϵ in order to get the expected number of messages, since the number of messages per round is not constant. With a level of accuracy equal to (3.1623e^{-016}), the expected cost is almost 4 messages per iteration for GGA and thus we can deduce the expected cost for our proposed algorithm GGTGA. We calculate the expected cost per iteration, since the greedy routing is found at each iteration. There is a trade-off between the level of accuracy and the number of both iterations and message-passings.

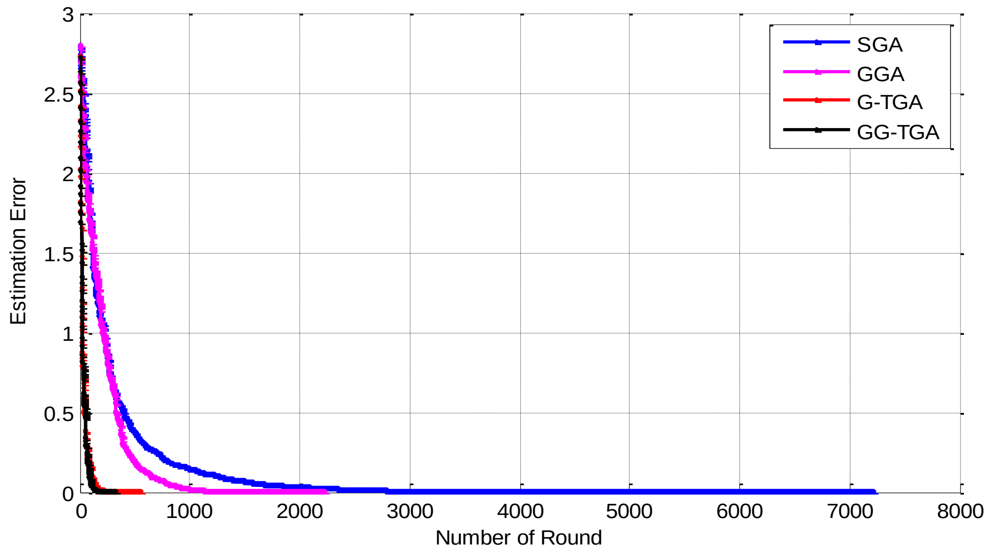


Figure 2. Convergence of Various Gossip Algorithms.

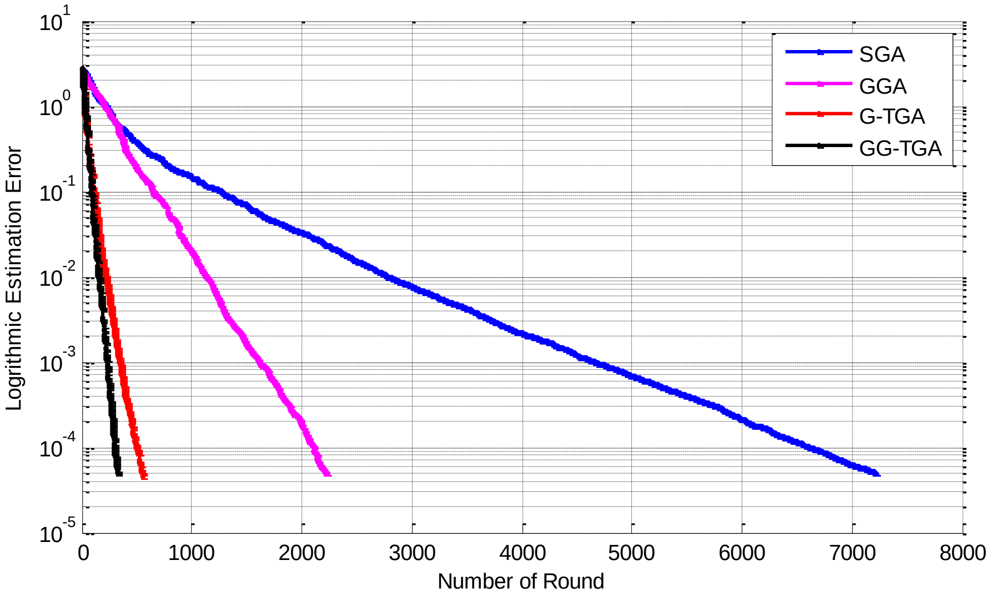


Figure 3. Logarithmic Convergence of Various Gossip Algorithms.

Table 2. The behaviour of different gossip algorithms.

Aggregate Algorithm				
Parameter evaluation	**SGA**	**GGA**	**G-TGA**	**GGTGA**
Number of iterations	7213	2224	552	323
ϵ	/	3.1623e^{-016}	/	/
Number of messages	14426	8896	3312	2584

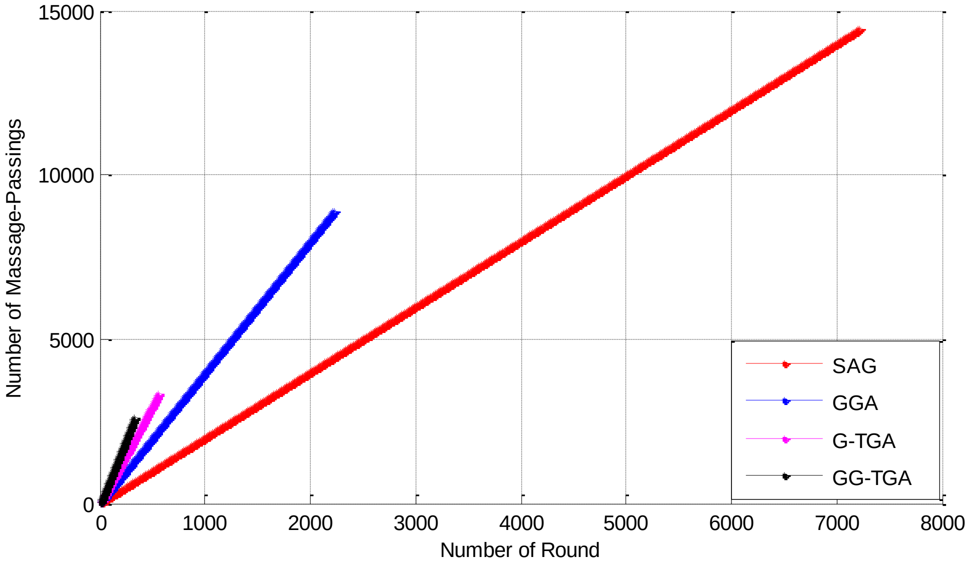


Figure 4. Linear Representation of Communication Overhead for Different Gossip Algorithms.

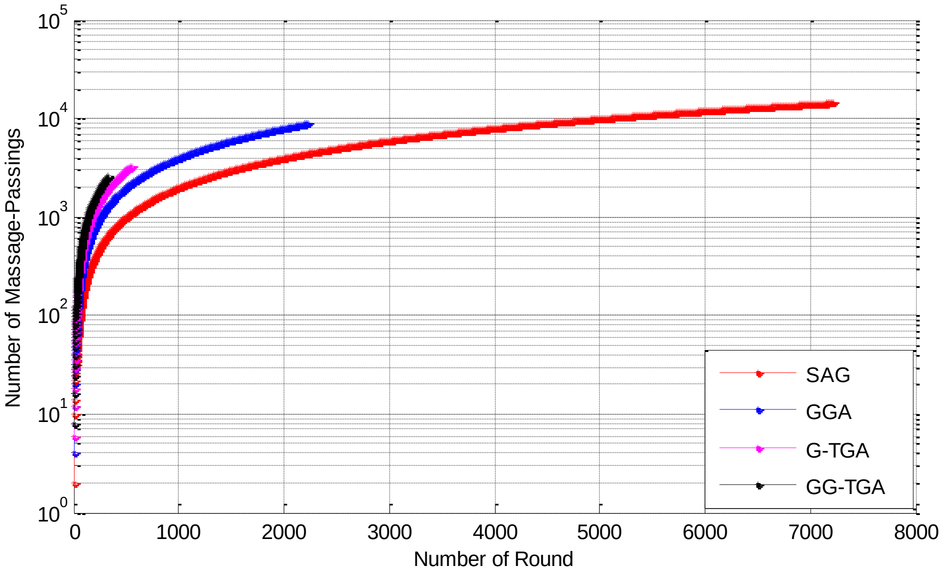


Figure 5. Logarithmic Representation of Communication Overhead for Different Gossip Algorithms.

5. Conclusion

Deterministic averaging algorithms are faster than randomized averaging algorithms (gossip algorithms). Putting conditions for the selection of destination nodes helps accelerate convergence, while full randomization increases redundancy and hence slows convergence. If the selection of nodes is not fully random, the convergence time will be enhanced with little cost. We propose a novel gossip algorithm that accelerates the time of convergence and reduces the number of radio transmissions needed to perform distributed averaging in WSNs. These two parameters determine the amount of power consumption in distributed WSNs. Our algorithm greatly reduced both convergence time and number of radio transmissions. Therefore, it was shown to outperform other existing gossip algorithms in terms of energy saving. Nodes require memory to save this information. Sensor nodes that need to save node locations call for additional memory requirements. This is the weakness point of our proposed algorithm GGTGA and GGA.

REFERENCES

[1] F. Benezit, Distributed Average Consensus for Wireless Sensor Networks, Ph.D. dissertation, vol. 4509, 13 November 2009.

[2] M. I. Raheem and N. A. S. Alwan, "Performance Evaluation of Different Gossip Algorithms for Distributed QR Factorization in Wireless Sensor Networks," Emirates Journal for Engineering Research, vol. 20 , no. 1, pp. 93-100, 2015.

[3] O. Sluciak, H. Strakova, M. Rupp and W. N. Gansterer, "Distributed Gram-Schmidt Orthogonalization Based on Dynamic Consensus," In: IEEE 2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), pp. 1207-1211, 4-7 Nov. 2012.

[4] N. A. S. Alwan and A. S. Mahmood, "Distributed Gradient Descent Localization in Wireless Sensor Networks," Arabian Journal for Science and Engineering, vol. 40, no. 3, pp 893-899, March 2015.

[5] A. G. Dimakis, S. Kar, J. M. F. Moura, M.G. Rabbat and A. Scaglione, "Gossip Algorithms for Distributed Signal Processing," Proceedings of the IEEE, vol. 98, no. 11, pp. 1847-1864, 2010.

[6] S. Boyd, A. Ghosh, B. Prabhakar and D. Shah, "Gossip Algorithms: Design, Analysis and Applications," Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies, vol. 3, pp. 1653–1664, 13-17 March 2005.

[7] S. Boyd, A. Ghosh, B. Prabhakar and D. Shah, "Randomized Gossip Algorithms," IEEE Transactions on Information Theory, vol. 52, no. 6, pp. 2508–2530, June 2006.

[8] A .G. Dimakis, A. D. Sarwate and M. J. Wainwright, "Geographic Gossip: Efficient Averaging for Sensor Networks," Proceedings of the 5th International Conference on Information Processing in Sensor Networks, pp. 69-76, 19-21 April 2006.

[9] F. Benezit, P. Denantes, A. G. Dimakes, P. Thiran and M. Vetterli, "Reaching Consensus about Gossip: Convergence Time and Costs," Information Theory and Applications, NCCR-MICS, 2009.

[10] B. Yang, W. Wu and G. Zhu, "Distributed Averaging in Wireless Sensor Networks with Triplewise Gossip Algorithms," In: IEEE 2013 TENCON - Spring Conference, pp. 178-182, 17-19 April 2013.

[11] D. Bertsekas, M. Athans and J. Tsitsiklis, "Distribution Asynchronous Deterministic and Stochastic Gradient Optimization Algorithm," IEEE Transactions on Automatic Control, vol. AC-31, no. 9, Sept. 1986.

[12] F. Benezit, A. G. Dimakis, P. Thiran and M. Vetterli, "Order–Optimal Consensus Through Randomized Path Averaging," IEEE Transactions on Information Theory, vol. 56, no. 10, pp. 5150-5167, October 2010.

[13] P. Denantes, F. Benezit, P. Thiran and M. Vetterli, "Which Distributed Averaging Algorithm Should I Choose For My Sensor Network?," In: The 27th Conference on Computer Communications, IEEE INFOCOM 2008, 13-18 April 2008.

[14] F. Benezit, V. Blondel, P. Thiran, J. Tsitsiklis and M. Vetterli, "Weight Gossip: Distributed Averaging Using Non-Doubly Stochastic Matrices," In: 2010 IEEE International Symposium on Information Theory, pp. 1753-11757, 13-18 June 2010.

[15] C. Ming, A. S. Daniel and M. Y. Edmund, "Accelerate Gossip Algorithms for Distributed Computation," 44th Annual Allerton Conference, Allerton House, UIUC, USA, pp. 952-959, 27-29 September 2006.

[16] T. C. Aysal, M. E. Yildiz, A. D. Sarwate and A. Scaglione, "Broadcast Gossip Algorithm for Consensus," IEEE Transaction on Signal Processing, vol. 57, no. 7, pp. 2748-2761, July 2009.

ملخص البحث:

في هذه الورقة البحثية، يتم اقتراح خوارزمية غوسِب (Gossip) جديدة لإيجاد المعدَّل الموزَّع وتمتاز بتقاربٍ سريع وكلفة منخفضة للاتصال مقارنةً بشبكات المجسّات اللاسلكية (WSNs).

وقد أثبتت الخوارزمية المقترحة تحسيناً للسلوك بالنسبة لخوارزمية غوسِب القياسية (SGA)، وخوارزمية غوسِب الثلاثية (TGA)، وخوارزمية غوسِب الجغرافية (GGA)؛ وذلك عن طريق الاستفادة من المعلومات الجغرافية للشبكة. وتقدم هذه الدراسة تحليلاً لزمن التقارب وكلفة الاتصال للخوارزمية المقترحة، إلى جانب مقارنة مع طرقٍ أخرى قائمة.

A 1.2-V Low-Power Full-Band Low-Power UWB Transmitter With Integrated Quadrature Voltage-Controlled Oscillator and RF Amplifier in 130nm CMOS Technology

Fadi R. Shahroury

Department of Electrical Engineering, Princess Sumaya University for Technology,
Amman, Jordan.
fadi@psut.edu.jo

(Received: 10-Feb.-2016, Revised: 11-Apr.-2016, Accepted: 24-Apr.-2016)

Abstract

This paper presents the design and simulation of a low-power full-band UWB transmitter with on-chip quadrature voltage-controlled oscillator (QVCO) in 130 nm CMOS technology. The proposed transmitter consists of a passive poly-phase filter (PPF), QVCO, a quadrature modulator core and an RF power amplifier. The QVCO uses the differential delay cell architecture with four cascaded stages. The transmitter has the following specifications: a 15.28 dB average conversion gain with a ripple of ±1dB from 2 GHz to 11 GHz, the average input 1-dB compression point (IP1dB) is –10 dBm and the average output 1-dB compression point (OP1dB) is 4.35 dBm. The QVCO achieves a wide frequency range (2-11 GHz) with a –80 dBc/Hz phase noise. In addition, the supply voltage of the proposed transmitter is 1.2 V with power consumption of 77.8 mW.

Keywords

UWB, QVCO, Up-conversion mixer, 130 nm CMOS technology.

1. Introduction

Ultra-Wideband (UWB) technology has been around since the 1980s, but it has been used for radar applications [1], since it gives accurate timing information due to the signal wideband nature. However, the need for high data rate has made short-range UWB wireless communications quite popular. Besides, the UWB is becoming more attractive for low-cost communication applications, because it provides low power consumption, large bandwidth and high data rates (up to 480 Mbps within 10m distance) [2]. Multi-band orthogonal frequency division multiplexing (MB-OFDM) and impulse radio communication are both under the UWB standards. However, the impulse radio approach faces the potential problem of many narrow-band systems co-existing as well as several technical challenges related to generation of short pulses [3]. Thus, overcoming these challenges enhanced the MB-OFDM approach which is used in this work.

So far, a variety of transmitters have been reported to implement MB-OFDM UWB transmitter [3]-[7]. However, scrutiny of these papers revealed that all UWB transmitters proposed have relatively high power consumption, although they did not contain an on-chip QVCO. Besides, they did not cover the full-band of MB-OFDM (2-11GHz), except the proposed work in [7]. However, the transmitter in [7] is designed with 14 on-chip spiral inductor.

In this paper, a UWB CMOS transmitter for multi-band OFDM applications is implemented using 130 nm CMOS technology with power consumption of 77.8 mW from a 1.2 V supply voltage. The designed transmitter covers the full-band of MB-OFDM (2-11GHz). It consists of a passive poly-phase filter (PPF), a quadrature voltage-controlled oscillator (QVCO), a quadrature modulator core and an RF power amplifier. In addition, only 8 on-chip inductors are used in the proposed transmitter.

This paper presents the design and simulation of a low-power full-band UWB transmitter. The system overview for UWB is discussed in Section 2, the operational principles and the design of the building blocks is addressed in Section 3. The simulation results of the proposed transmitter circuit are reported in Section 4, followed by a conclusion in Section 5.

2. System Overview

MB-OFDM UWB system is a system that enables transmitting data over multiple carriers at precise frequencies at the same time. Multi-band OFDM system consolidates the multi-band (MB) approach with the OFDM modulation technique. This approach divides the spectrum into smaller sub-bands each with a bandwidth greater than 500 MHz (Federal Communications Commission (FCC) requirement for a UWB system) and uses one of the sub-bands in each time-slot to transmit the OFDM symbols. The system is denoted as a UWB-OFDM system since it is an OFDM system, but operates over a very wide bandwidth, much larger than the conventional OFDM system bandwidth.

In order to uphold the UWB standards based on OFDM, the MB-OFDM Alliance (MBOA) was formed in 2003. As stated in the MBOA and WI Media standard [8], the UWB spectrum extends over the frequency range from 3.168 GHz to 10.56 GHz and is divided into 14 bands, each band of which has a bandwidth of 528 MHz and consists of 128 sub-channels, where the bandwidth of each sub-channel is 4.125 MHz. The bands are gathered into five groups, the first four groups consist of three bands, so they contain the first 12 bands, and the fifth one contains the last 2 bands, as illustrated in Figure 1.

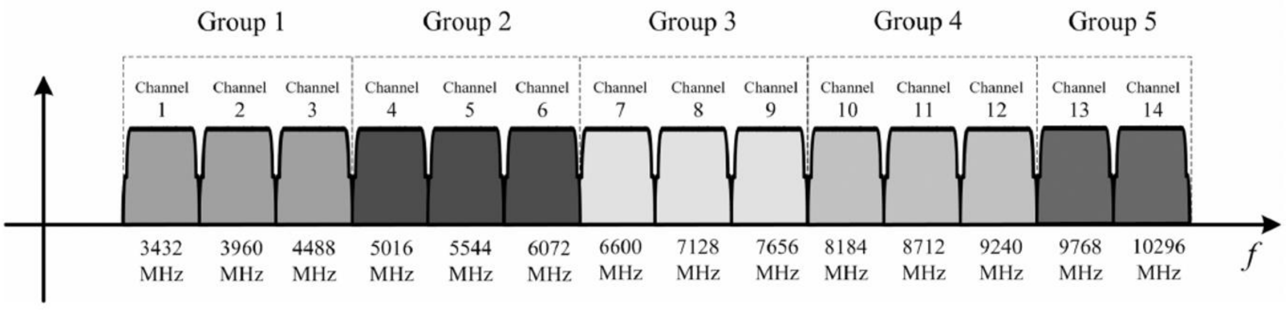


Figure 1. The allocation of the bands in the UWB MB-OFDM system.

3. PRINCIPLE OF OPERATION

Figure 2 shows the block diagram of the proposed UWB transmitter system which consists of a passive poly-phase filter, a quadrature modulator core, a quadrature voltage-controlled oscillator and an RF power amplifier. The quadrature modulator core is formed by two mixers and one adder and converts the quadrature intermediate frequency (IF) signals and the quadrature local oscillator (LO) signals into the single-side-band signal. The quadrature LO signal is obtained from a QVCO which operates from 2 GHz to 11 GHz, and the quadrature of the input IF signal is achieved by the passive poly-phase filter. Then, the amplification of the output signal of the quadrature modulator is carried out using an RF power amplifier, in order to drive the load of the output. For the measurement purpose, the single ended to differential and differential to single ended off-chip transformation at the IF input port and RF output port are used, respectively.

The following subsections describe the operational principles and the circuits' realization of the quadrature modulator core, the PPF, the QVCO and the power amplifier.

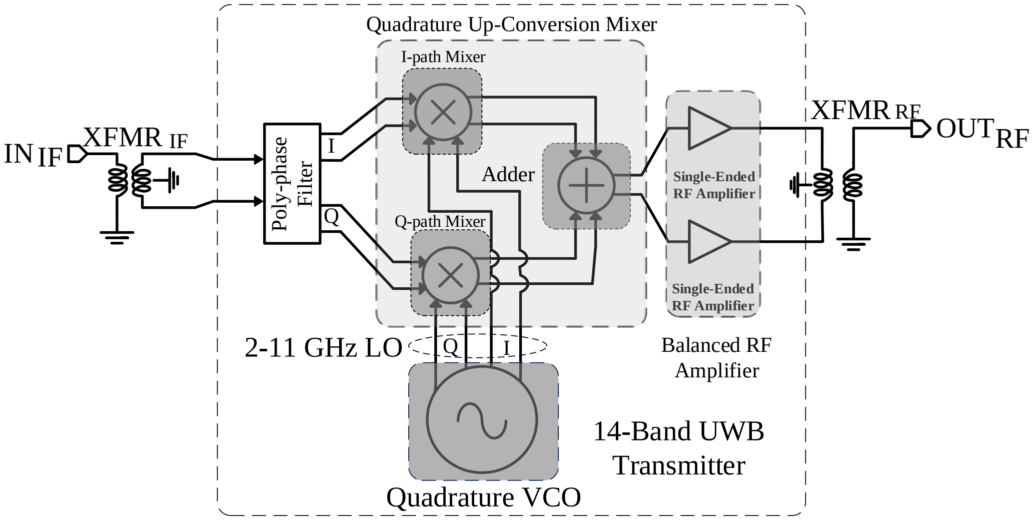
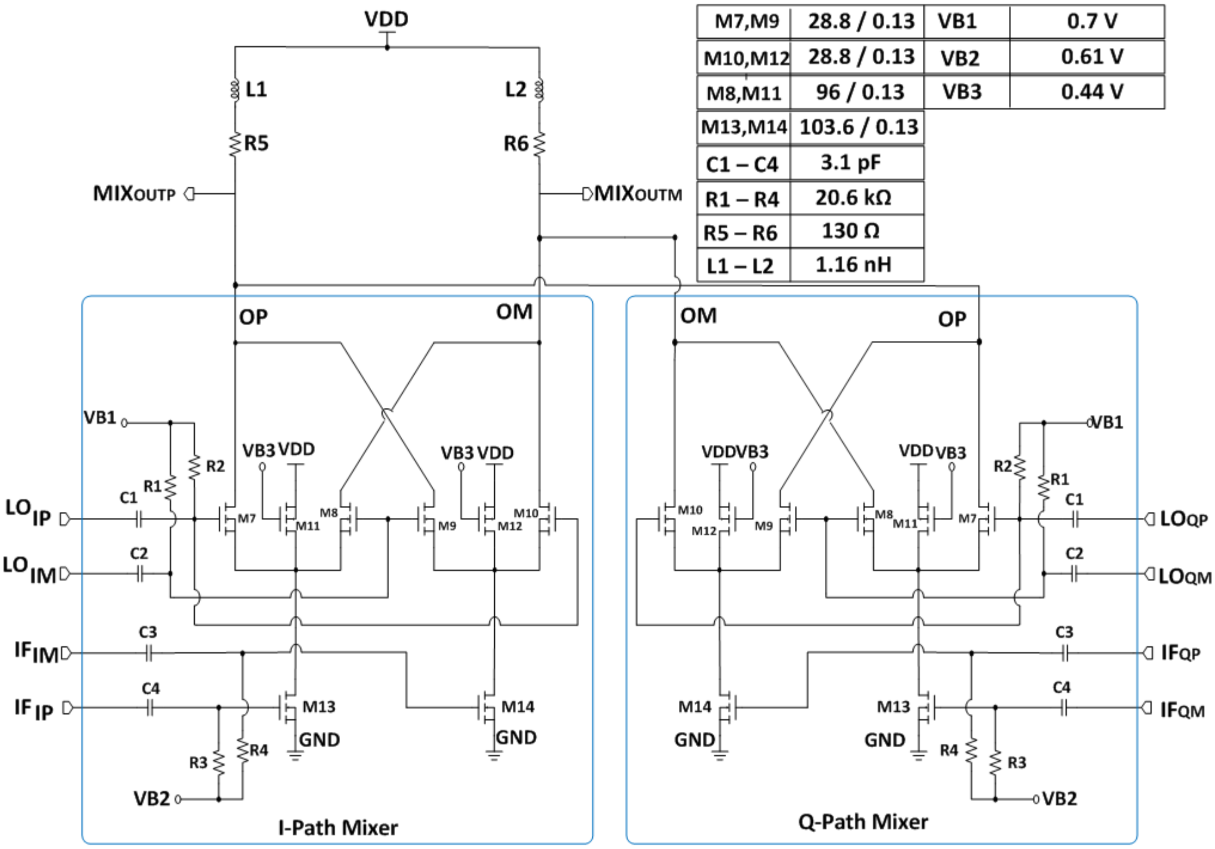


Figure 2. The block diagram of the full-band UWB transmitter.

3.1 Quadrature Modulator Core

Figure 3. shows the schematic diagram of the quadrature modulator core which is composed of two paths: the I-path and the Q-path. The output of each path is added in current domain at nodes OP and OM. Every path contains a double-balanced quadrature up-conversion mixer.

In each double-balanced quadrature up-conversion mixer, transistors M7-M10 form the Gilbert cell core [9], where the LO signal is injected. Transistors M13 and M14 are used at the IF input, while M11 and M12 are used as a current source to enhance the conversion gain of the mixer by realizing bleeding technique. The capacitors C1- C4 are used as DC block capacitors, and resistors R1- R4 are used for biasing. Each mixer consumes 16.72 mA from a 1.2 V power supply. The entire mixer including both paths has a total power consumption of about 40.1 mW.



M7,M9	28.8 / 0.13	VB1	0.7 V
M10,M12	28.8 / 0.13	VB2	0.61 V
M8,M11	96 / 0.13	VB3	0.44 V
M13,M14	103.6 / 0.13		
C1 – C4	3.1 pF		
R1 – R4	20.6 kΩ		
R5 – R6	130 Ω		
L1 – L2	1.16 nH		

Figure 3. Simplified schematic diagram of the quadrature modulator core.

3.2 Passive Poly-phase Filter (PPF)

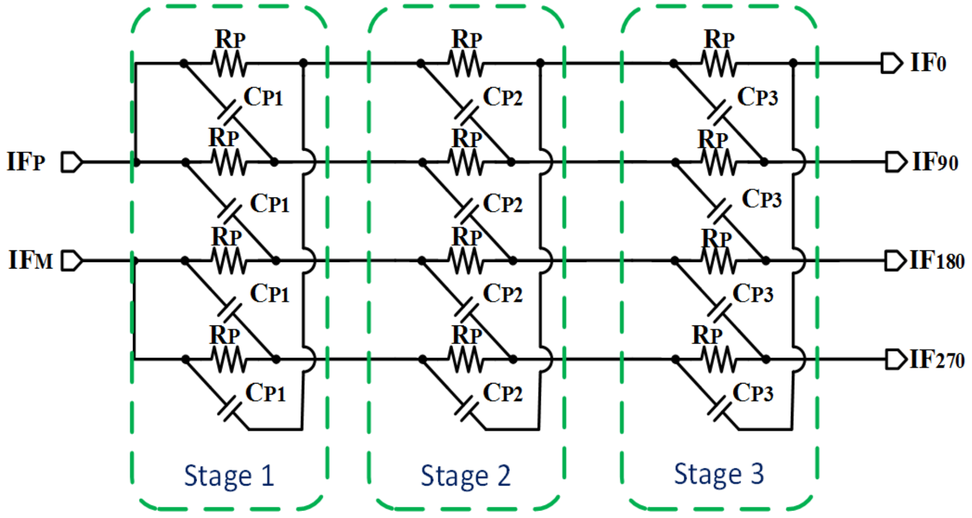
PPF is designed to provide the quadrature modulator core with differential and quadrature inputs, thereby allowing the use of a double-balanced mixer topology. This property proves critical in reducing the LO-to-IF feedthrough. To achieve wider range of operation, three-stage poly-phase RC filter is employed, as illustrated in Figure 4. The unit resistor is 5kΩ in the three-stages, while the unit capacitor is 440 fF, 240 fF and 120 fF in the first, second, and third stage, respectively. To verify the phase precision of PPF outputs, the image rejection ratio simulation methodology is used [10].

3.3 Quadrature Voltage Controlled Oscillator (QVCO)

Figure 5 illustrates the block diagram of the QVCO. A ring QVCO with four differential delay cells is chosen. This structure can effectively reduce chip area and cost as compared to QVCO based on LC ring oscillator structure.

The schematic diagram of each differential delay cell is depicted in Figure 6. In Figure 6, the differential delay cell consists of a differential transistor pair (M3 and M4) with load transistors (M1 and M2) and tail current transistors (M5 and M6). The operation of the QVCO is divided into two modes, high frequency mode (3.28-11.16 GHz) and low frequency mode (1.5-6.2 GHz) based on which tail current transistor (M5 or M6) of the differential delay cell is enabled. In each mode of operation, the oscillation frequency of the ring QVCO is controlled by the gate voltage of M3 and

M4. The designed QVCO based on proposed delay cell covers the frequency range from 2 GHz to 11 GHz and has a phase noise of –80 dBc at 1MHz offset frequency with a power consumption of 19.4mW.



RP	5 kΩ
CP1	440 fF
CP2	240 fF
CP3	120 fF

Figure 4. The circuit schematic diagram of the poly-phase filter.

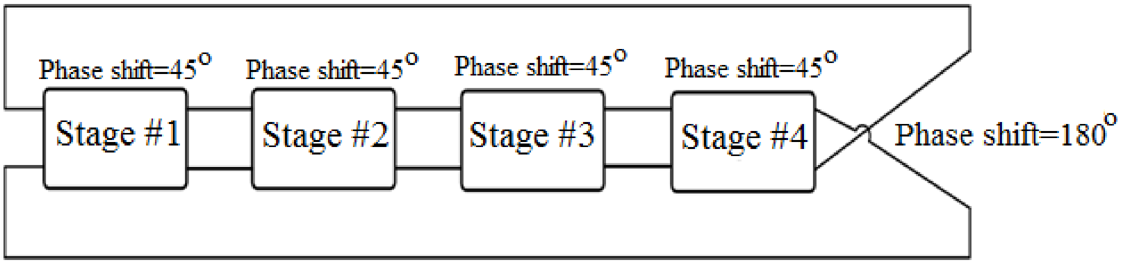
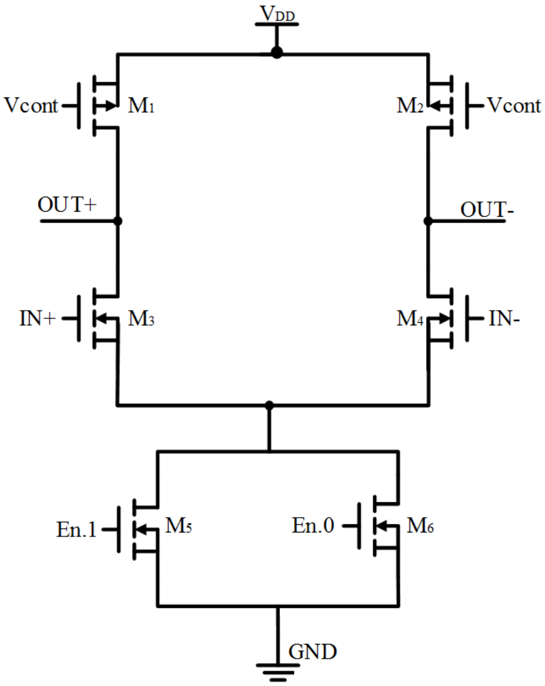


Figure 5. The four-phase LO generator block diagram.

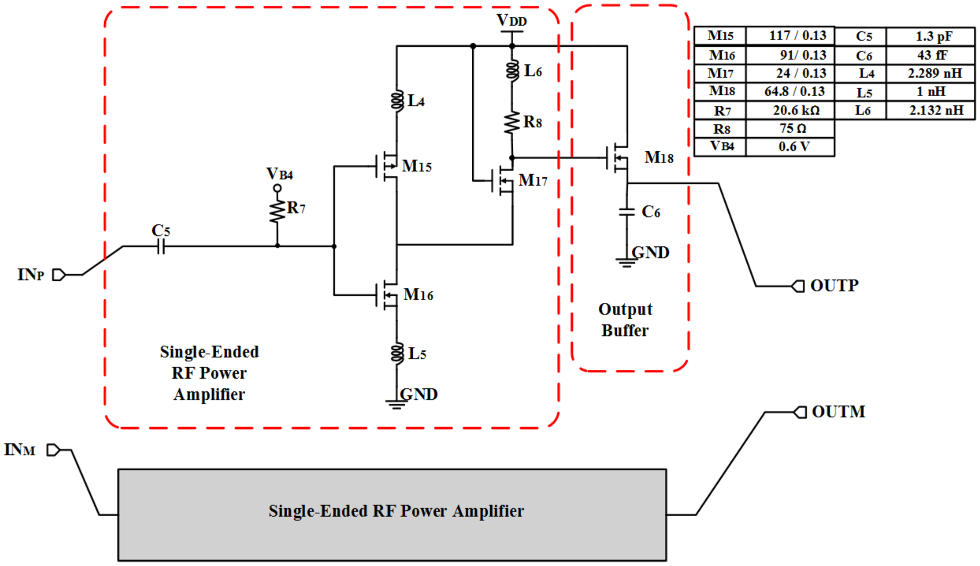
3.4 RF Power Amplifier

The circuit diagram of the RF power amplification stage is shown in Figure 7. In each amplifier, M15 and M16 are used as an amplifier with source degenerative inductors L4 and L5. Also the current-reuse technique is applied to (M15, M16) so a larger gain is achieved without an increase in power consumption. C5 is used for DC blocking. M17 is a common-gate amplifier following the input stage. Its main purpose is biasing the drains of (M15, M16), so that both stay in the saturation region. Inductance peaking techniques are also used here to improve the operating frequency [11]. M18 is designed as a buffer to drive a 50 ohm output load for the measurement purpose, and C6 represents the output pad parasitic capacitance. A single power amplifier and its output buffer drain 7.625 mA from a 1.2 V power supply.



M_1,M_2	24 / 0.13
M_3,M_4	20.0 / 0.13
M_5	8.6 / 0.13
M_6	4.6 / 0.28

Figure 6. The simplified schematic diagram of unit differential delay cell.



M_{15}	117 / 0.13	C_5	1.3 pF
M_{16}	91/ 0.13	C_6	43 fF
M_{17}	24 / 0.13	L_4	2.289 nH
M_{18}	64.8 / 0.13	L_5	1 nH
R_7	20.6 kΩ	L_6	2.132 nH
R_8	75 Ω		
V_{B4}	0.6 V		

Figure 7. The simplified schematic diagram of RF power amplifier.

4. Simulation Results

The proposed UWB transmitter with on-chip QVCO was designed, simulated and verified using Cadence Spectre Circuit Simulator in TSMC 130nm 1P8M CMOS technology. It consumes a total power of 77.8 mW. The transmitter covers the 14 bands of UWB MB-OFDM system, so that it can be used for UWB full-band application.

The design and optimization of UWB transmitter require precise RF modeling for both active and passive devices over a wide range of operation. Thus, in our simulation, we have used the RF models extracted from the physical layout.

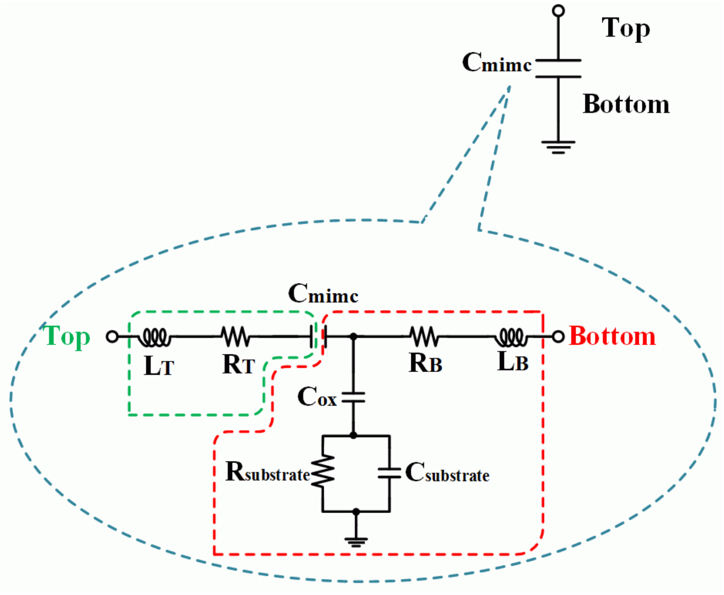


Figure 8. The RF model of the capacitor.

The capacitor RF model is shown in Figure 8. The main capacitance in the model is C_{mimc} which has two plates: the top plate and the bottom plate. Between the main capacitor and the top plate, there will be a parasitic inductance (L_T) and resistance (R_T), as well as on the bottom plate side (L_B) and (R_B). Since the bottom plate is close to the silicon substrate, this will create an oxide capacitance modeled as C_{OX}. $R_{substrate}$ and $C_{substrate}$ present the silicon substrate resistance and capacitance, respectively.

The QVCO tuning curves for the typical process are shown in Figure 9. From Figure 9, it can be seen that the QVCO covers all 14 bands in the UWB system. The QVCO works in two modes: the high frequency mode (when En.1 is enabled) which covers the frequency range from 3.28 GHz to 11.16 GHz and the low frequency mode (when En.1 is enabled) which covers the range from 1.5 GHz to 6.2 GHz. There is an overlap between each mode, which starts at 6.2 GHz and ends at 3.28 GHz with a total overlap of 2.9 GHz. This overlap is designed to be high enough in order to alleviate the frequency shift in the QVCO due to different corner processes (FF, SS, FS and SF). From the simulation, the critical cases for overlap frequency and QVCO frequency coverage range were in fast and slow process, as depicted in Figure 10 and Figure 11. From the results illustrated in Figure 10 and Figure 11, for fast and slow process, respectively, the QVCO still covers the entire band of the UWB system in both corners. The transient time simulation result of the proposed QVCO at 10 GHz under typical process is shown in Figure 12.

Table 1. The simulated power consumption in all circuit blocks of the proposed transmitter.

Transmitter	**Power**
QVCO	19.4 mW
Quadrature modulator core	40.1 mW
RF power amplifier	18.3 mW

Table 2. Performance and result comparison of published UWB transmitter.

	This work	[3]	[4]	[5]	[6]	[7]
Technology	130 nm CMOS	180 nm CMOS	90 nm CMOS	130 nm CMOS	130 nm CMOS	130 nm CMOS
Supply voltage	1.2 V	1.8 V	1.1 V	1.2 V	1.5 V	1.2 V
Bandwidth	2GHz-11GHz	3GHz-8GHz	3.1GHz-9.5GHz	3GHz-8GHz	3GHz-5GHz	3GHz-11GHz
Number of bands	14	9	12	9	3	14
OP_{-1dB}	4.35 dBm	–8.2 dBm	–2.8 dBm	1.5dBm	5 dBm	–0.4 dBm
Containing VCO	Yes	No	No	No	No	No
Power	77.8 mW	139 mW	131 mW	66 mW	97.5 mW	53.1 mW

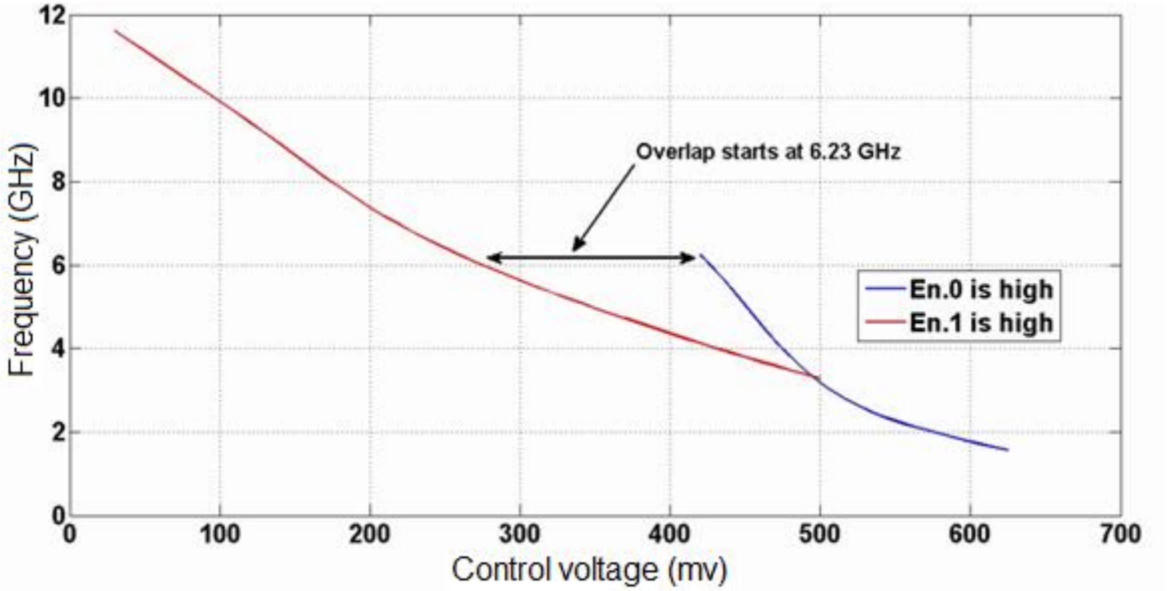


Figure 9. QVCO tuning curves, typical process.

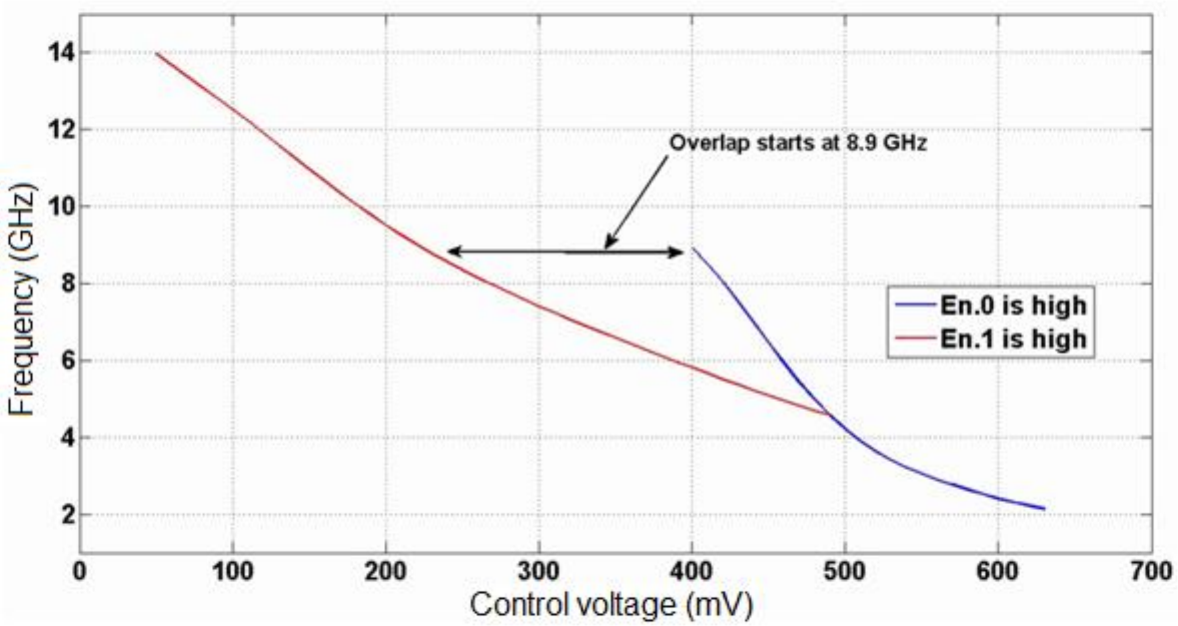


Figure 10. QVCO tuning curves, fast process.

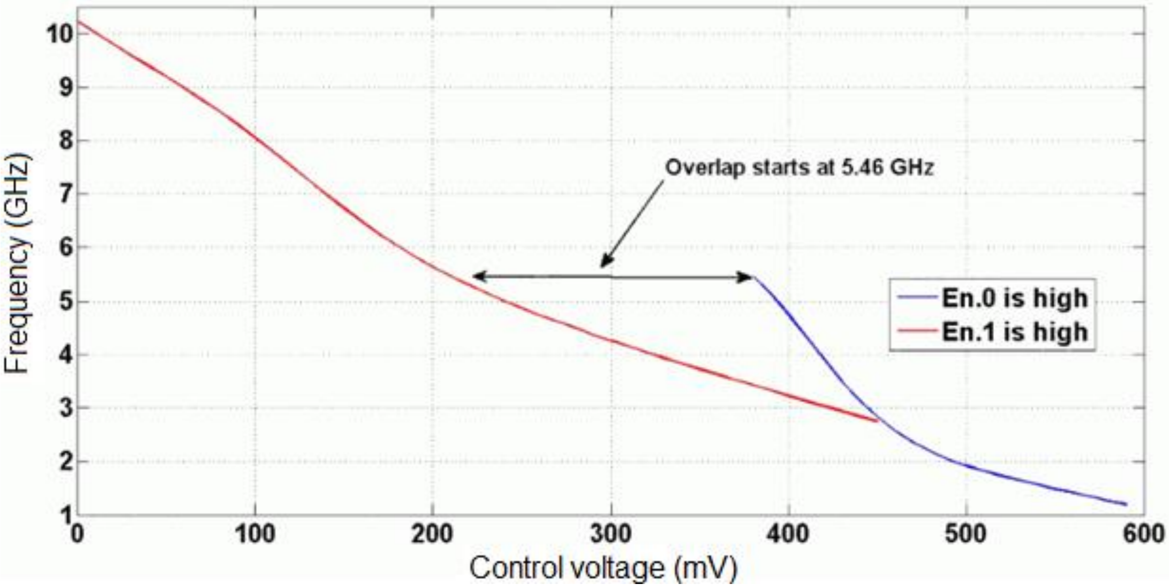


Figure 11. QVCO tuning curves, slow process.

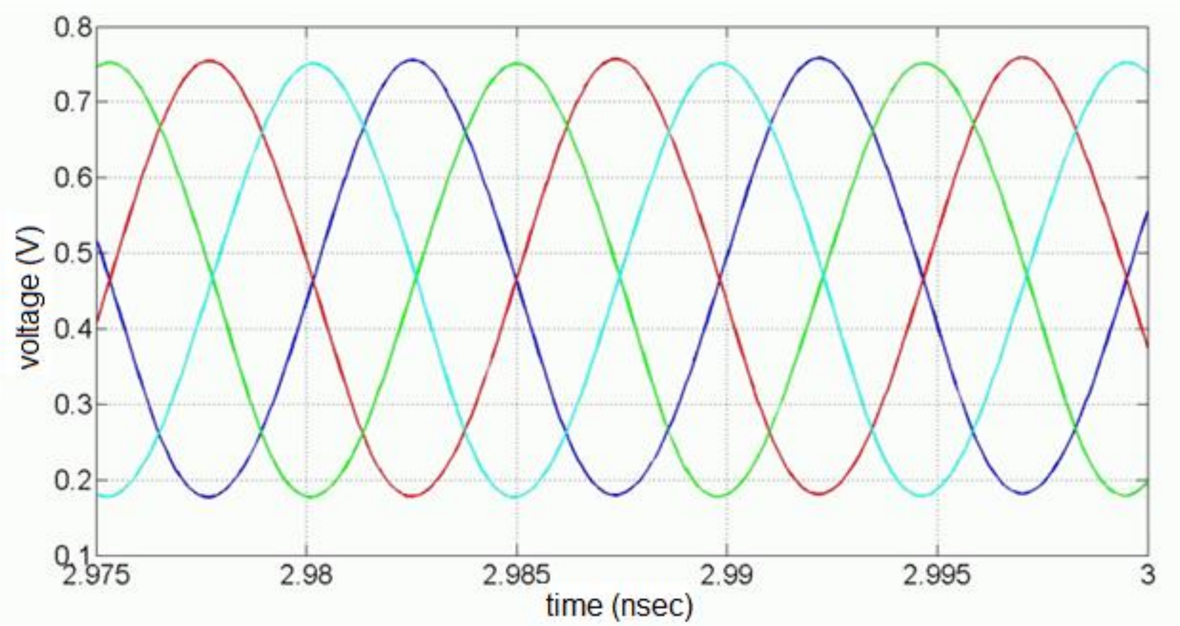


Figure 12. The transient time simulation of QVCO at 10 GHz for typical process, 25°C, $V_{cont.}$ = 0 V.

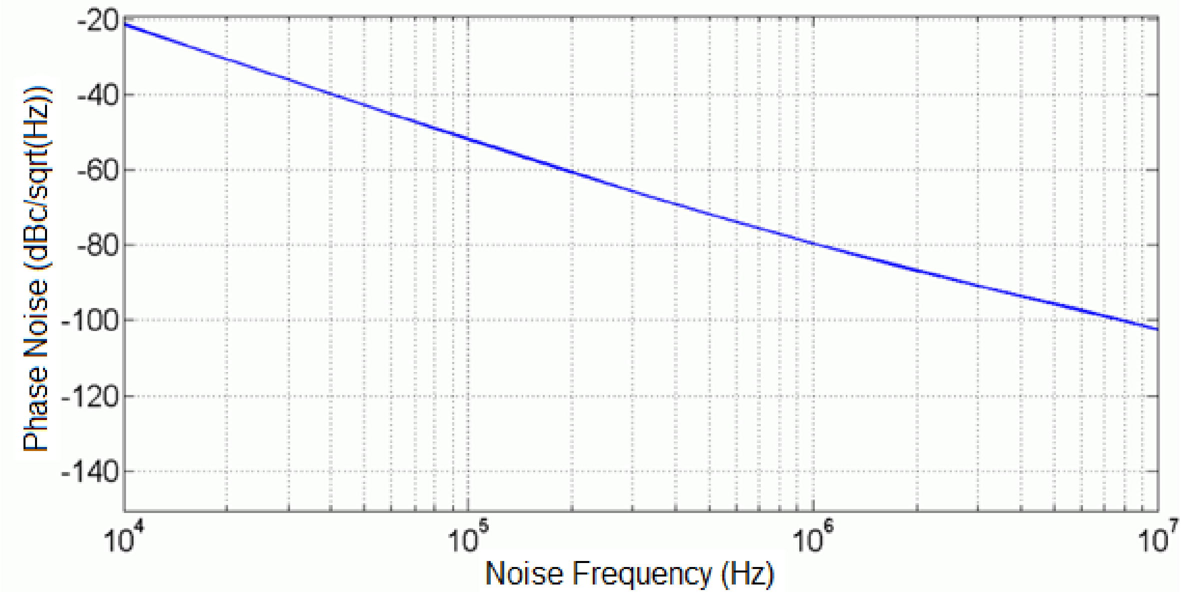


Figure 13. The simulated phase noise of the QVCO.

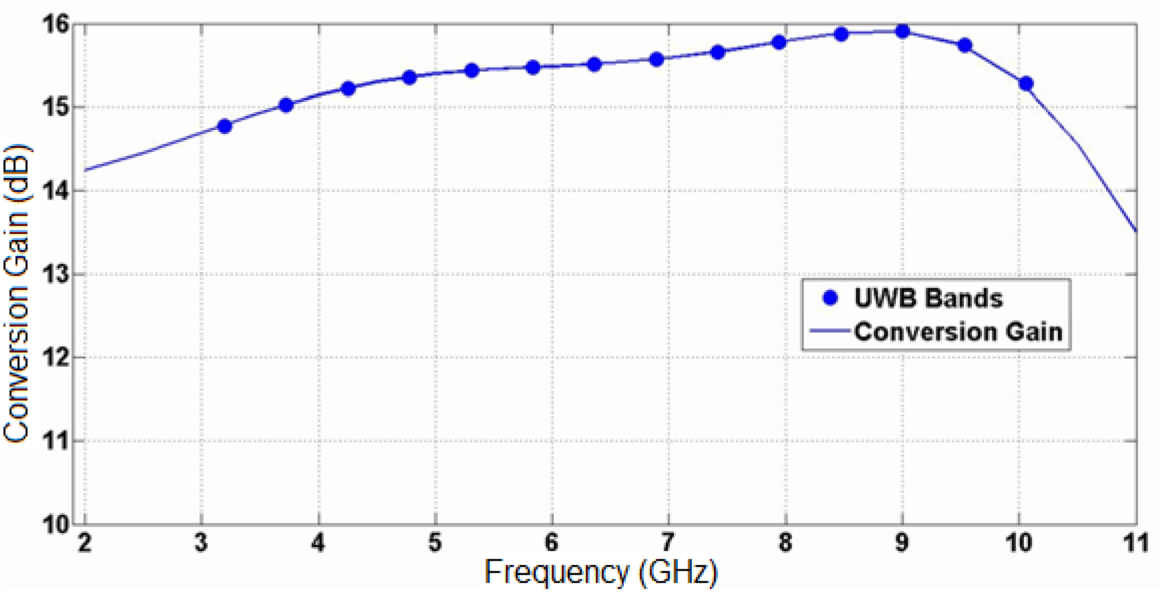


Figure 14. The simulated UWB transmitter conversion gain.

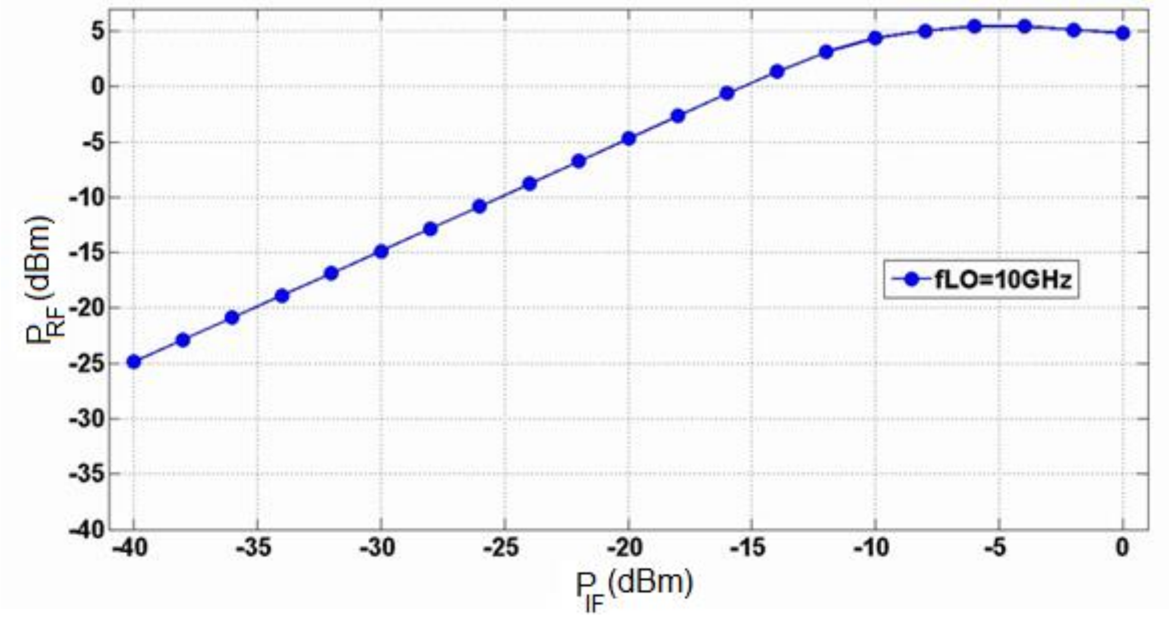


Figure 15. Linearity performance of the UWB transmitter at 10 GHz.

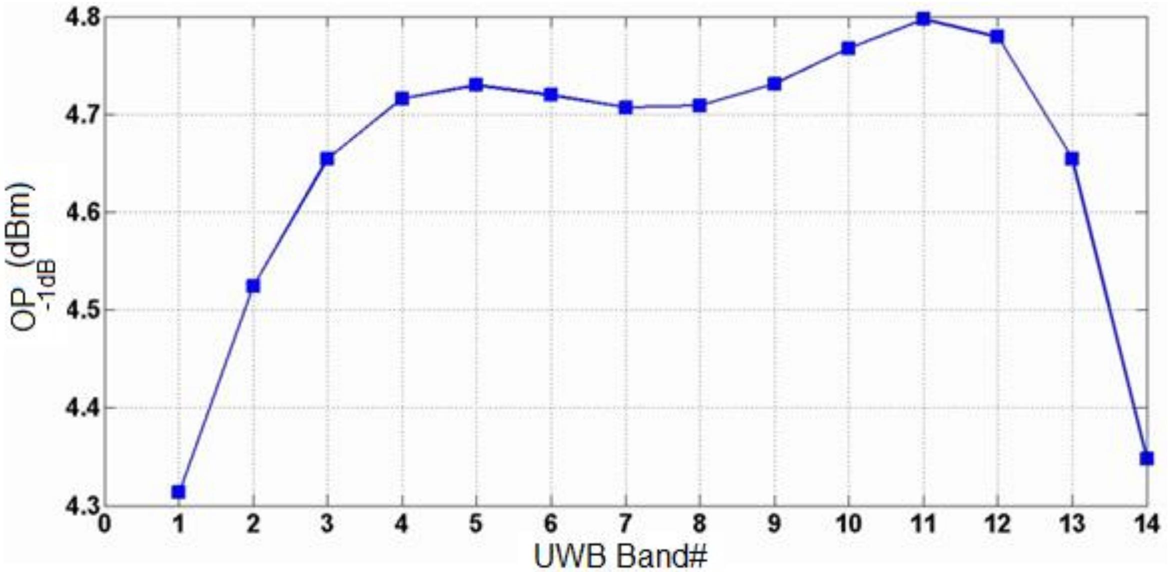


Figure 16. The simulated OP1dB of the UWB transmitter over each band.

The simulated phase noise of the proposed QVCO is shown in Figure 13. The worst case for phase noise of QVCO is –80 dBc at 1 MHz offset frequency with an operating frequency of 11 GHz.

The transmitter conversion gain in the frequency range from 2 GHz to 11 GHz is shown in Figure 14. As can be seen from Figure 14, the simulated average conversion gain of 15.28 dB is achieved with a gain ripple of ±1dB.

The simulated output power (PRF) *versus* the input IF power (PIF) is shown in Figure 15, where the corresponding IP1dB and OP1dB are –10 dBm and 4.35 dBm, respectively. The average simulated OP1dB is 4.65 dBm over the entire 14 bands as depicted in Figure 16.

The power consumption of each circuit block is shown in Table 1, while the simulated performance parameters are summarized in Table 2. In addition, comparisons with other published works are listed. Based upon Table 2, it is clear that the proposed transmitter with implemented QVCO covers the full-band of MB-OFDM (2 GHz-11 GHz) under low power dissipation. It drains 64.83 mA from the supply voltage of 1.2 V. Besides, the OP1dB achievement of the proposed UWB transmitter is the highest except [6], and it conforms with the specifications of UWB applications [12].

5. CONCLUSION

In this paper, a UWB full-band MB-OFDM transmitter with implemented QVCO is designed. The transmitter covers the frequency range from 2 GHz to 11 GHz and can cover all of the frequency bands of the UWB MB-OFDM system (14 bands) due to the use of inductance peaking technique. The simulation results have shown that the proposed transmitter can achieve a conversion gain of 15.28 dB with a ripple of ±1dB. In addition, the power dissipation of the proposed transmitter is 77.8 mW from a 1.2 V supply voltage. Future search will be conducted on the design of frequency synthesizer to reduce the QVCO phase noise and to control the output frequency [12]. In addition, the carrier leakage and the sideband suppression of the proposed transmitter will be explored.

ACKNOWLEDGMENT

The author would like to thank EUROPRACTICE for their kind support by providing the Cadence Spectre Circuit Simulator and the technology files.

REFERENCES

[1] R. J. Fontana, "Recent System Applications of Short-pulse Ultra-wideband (UWB) Technology," IEEE Transactions on Microwave Theory and Techniques, vol. 52, pp. 2087-2104, 2004.

[2] F. W. Chia-Chin Chong and H. Inamura, "Potential of UWB Technology for the Next Generation Wireless Communications," in: Proc. IEEE 9th International Symposium on Spread Spectrum Techniques and Applications, Manaus-Amazon, pp. 422-429, 2006.

[3] D. L. C. S. Hui Zheng, S. Lou and L. Tatfu Chan, "A 3.1 GHz-8.0 GHz Single-Chip Transceiver for MB-OFDM UWB in 0:18m CMOS Process," IEEE Journal of Solid-State Circuits, vol. 44, pp. 414-426, 2009.

[4] H. K. A. Tanaka, H. Okada and H. Ishikawa, "A 1.1V 3.1 GHz-9.5GHz MB-OFDM UWB Transceiver in 90nm CMOS," in: Proc. IEEE International Solid-State Circuits Conference, p.398-407, San Francisco, CA, 2006.

[5] H.-Y. Shih and C.-W. Wang, "A Highly-Integrated 38 GHz Ultra-Wideband RF Transmitter with Digital-Assisted Carrier Leakage Calibration and Automatic Transmit Power Control," IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 20, pp. 1357-1367, 2012.

[6] E. A. C. Sandner, "A WiMedia/MBOA-Compliant CMOS transceiver for UWB," IEEE Journal of Solid-State Circuits, vol. 41, pp. 2787-2794, 2006.

[7] Y.-K. L., Z.-D. H., Fadi R. Shahroury, Wen-Chieh Wang, Chang-Ping Liao and C.-Y. Wu, "The Design of Integrated 3-GHz to 11-GHz CMOS Transmitter for Full-Band Ultra-Wideband (UWB) Applications," in: Proc. IEEE International Symposium on Circuits and Systems, Seattle, WA, pp. 2709-2712, May 2008.

[8] W. Alliance, "Multi-band OFDM Physical Layer Specication," WiMedia Alliance, Tech. Rep. Release 1.1, 2005.

[9] B. Gilbert, "A Precise Four Quadrant Multiplier With Sub-nanosecond Response," IEEE Journal of Solid-State Circuits, vol. 3, pp. 365-373, 1968.

[10] C.-Y. Chou and C.-Y. Wu, "The Design of Wideband and Low-power CMOS Active Polyphase Filter and Its Application in RF Double Quadrature Receiver," IEEE Transactions on Circuits and Systems I, vol. 52, pp. 825-833, May 2005.

[11] T. H. Lee, The Design of CMOS Radio-Frequency Integrated Circuits, Second Edition, 2003.

[12] E. A. B. Razavi, "A UWB CMOS Transceiver," IEEE Journal of Solid-State Circuits, vol. 40, pp. 2555-2562, 2005.

"A 1.2-V Low-power Full-band Low-power UWB Transmitter with Integrated Quadrature Voltage-controlled Oscillator and RF Amplifier in 130nm CMOS Technology", Fadi R. Shahroury.

ملخص البحث:

تقدم هذه الورقة البحثية تصميماً ومحاكاةً لمرسلٍ منخفض القدرة كامل النطاق في مجال النطاق فائق الاتساع (UWB) مع مذبذب رباعي مُتحكّم به بفرق الجهد (QVCO)، يستخدم تكنولوجيا أشباه الموصلات ذات الأكسيد المعدني المتممة (CMOS)/١٣٠ نانومتراً. ويتكوّن المرسل المقترح من مرشح سلبي متعدد الأطوار (PPF)، ومذبذب رباعي مُتحكَّم به بفرق الجهد (QVCO)، وقلب معدِّل رباعي، ومضخم قدرة للإشارات ذات الترددات الراديوية (RF).

ويستخدم المذبذب بنية خلايا التأخير التفاضلية بأربع مراحل متعاقبة. ويمتلك المرسل المقترح المواصفات الآتية: كسب تحويل متوسط مقداره ١٥,٢٨ ديسيبل بتعرُّج مقداره ± ١ ديسيبل في النطاق الترددي الممتدّ من ٢ جيجاهيرتز إلى ١١ جيجاهيرتز؛ نقطة ضغط الدّخل (IP1dB) تساوي -١٠ ديسيبل (dBm)؛ نقطة ضغط الخرج (OP1dB) تساوي ٤,٣٥ ديسيبل (dBm).

ويحقق المذبذب مدى ترددياً واسعاً يتراوح من ٢ جيجاهيرتز إلى ١١ جيجاهيرتز بضجيج طور مقداره -٨٠ ديسيبل (dBc)/هيرتز. إضافة إلى ذلك، فإن فرق جهد التشغيل للمرسل المقترح يبلغ ١,٢ فولت، في حين تبلغ القدرة التي يستهلكها ٧٧,٨ ميلي واط.

HIGHER LEVEL SECURITY APPROACH FOR DATA COMMUNICATION SYSTEM BASED ON AES CRYPTOGRAPHY AND DWT STEGANOGRAPHY

Saja M. Saraireh[1] and Aser M. Matarneh[2]

Department of Electrical Engineering, Mutah University, Jordan
aser_m2002@yahoo.com[2], aser.matarneh@mutah.edu.jo[2]

(Received: 27-Mar.-2016, Revised: 23-Jun.-2016, Accepted: 26-Jun.-2016)

ABSTRACT

Cryptography is used for secured data transmission, but the resulting unreadable messages usually attract other's attention, so steganography is employed to hide the secret information to prevent attackers from discovering the presence of secret data. This paper proposes an improved technique that combines both Advanced Encryption Standard (AES) algorithm for cryptography and steganography and takes the advantages of using the high frequency coefficients of the cover image by applying the Discrete Wavelet Transform (DWT).The proposed technique is employed to study the effect of hiding the encrypted secret message in 24-bit RGB image. The performance of the proposed method is evaluated in terms of the Peak Signal to Noise Ratio (PSNR) analysis, the payload embedding capacity and the histogram distribution analysis. Comparison to other four associated works will be offered. Experimental results reveal that the proposed method gives a secure technique for data hiding and shows robustness against different attacks.

KEYWORDS

DWT, Steganography, AES, Cryptography, PSNR, Payload, Histogram.

1. INTRODUCTION

Security problems become an essential issue due to exchanging large amounts of data on computer networks [1]. This has resulted in the appearance of a lot of applications that are concerned with the data hiding field, such as; cryptography, watermarking, fingerprinting and steganography. The most two preferred techniques are cryptography and steganography [2]-[3]. Although these two are related, there is a difference between them; cryptography encrypts the secret text, but its presence is still detectable, so that the message can be intercepted and altered by the attacker. In contrast, steganography conceals the existence of the secret data in another medium [4]-[5], so that the development of steganography provides secure transfer of data without stimulating any doubt, where different techniques for data hiding are used to hide the secret data into a cover medium [6]-[7].

Steganography methods are classified into; spatial domain-based steganography and frequency domain-based steganography [8]. Spatial domain techniques use the pixel's least significant bits (LSBs) to embed the secret message bits. The resulting stego-image is susceptible to many noisy operations because of the simplicity of LSB techniques [9]. Frequency domain techniques utilize the properties of the cover image and then the steganography robustness is improved [5].

There have been a lot of works using the combination of both cryptography and steganography or one of them. In [10], DWT is used to embed the secret message in a gray image. It provides high stego-image quality although it embeds large capacity of data. In [11], DWT steganography method was used, in which the data is hidden in the main components of the sub-image. These are; the horizontal, the vertical and the diagonal components.

In [12], a method was presented that uses a combined application of steganography and visual cryptography. It provides the protection for the customer identity, then a limited amount of information is given for money transfer through the shopping on internet. In [13], AES Rijndeal method and secret key steganography are combined to provide higher security and higher hidden data rate. In AES, the message is powerfully encrypted. By steganography, Discrete Cosine Transform (DCT) of an image is used, where two secret keys are generated in such a manner that makes the system highly secured. In [14], a symmetric key RSA and symmetric key AES algorithms are employed to encrypt the data, then LSB steganography technique is used to embed the encrypted data into the cover image. A combination of RSA algorithm and DCT with LSB technique is proposed in [15], where the message is encrypted using RSA, then embedded using DCT with LSB in digital media. Custom neural network is used for extracting the encrypted data.

The main aim of AES encryption/decryption process integrated with steganography is to offer a high level of security with moderate capacity albeit the complexity added by such model. The system would become robust against attacks if the security issue is properly tackled.

The proposed algorithm is based on DWT using colored digital images and AES encryption techniques, whereas the similar algorithm has been proposed in [16] for gray images based on Double Density Dual Tree (DD DT) DWT using AES and T-codes. However, although the proposed algorithm in [16] offers a higher degree of security, the quality of the stego-image is still a critical issue.

In this paper, the advanced encrypted standard (AES) cryptography technique is employed to encrypt the secret message, and (DWT) steganography is used to embed the encrypted secret message in the cover image, with utilizing the advantages of 24-bit RGB image as it is a cover medium and using the pixel indicator technique (PIT) to embed the encrypted secret message in the cover image which is proposed in [17]. The proposed method is evaluated by comparing it to other four methods. It achieves higher Peak Signal to Noise Ratio (PSNR) values and makes a trade-off between capacity and security. The rest of the paper is organized as follows; in section 2 the proposed system is introduced, while section 3 presents and discusses the experimental results and then, the paper is concluded in section 4.

2. Proposed System

The main idea of the proposed algorithm is to combine both AES cryptography and DWT steganography, with utilizing the advantages of 24-bit RGB images by using PIT. Next sub-sections briefly introduce these techniques.

2.1 Two-Dimensional Haar-DWT

The operations of the two-dimensional Haar-DWT result of four different sub-bands are denoted as LL, HL, LH and HH, as shown in Figure 1. These four sub-bands represent four sub-images of the same size (M/2, N/2) that are obtained from an image of size (M, N) [9, 19], as shown in Figure 2. Some information is obtained from these sub-bands (LL, HL, LH, HH), since the cover image pixels are classified into less important and more important pixels [5]. According to this, the LL sub-band is not used for embedding, since it contains vital information about the original image, and any changes in it will affect the quality of the image and can be easily noticed by the human eye [20].

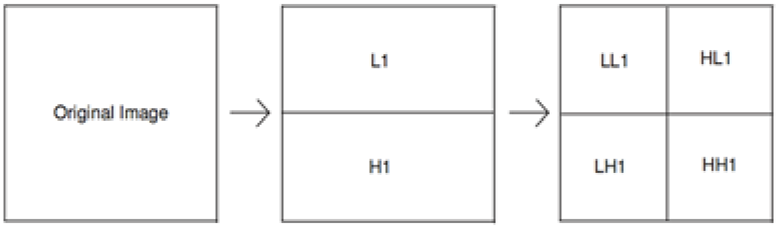


Figure 1. The horizontal operation and the vertical operation of DWT.

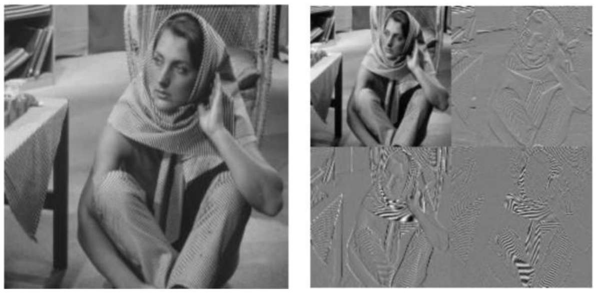


Figure 2. The original image to the left and its DWT to the right.

2.2 Pixel Indicator Technique (PIT)

Pixel indicator technique was used for 24-bit RGB image steganography. The RGB channels of the image are classified by this technique to be: the indicator, channel 1 and channel 2 in specified order. The 2 LSBs of the indicator channel are used to specify the embedding of data in channel 1 and channel 2 [17]. Table 1 shows this technique. An improvement on this technique is proposed in [20], where instead of embedding only 2 bits in the channel selected for the embedding, (1, 2 or 3) bits can be embedded in that channel according to the number of zeros in the most significant part MSB, where 1 bit is embedded if the number of zeros in the MSB is 0 or 4, 2 bits are embedded if the number of zeros in the MSB is 2 and 3 bits are embedded if the number of zeros in the MSB is 1 or 3, as shown in Table 2. The proposed technique employs this method in the transform domain instead of using it in the spatial domain to improve security.

2.3 The Embedding Process

As the proposed technique idea is derived from [17] and [21], some improvements should be mentioned, where instead of embedding the secret text clearly, it is embedded after encrypting it by the AES algorithm, then as an alternative of using the spatial domain of the cover image for hiding, the transform domain is employed, and finally, the location of the indicator channel is not fixed, where it depends on the transform band used for embedding.

In the proposed algorithm, the cover image is firstly separated into its RGB color components, then the DWT is applied on each component separately. For the secret text message, the first step is to encrypt it by the AES algorithm, and convert it into its binary representation. The next step is to store its length to variable X and in the first 8 pixels of HL band in the Red channel. To improve security by using the PIT method, the indicator channel is variable. The indicators are chosen according to the transform band that is used to embed the encrypted secret text in it. In the HL band, the Red channel is the indicator. Green and Blue are channel one and channel two, respectively. In the LH band, the Green channel is the indicator. Red and Blue are channel one and channel two, respectively. In the HH band, the Blue channel is the indicator. Red is channel one and Green is channel two.

Table 1. The relation between the indicator and other channels [17].

The 2 LSBs of the indicator	Channel 1 (green channel)	Channel 2 (blue channel)
00	No hidden data	No hidden data
01	No hidden data	2 bits of hidden data
10	2 bits of hidden data	No hidden data
11	2 bits of hidden data	2 bits of hidden data

Table 2. The hiding process [21].

Number of zeros in the MSB	LSB			
	b5	b6	b7	b8
0 or 4	×	×	×	✓
2	×	×	✓	✓
1 or 3	×	✓	✓	✓

The embedding process is started from pixel nine in the HL band of the Red channel. The embedding of the encrypted secret message in the selected channel is executed by the hiding process, as shown in Table 2. The sequence of the embedding algorithm is flowcharted in Figure 3. The process is stopped when the secret message is completely embedded. When the embedding process is completed, the inverse discrete wavelet transform (IDWT) is applied for each RGB channel, then the RGB components are combined in order to obtain the stego-image.

2.4 The Extraction Process

The extraction process is flowcharted in Figure 4. When the secret text message is extracted, which is in binary representation, it is converted into its character representation and decrypted by the AES algorithm. The algorithm will stop based on the length of the secret message which is stored in the first 8 bytes of the Red channel in the HL band.

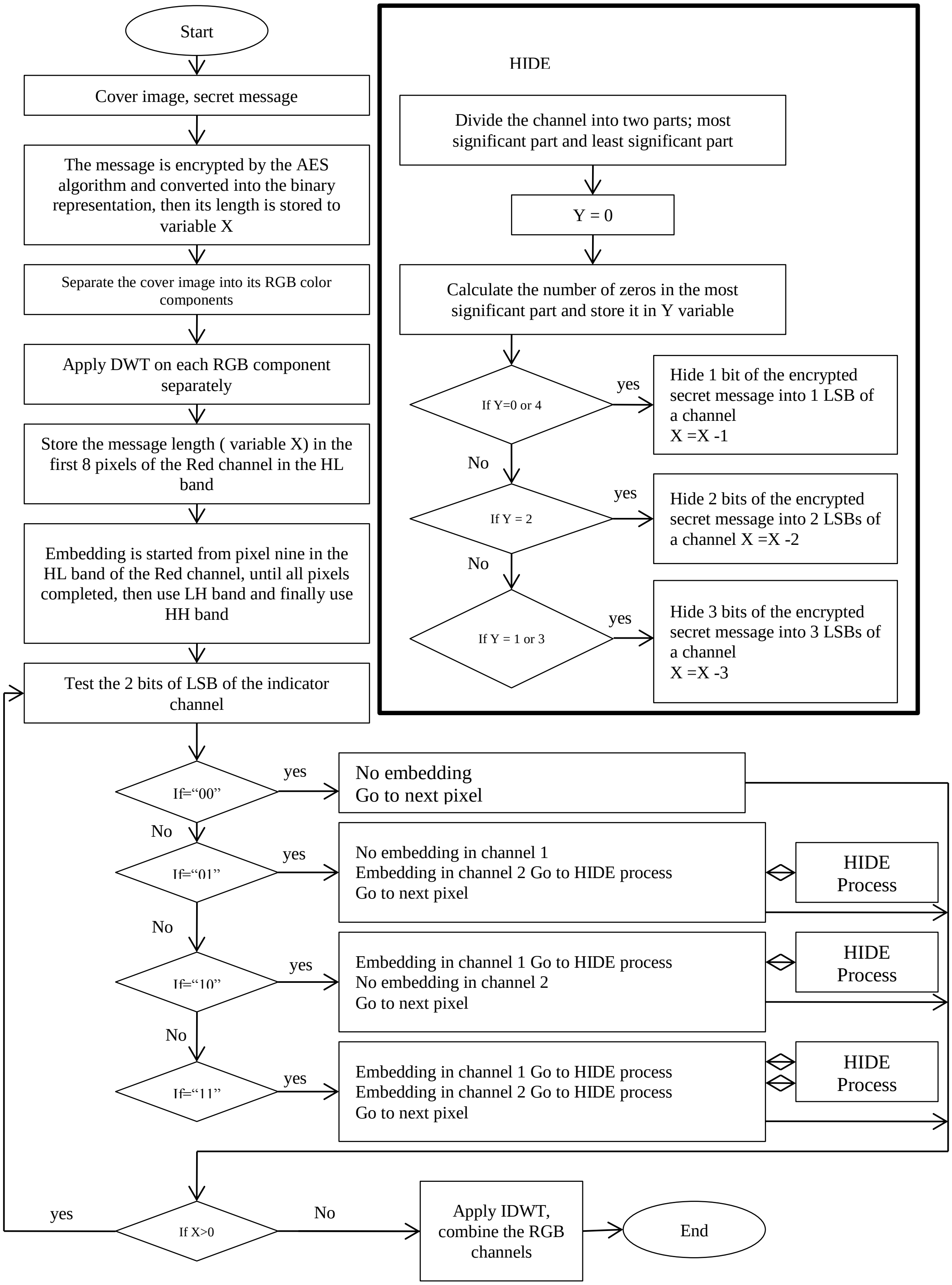

Start
Cover image, secret message
The message is encrypted by the AES algorithm and converted into the binary representation, then its length is stored to variable X
Separate the cover image into its RGB color components
Apply DWT on each RGB component separately
Store the message length (variable X) in the first 8 pixels of the Red channel in the HL band
Embedding is started from pixel nine in the HL band of the Red channel, until all pixels completed, then use LH band and finally use HH band
Test the 2 bits of LSB of the indicator channel
If=“00”
yes
No embedding
Go to next pixel
No
If=“01”
yes
No embedding in channel 1
Embedding in channel 2 Go to HIDE process
Go to next pixel
HIDE Process
No
If=“10”
yes
Embedding in channel 1 Go to HIDE process
No embedding in channel 2
Go to next pixel
HIDE Process
No
If=“11”
yes
Embedding in channel 1 Go to HIDE process
Embedding in channel 2 Go to HIDE process
Go to next pixel
HIDE Process
yes
If X>0
No
Apply IDWT, combine the RGB channels
End
HIDE
Divide the channel into two parts; most significant part and least significant part
Y = 0
Calculate the number of zeros in the most significant part and store it in Y variable
If Y=0 or 4
yes
Hide 1 bit of the encrypted secret message into 1 LSB of a channel
X =X -1
No
If Y = 2
yes
Hide 2 bits of the encrypted secret message into 2 LSBs of a channel X =X -2
No
If Y = 1 or 3
yes
Hide 3 bits of the encrypted secret message into 3 LSBs of a channel
X =X -3

Figure 3. The embedding process flowchart.

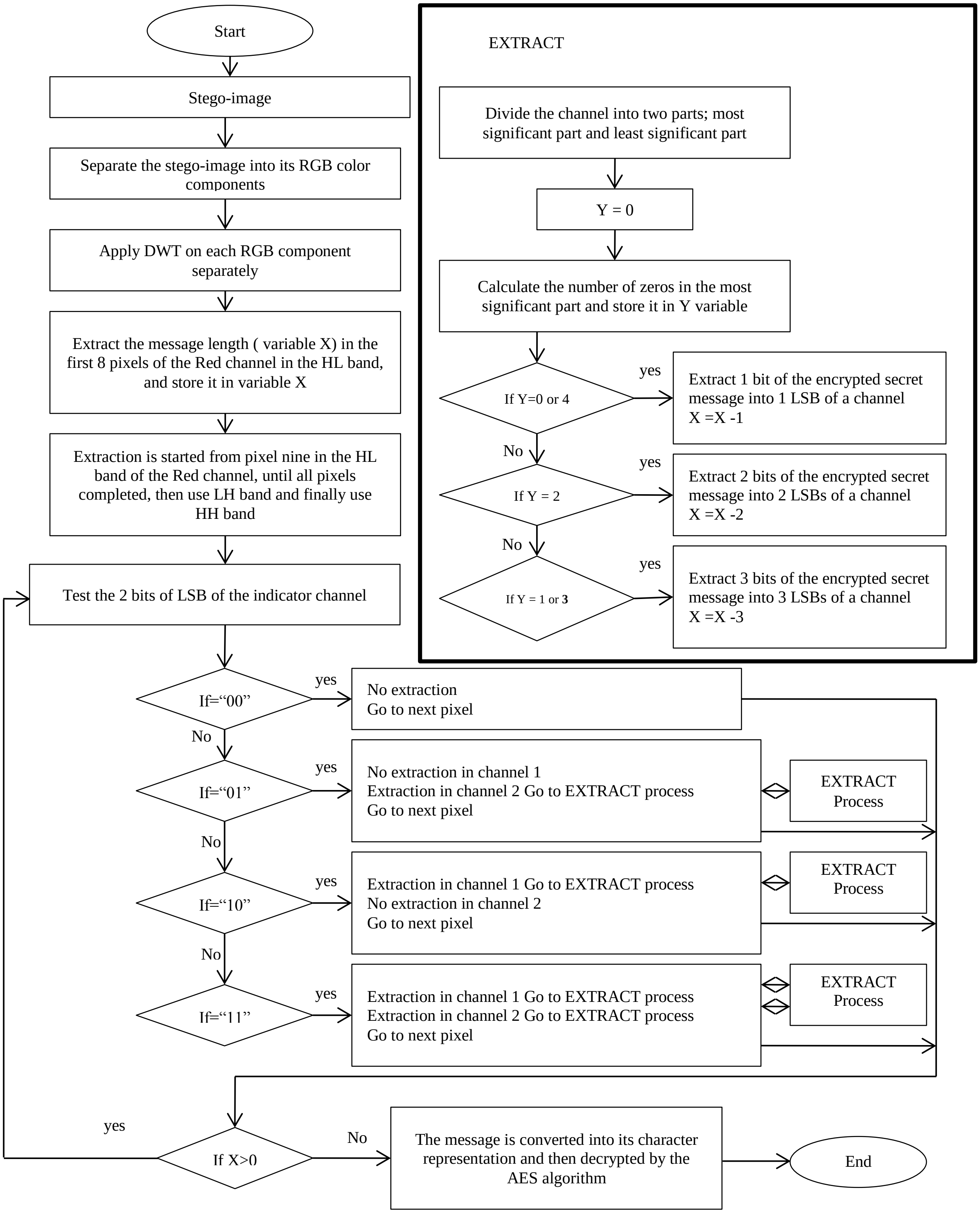

Start
Stego-image
Separate the stego-image into its RGB color components
Apply DWT on each RGB component separately
Extract the message length (variable X) in the first 8 pixels of the Red channel in the HL band, and store it in variable X
Extraction is started from pixel nine in the HL band of the Red channel, until all pixels completed, then use LH band and finally use HH band
Test the 2 bits of LSB of the indicator channel
EXTRACT
Divide the channel into two parts; most significant part and least significant part
Y = 0
Calculate the number of zeros in the most significant part and store it in Y variable
If Y=0 or 4
yes
Extract 1 bit of the encrypted secret message into 1 LSB of a channel X =X -1
No
If Y = 2
yes
Extract 2 bits of the encrypted secret message into 2 LSBs of a channel X =X -2
No
If Y = 1 or 3
yes
Extract 3 bits of the encrypted secret message into 3 LSBs of a channel X =X -3
If="00"
yes
No extraction Go to next pixel
No
If="01"
yes
No extraction in channel 1 Extraction in channel 2 Go to EXTRACT process Go to next pixel
EXTRACT Process
No
If="10"
yes
Extraction in channel 1 Go to EXTRACT process No extraction in channel 2 Go to next pixel
EXTRACT Process
No
If="11"
yes
Extraction in channel 1 Go to EXTRACT process Extraction in channel 2 Go to EXTRACT process Go to next pixel
EXTRACT Process
yes
If X>0
No
The message is converted into its character representation and then decrypted by the AES algorithm
End

Figure 4. The extraction process flow chart.

3. Experimental Results

The proposed DWT-based image data hiding technique is implemented using Matlab. To evaluate the performance of the proposed scheme, a number of different types of colored image formats (BMP, JPEG and TIF) are tested. Peak signal to noise ratio PSNR is used for comparing the quality of stego-images with the original images, the payload for embedding capacity and the histogram have been shown to check the degradation of the quality of images. Further, some image processing operations are applied to check robustness. However, the results of only three RGB over images are included in this paper. In addition to that, the effect of increasing the embedded text size on the PSNR and the mean square error MSE is also tested.

The employed images are colored images with a size of 1024×1024x3. Figures (5, 6 and 7) show the original images and the stego-images after embedding an encrypted secret text with a size of 2120 bytes.

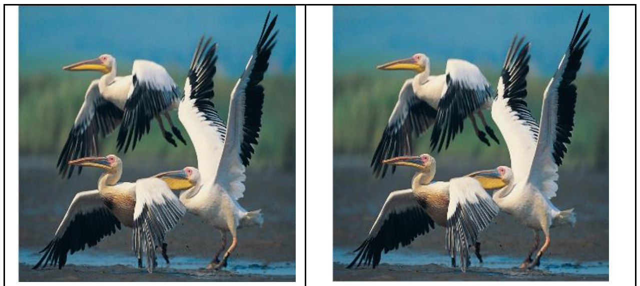


Figure 5. (a) cover-image for (image1) with a size of 1024X1024x3

Figure 5. (b) stego-image for (image1) with a size of 1024X1024x3

Figure 5. (a, b) show cover-image, stego-image after embedding 2120 bytes inside (image 1) picture by the proposed algorithm.

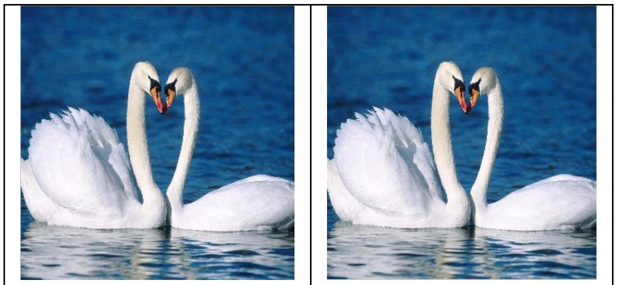


Figure 6. (a) cover-image for (image2) with a size of 1024X1024x3

Figure 6. (b) stego-image for (image2) with a size of 1024X1024x3

Figure 6. (a, b) show cover-image, stego-image after embedding 2120 bytes inside (image 2) picture by the proposed algorithm.

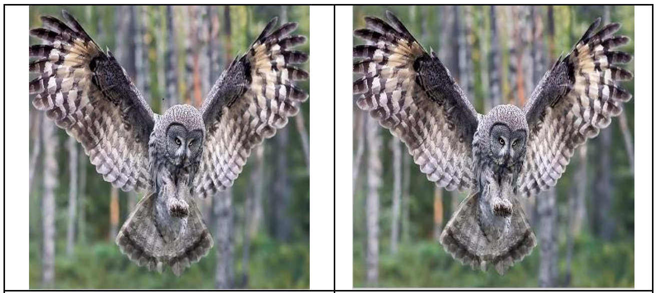


Figure 7. (a) cover-image for (image3) with a size of 1024X1024x3

Figure 7. (b) stego-image for (image3) with a size of 1024X1024x3

Figure 7. (a, b) show cover-image, stego-image after embedding 2120 bytes inside (image 3) picture by the proposed algorithm.

The stego-images are looking intact, which means that the proposed algorithm provides high quality of the stego-image. This can be measured with some tests. The PSNR test measures the image quality by comparing the original image with the stego-image. PSNR can be obtained using Equation 1 [22]. As a high level of security can be obtained for high PSNR values, the human eye will not be able to discriminate between the original image and the stego-image, because the stego-image will be very similar to the original cover image and the attacker would not detect the hidden information. The PSNR values of the proposed system for different images are summarized in Table 3, including a comparison with other ones that use 24-bit color images ([17] and [21]), which are described briefly in section 2.3, in addition to [23] which includes an embedding process that depends on calculating the number of zeros and ones in the (red) indicator channel, then obtaining the absolute difference between them. The resulting value is used to determine the number of bits that should be embedded in channel 1 and channel 2, as well as another technique called simple DWT-based steganography, which embeds one bit of the secret text on each LSB of the pixels in the high frequency bands (HL, LH, HH) of the cover image. Simple DWT is added for the comparison purpose where it is not related to the specified work, however, it is related to similar works such as [24]-[25].

$$\text{PSNR} = 10log_{10}\frac{255^2}{\text{MSE}}\ (\text{dB}) \qquad (1)$$

The mean square error (MSE) is :

$$\text{MSE} = \left[\frac{1}{N\times N}\right]^2 \sum_{i=1}^{N}\sum_{j=1}^{N}\left(X_{ij} - \bar{X}_{ij}\right)^2 \qquad (2)$$

where:

N is the image size;

X_{ij} represents the original image pixels;

$\bar{\text{X}}_{\text{ij}}$ represents the stego-image pixels.

Table 3. The PSNR results.

Image	**Image Size**	**Gutub *et al.* [17]**	**Ghosal *et al.* [23]**	**Yazan *et al.* [21]**	**Simple DWT based Steganography**	**Proposed Algorithm**		
						BMP	**JPEG**	**TIFF**
Image 1	1024×1024 x3	57.73	56.29	57.94	74.1793	68.5244	68.0753	68.0791
Image 2	1024×1024 x3	58.10	56.68	58.20	73.9965	64.6591	68.4558	68.4558
Image 3	1024×1024 x3	57.90	57.02	57.93	74.0210	67.9923	67.6653	67.6653

It can be noted from Table 3 that the proposed scheme showed better results for all image types (BMP, JPEG and TIFF) when compared to the techniques in [17], [21] and [23], where it provides higher PSNR values with an average difference equal to 8 dB approximately, which makes the stego-image indistinguishable from the cover image, except the simple DWT-based steganography which gives higher PSNR values than the proposed technique. However, this does not mean that it is better, because the embedding procedure for the simple DWT is very easy, where the secret text can be obtained only by applying the DWT on the stego-image, then removing the 7left-most bits from each byte, where that threatens its security. The proposed scheme gives better security, because the steganography embedding procedure is difficult to be expected, in addition to that the indicator is not the same in the three high frequency bands (HL, LH and HH). Also, in case of the extraction of the embedded text, it becomes difficult to obtain the secret message due to high security added by the AES algorithm. As a result, AES encryption technique provides higher quality of the stego-image, but at the cost of higher complexity.

Table 4 shows the data load which can be embedded inside different loads of the images. It represents the maximum amount of secret data in bytes which can be embedded in each image.

Table 4. The payload results.

Image	**Image Size**	**Gutub et al,.. [17]**	**Ghosal's et al,.. [23]**	**Yazan et al,.. [21]**	**Simple DWT based Steganography**	**Proposed Algorithm**		
						BMP	**JPEG**	**TIFF**
		Hiding Capacity (Bits)	Hiding Capacity (Bits)	Hiding Capacity (Bits)	Hiding Capacity (Bits)	Hiding Capacity (Bits)	Hiding Capacity (Bits)	Hiding Capacity (Bits)
Image 1	1024×1024x3	1583486	1811022	1833746	2359272	485021	477122	477122
Image 2	1024×1024x3	1575400	1869612	2087466	2359272	700158	655622	655622
Image 3	1024×1024x3	1572370	1456352	1572900	2359272	649353	632528	632528

Table 4 presents the payload which is defined as "the maximum message size that can be embedded subject to certain constraints" [26]. The proposed algorithm has been tested using three different color image types (BMP, JPEG and TIFF). The hiding capacity is better in case of TIFF image than in BMP or JPEG. It is obvious that all the other works give higher payloads than the proposed algorithm. This is related to two reasons that are; using only three bands of the cover image (HL, LH and HH), which means that 25% of the image is not used for the embedding process, as well as using the PIT, where the indicator channel is not employed for hiding the secret message. However, high PSNR values can compensate this restriction -which will be explained later- where a trade-off between payload and PSNR is obtained, as shown in Figure 11, which resulted in high security with moderate capacity.

The degradation of quality of images can also be visually noticed by applying the histogram analysis that depends on the comparison between the cover image and the stego-image through the statistical tool histogram shown in Figures (8-10) which present the histogram comparison between the cover image and its corresponding stego-image, where the histograms are calculated for Red, Green and Blue channels separately. It can be shown that there are no visual changes between the original image histograms and the stego-image histograms that are detected, so that the proposed scheme becomes superior to other schemes in terms of high degree of security with moderate capacity.

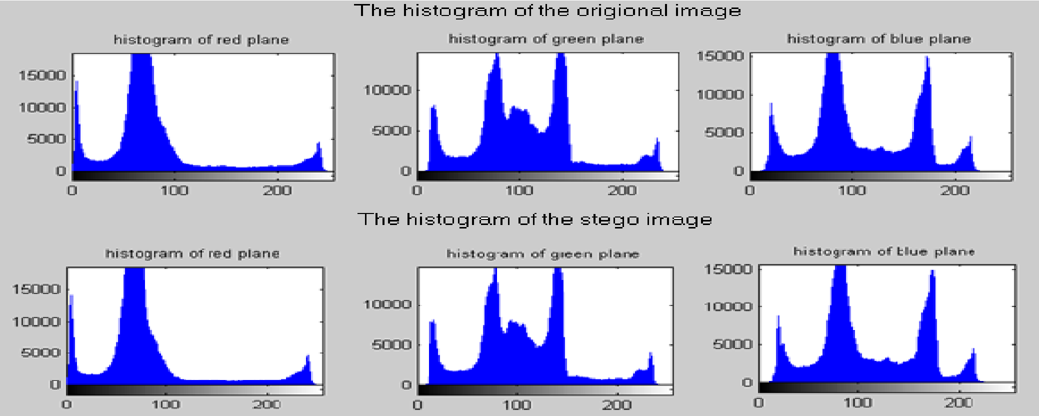


Figure 8. The histogram of the Red, Green and Blue channels of image 1. The upper part is for the original image and the bottom one is for the stego-image.

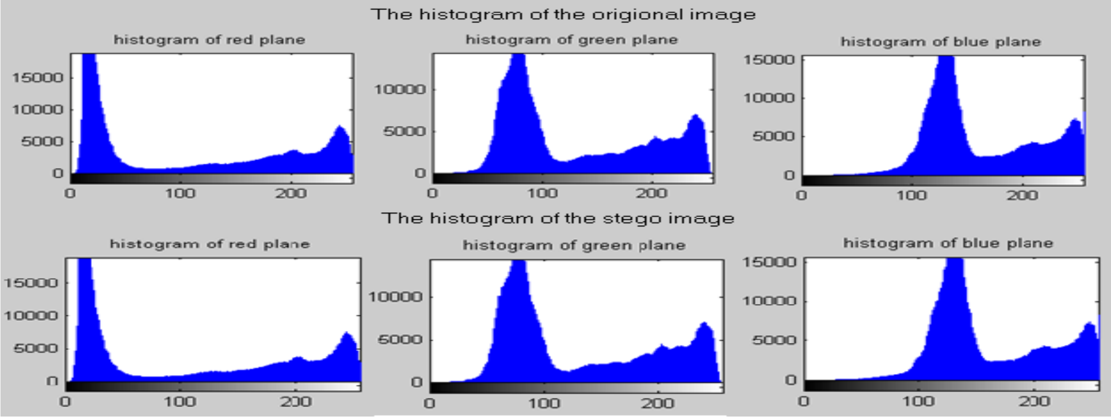


Figure 9. The histogram of the Red, Green and Blue channels of image 2. The upper part is for the original image and the bottom one is for the stego-image.

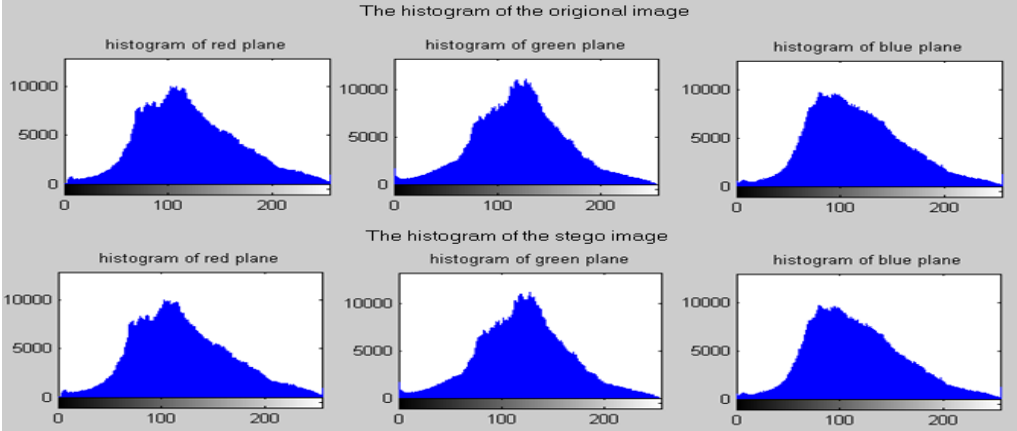


Figure 10. The histogram of the Red, Green and Blue channels of image 3. The upper part is for the original image and the bottom one is for the stego-image.

Table 5. The PSNR in dB for image 1, image2 and image 3 after the attacks.

Image after attack	Image Processing Operations		
	Addition of Gaussian noise	Radial blur	Median blur
Image 1	20.2184	24.5519	22.1014
Image 2	20.4400	20.1862	21.9934
Image 3	20.5697	27.2037	21.7112

The effects of increasing the payload on the PSNR and MSE for image 1, image 2 and image 3 are shown in Table 6 and represented in Figure 11 and Figure 12. The horizontal axis for Figure 11 shows the payload normalized to the value (2120 bytes) and the vertical axis shows the PSNR in dB. The curve shows that a degradation of the PSNR values is obtained as the payload increases, where the embedded text size is increased from 16960 bits to 200264 bits. For Figure 12, the horizontal axis shows the payload normalized to the value (2120 bytes) and the vertical axis shows the MSE.

It can be noticed that for image 1 in Table 3; algorithms in [17] and [21] give PSNR values which are approximately equal to 58 dB for an embedded text size of 2120 bytes. The proposed algorithm can embed approximately 5 times the value 2120 at the same PSNR (58 dB). Moreover, for the algorithm in [23], it gives a PSNR value equal to 56.29 dB when the embedded text message is 2120 bytes, which is approximately the same value obtained by the proposed algorithm, but when the size of the embedded text message is 9 times of 2120 bytes. This result shows the security of the proposed system, since it maintains high PSNR values albeit of increasing the payload to the maximum capacity, which means that the proposed algorithm is superior to other algorithms, where it maintains high security in spite of using the maximum capacity of the cover image. Such improvement in the proposed algorithm is due to employing the less significant bands in the DWT of the cover image.

Table 6. The effects of increasing the payload on the PSNR and MSE for image 1, image 2 and image 3.

Text Size (Bits)	**Image 1**		**Image 2**		**Image 3**	
	PSNR (dB)	**MSE**	**PSNR (dB)**	**MSE**	**PSNR (dB)**	**MSE**
16960	68.5244	0.0091	64.6591	0.0222	67.9923	0.0103
50288	62.8501	0.0337	59.3996	0.0747	63.2162	0.0310
83616	59.6919	0.0698	57.2645	0.1221	60.9395	0.0524
116944	57.2748	0.1218	55.4421	0.1857	59.4579	0.0737
150272	55.6721	0.1761	54.0982	0.2531	58.3095	0.0960
200264	53.9851	0.2598	52.8852	0.3346	56.7209	0.1384
266920	52.4574	0.3693	51.6802	0.4416	54.9500	0.2080

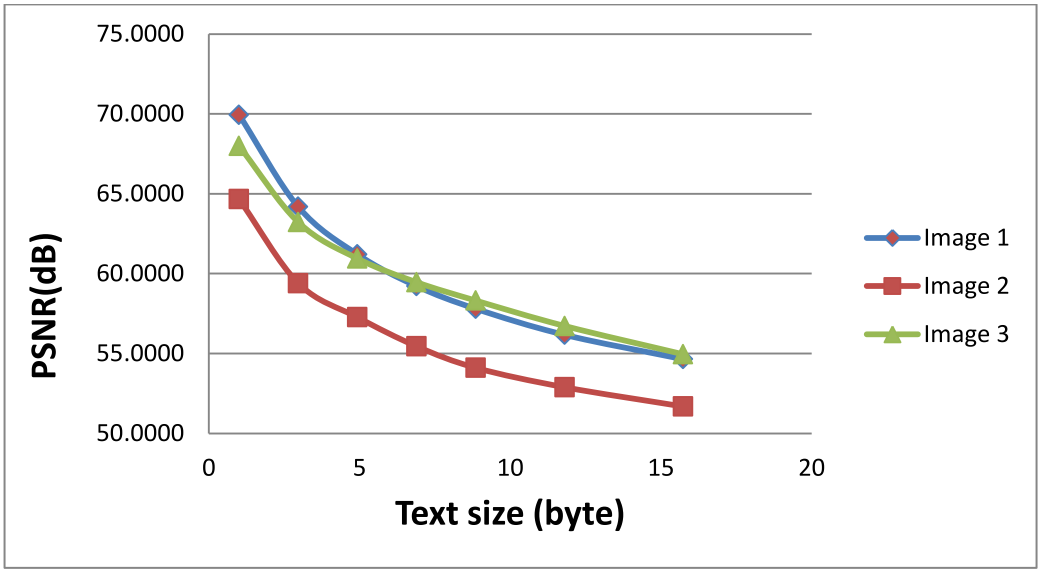


Figure 11. The effect of increasing the embedded text size on the PSNR of image 1, image 2 and image 3.

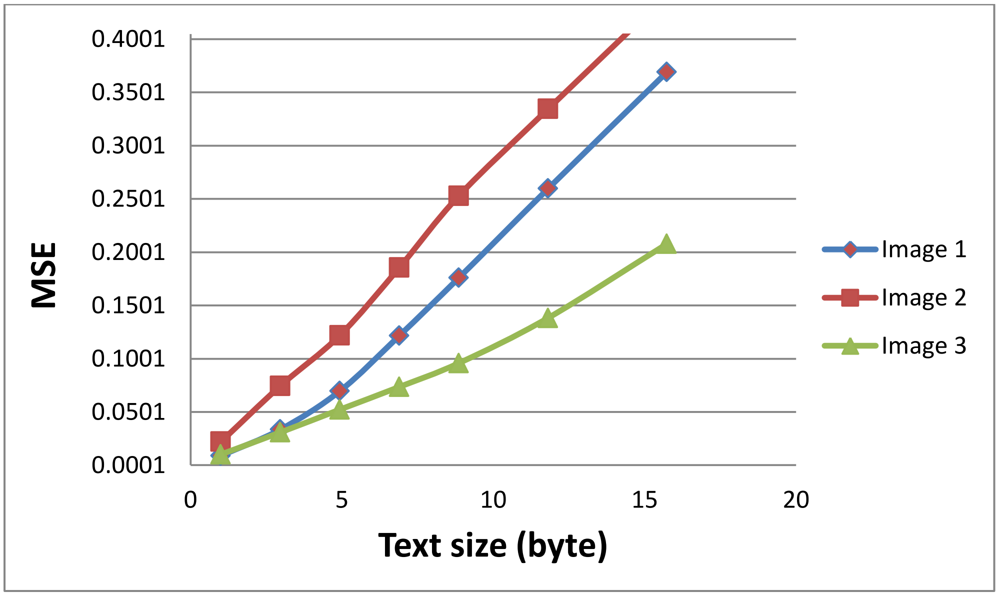


Figure 12. The effect of increasing the embedded text size on the MSE of image 1, image 2 and image 3.

4. CONCLUSION

In this paper, a combination between AES and DWT-based steganography with using PIT has been proposed. The stego-images are looking similar to the cover images and have high PSNR values. The histograms of the original images and stego-images have not shown visual changes. Therefore, an unauthorized observer will not be conscious of the existence of the hidden message. The payload that can be obtained by the proposed scheme is somewhat moderate with respect to other techniques. Moreover, the proposed scheme achieves higher PSNR, in addition to that the trade-off between payload and PSNR confirms the system security with good available capacity. Comparative analysis between the proposed technique and other existing ones has shown the superiority of the proposed technique from the perspective of providing high level of security with moderate capacity. Also, the robustness of the system is verified by applying some attacks. Future work would be directed towards building up special algorithms that improve both the security and capacity as well as increasing PSNR.

REFERENCES

[1] Gurpreet Kaur and Kamaljeet Kumar, "Digital Watermarking and other Data Hiding Techniques," International Journal of Innovation Technology and Exploring Engineering (IJJTEE), vol. 2, issue 5, 2013.

[2] Deepali V. Patil and Shatendra Dubey, "Review Paper on Image Steganography," International Journal of Research in Computer Applications and Robotics, vol. 2, issue 6, pp. 35-40, 2014.

[3] M. A. B. Younes and A. Jantan, "Image Encryption Using Block-based Transformation Algorithm," International Journal of Computer Science, vol. 35, issue 1, pp.15-23, 2008.

[4] Liu Tong and Qiu Zheng-Ding, "A DWT-based Color Image Steganography Scheme," Proceedings of IEEE 6th International Conference on Signal Processing, vol. 2, pp. 1568-1571, 2002.

[5] Anjali A. Shejul and U. L. Kulkarni, "A DWT-based Approach for Steganography Using Biometrics," Proceedings of IEEE 2010 International Conference on Data Storage and Data Engineering (DSDE), pp. 39-43, Feb. 2010.

[6] B. Raja Rao *et al.*, "A Novel Information Security Scheme Using Cryptic Steganography," Indian Journal of Computer Science and Engineering, vol. 1, no. 4, pp. 327-332, 2010.

[7] Lokesh Kumar, "Novel Security Scheme for Image Steganography Using Cryptography Technique," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 2, issue 4, April 2012.

[8] Falesh M. Shelke, Ashwini A. Dongre and Pravin D. Soni, "Comparison of Different Techniques for Steganography in Images," International Journal of Application or Innovation in Engineering and Management (IJAIEM), vol. 3, issue 2, 2014.

[9] Ahmed A. Abdelwahab and Lobha A. Hassan, "A Discrete Wavelet Transform Based Technique for Image Data Hiding," Proceedings of the 2nd National Radio Science Conference, pp. 1-9, Egypt, 2008.

[10] Vladimir Banoci, Gabriel Bugar and Dusan Levicky, "A Novel Method of Image Steganography in DWT Domain," Proceedings of IEEE 21st International Conference on Radioelektronika, pp. 1-4, April 2011.

[11] J. K. Mandal and M. Sengupta, "Authentication Secret Message Transformation through Wavelet Transform-based Sub-band image Coding (WTSIC)," Proceeding of IEEE 2010 International Symposium on Electronic System Design (ISED), pp. 225-229, Dec. 2010.

[12] Souvik Roy and P. Venkateswaran, "Online Payment System Using Steganography and Visual Cryptography," 2014 IEEE Students' Conference on Electrical, Electronics and Computer Science (SCEECS), pp. 1-5, Mar. 2014.

[13] Pye Pye Aung and Tun Min Naing, "Implementation of Cryptography and Image Steganography with New Security Feature," International Conference on Advances in Engineering and Technology (ICAET'2014), 2014.

[14] F. M. Septimin, Mircea Valdutin and P. Lucian, "Secret Data Communication System using Steganography, AES and RSA," 2011 IEEE 17th International Symposium for Design and Technology in Electronic Packaging (SIITME), pp. 339-344, Oct. 2011.

[15] Kamal and Lovnish Bansal, "Enhancement Key of Cryptography and Steganography Using RSA and Neural Network," International Journal of Advanced Research in Computer Engineering and Technology (IJARCET), vol. 3, issue 5, 2014.

[16] S. K. Muttoo and Sushil Kumar, "A Multilayered Secure, Robust and High-Capacity Image Steganographic Algorithm," World of Computer Science and Information Technology Journal (WCSIT), vol. 1, no. 6, pp. 239-246, 2011.

[17] Adnan Gutub, Mahmoud Ankeer, Muhammad Abu-Ghalioun, Abdulrahman Shaheen and Aleem Alvi, "Pixel Indicator High-Capacity Technique for RGB Image based Steganography," IEEE International Workshop on Signal Processing and Its Applications, University of Sharjah, Sharjah, U.A.E., 2008.

[18] Ruchi R. Vairagade, Shubhangini Ugale and Prachi Pendke, "Review on 128-Bit Advanced Encryption Standard Algorithm with Fault Detection," International Journal of Advanced Information and Communication Technology (IJAICT), vol. 1, issue 7, 2014.

[19] Mohammad Abdullatif, Othman O. Khalifa, R. F. Olanrewaju and Akram M. Zeki, "Robust Image Watermarking Scheme by Discret Wavelet Transform," Proceedings of IEEE 5th International Conference on Computer and Communication Engineering (ICCCE), pp. 316-319, Sept. 2014.

[20] Swapnali Zagade and Smita Bhosale, "Secret Data Hiding in Images by Using DWT Techniques," International Journal of Engineering and Advanced Technology (IJEAT), vol. 3, issue 5, 2014.

[21] Yazan Abdallah and H. Seidan, Enhancement of a Steganographic Algorithm for Hiding Text Messages in Images, M. Sc. Thesis, Middle East University, 2013.

[22] Saleh Saraireh, "A Secure Data Communication System Using Cryptography and Steganography," International Journal of Computer Networks and Communications (IJCNC), vol. 5, no. 3, 2013.

[23] S. K. Ghosal, "A New Pairwise Bit-based Data Hiding Approach on 24-Bit Color Image Using Steganographic Technique," Greater Kolkata College of Engineering & Management, Kolkata, India, 2011.

[24] T. Vanitha, Anjalin D'Souza, B. Rashmi and Sweeta D'Souza, "A Review on Steganography – Least Significant Bit Algorithm and Discrete Wavelet Transform Algorithm," International Journal of Innovative Research in Computer and Communication Engineering, vol. 2, special issue 5, 2014.

[25] Amitava Nag, Sushanta Biswas, Debasree Sarkar and Partha Pratim Sarkar, "A Novel Technique for Image Steganography Based on DWT and Huffman Encoding," International Journal of Computer Science and Security (IJCSS), vol. 4, issue 6, 2011.

[26] R. Chandramouli and N. D. Memon, "Steganography Capacity: A Steganalysis Perspective," Proc. SPIE Security and Watermarking of Multimedia Contents, Special Session on Steganalysis, 2003.

ملخص البحث:

تستخدم تقنيات الإخفاء من أجل نقل البيانات بأمان، إلا أنّ الرسائل الناتجة غير القابلة للقراءة عادة ما تجذب انتباه الآخرين، ولهذا يتم إخفاء المعلومات السرية على نحوٍ لا يمكن معه للمقتحمين أن يكتشفوا وجود البيانات السرّية.

تقترح هذه الورقة تقنية محسّنة تجمع بين خوارزمية معيار الإخفاء المتقدم (AES) والاختزال، وتستفيد من مزايا استخدام معامِلات الترددات العالية لصورة التغطية عن طريق تطبيق التحويل المجرّد للمويجات (DWT). وقد تم تطبيق التقنية المقترحة لدراسة أثر إخفاء الرسالة السرّية المخفية في صورة الأحمر والأخضر والأزرق المؤلفة من ٢٤ بِت (bit). وتم تقييم أداء الطريقة المقترحة من حيث: المعدل الأقصى للإشارة إلى الضجيج (PSNR)، وسعة إخفاء الحمل الصافي، وتحليل توزيع الرسم البياني النسيجي.

من ناحية أخرى، تقدم هذه الورقة مقارنة بين التقنية المقترحة وأربعة من الأعمال السابقة الأخرى ذات العلاقة بموضوع الدراسة. وقد كشفت النتائج التجريبية أنّ الطريقة المقترحة في هذا البحث تعطي درجة عالية من الأمان في إخفاء البيانات، إضافة إلى أنها تبدي حصانة ضدّ الهجمات ومحاولات الاختراق المختلفة.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 2, No. 3, December 2016.

MINIMUM BIT ERROR RATE ASSISTED QPSK FOR PRE-FFT BEAMFORMING IN LTE OFDM COMMUNICATION SYSTEMS

Waleed Abdallah[1], Mohamad Khdair[2] and Mos'ab Ayyash[3]

Faculty of Technology and Applied Sciences
Al-Quds Open University, Jerusalem, Palestine.
wsalos@qou.edu[1], mkhdair@qou.edu[2], mayyash@qou.edu[3]

(Received: 31-May-2016, Revised: 25-Aug.-2016, Accepted: 27-Aug.-2016)

ABSTRACT

In this paper, a diamond shape pilot arrangement for OFDM channel estimation is investigated. Such arrangement will decrease the number of pilots transmitted over the communication channel, which will in turn increase data throughput while maintaining acceptable accuracy of the channel estimation. The adaptive antenna array (AAA) is combined with orthogonal frequency division multiplexing (OFDM) to combat the intersymbol interference (ISI) and the directional interferences.

In this paper, The optimum beamformer weight set is obtained based on minimum bit error rate (MBER) criteria in diamond-type pilot-assisted in 3GPP long term evolution (LTE) OFDM systems under multipath fading channel. The simulation results show that the quadrature phase shift keying signaling based on MBER technique utilizes the antenna array elements more intelligently than the standard minimum mean square error (MMSE) technique.

KEYWORDS

MBER beamforming, OFDM systems, Pre-FFT, Diamond shape pilot, 3GPP, LTE.

1. INTRODUCTION

Orthogonal Frequency Division Multiplexing (OFDM) is considered an efficient technique for high speed digital transmission over severe multipath fading channels, where the delay spread is larger than the symbol duration. When the inserting guard time is longer than the delay spread of the channel, this makes the system robust against inter-symbol interference (ISI). In addition to that, channel estimation and compensation can be achieved by inserting known pilot symbols between data symbols [1]-[3].

Over the past few years, Adaptive Antenna Array (AAA) has gained much attention due to its ability to increase the performance of wireless communication systems, in terms of spectrum efficiency, network scalability and operation reliability. Antenna arrays can mitigate the effect of ISI and relax the design of channel equalizer [2].

Adaptive beamforming can separate transmitted signals on the same carrier frequency, provided that they are separated in the spatial domain. The beamforming processing combines the signals received by the different elements of an antenna array to form a single output. The adapted weight set of each element of the antenna array is obtained by the processor, achieving certain criteria to suppress the co-channel interference; thus improving coverage quality.

For a communication system, it is the achievable bit-error rate (BER), not the MSE performance, that really matters. Ideally, the system design should be based directly on minimizing the BER, rather than the MSE. It is demonstrated in Ref. [3] that the MBER solution utilizes the array weights more intelligently than the MMSE approach.

One of the two main techniques which are used in OFDM systems is called Pre-FFT, where an optimum beamformer weight set is obtained in time domain before Fast Fourier Transform (Pre-FFT). The main motivation behind Pre-FFT scheme is reducing the cost due to FFT processing [1]-[7]. The weight obtained for each pilot subcarrier can be identically applied on all data subcarriers in the same OFDM symbol; thereby reducing the number of frequency domain narrow-band beamformers. Post FFT is not always better in performance than pre-FFT [2]-[3]. In [4], a pre-FFT least mean square (LMS) beamforming for OFDM systems was analyzed in additive Gaussian noise channel. An adaptive MBER beamforming was analyzed in [4] for single carrier modulation and in [2] for OFDM systems in additive Gaussian noise channel. A class of MBER algorithms were studied in [8] and combined with space time coding in [9]. Eigenvector combining was considered in [8]. MIMO MBER beamforming for OFDM was studied in [5]. A block by block post-FFT multistage beamforming was considered in [4].

In [1]-[2], the MMSE and MBER beamformers for Pre-FFT OFDM are presented, respectively, without investigating several factors affecting performance. The channel is assumed to be non-dispersive with additive Gaussian noise, which is not a practical channel.
Since new wireless standards, such as IEEE 802.11 and 802.16, use the pilot subcarriers in their structures, our focus in this paper will be given to suppress co-channel interference and mitigate the multipath interference in pilot-assisted OFDM systems.

In [3], the MMSE beamforming algorithm for Pre-FFT OFDM system is applied on a channel assumed to be frequency selective fading. A recent work [5]-[6] has suggested an adaptive MBER beamforming assisted receiver for binary phase shift keying OFDM communication systems. This paper first presents a novel beamforming technique based directly on minimizing the system's BER for broadband OFDM wireless systems with quadrature phase shift keying (QPSK) modulation. The main contribution in this paper is to show that the diamond type pilot aided channel estimation has better performance when the channel is time-variant and the reduced number of pilot increases efficiency. The paper is an extension to [2] with an improved channel estimator and the performance results are more applicable to 3GPP-LTE.

This paper is organized as follows: Section 2 describes the LTE pilot structure. Section 3 describes the Pre-FFT adaptive beamforming based on MMSE criteria. In Section 4, Pre-FFT adaptive beamforming based on MBER criteria is introduced. Sections 5 and 6 clarify the computational complexity and the convergence rate for the simulated system, respectively. System specification is shown in section 7. In section 8, simulation results are provided. Finally, conclusions and possible directions for future work are presented in section 9.

2. LTE Pilot Structures

As indicated earlier, wireless standards, such as IEEE 802.11 and 802.16, use the pilot subcarriers in their structures. This pilot signal is used to measure the channel quality and perform channel estimation at the end-user side. There are several types of pilot structures: Block-type pilot, Comb-Type pilot and Diamond-type pilot. In Table 1, we present a quick comparison between the system specifications for Block-type pilot [14] and Diamond-type pilot simulated in this paper:

In this paper, we research further the time domain MBER and LMS channel estimation based on diamond-type pilot structure. The aim is to achieve better performance and fewer computations and at the same time increase spectral efficiency and throughput by using less number of pilot signals.

Table 1. Simulation system specifications for Block and Diamond type Pilot.

	Block-type Pilot	Diamond-type Pilot
Cluster Size	4	6 (As shown in Figure 1)
Subcarrier	64 , 128 , 256 , 512	60 , 120 , 300 , 600
Number Pilot = (Subcarrier ÷ Cluster Size)	16 , 32 , 64 , 128	10 , 20 , 50 , 100

Figure 1 shows an example of the pilot pattern in every RB of a subframe at the first antenna port, but the location may be shifted in frequency domain for different subframes [10].

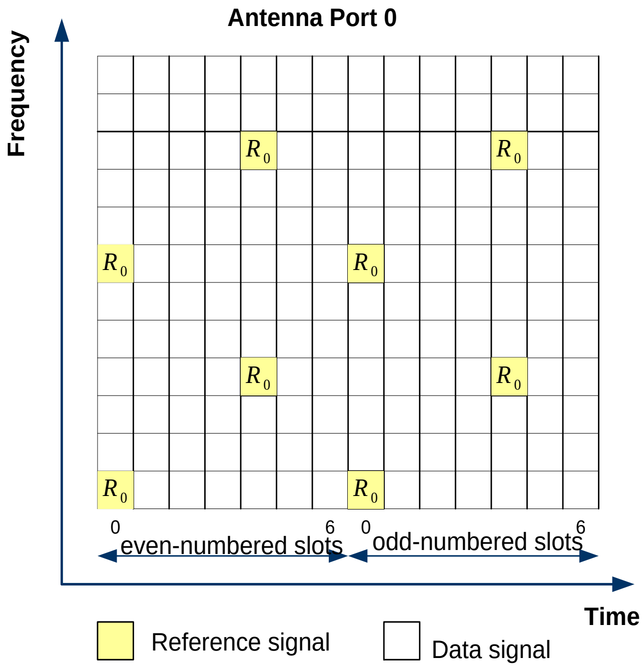


Figure 1. Pilot locations for the first transmit antenna in the 3GPP-LTE system [10].

Consider M-users, where each user transmits a QPSK signal and the OFDM system uses K subcarriers for parallel transmission [2]. The sample modulated by the kth subcarrier of the mth user is given by:

$$x_m(k) = b_m(k) \qquad 1 \le m \le M \qquad 1 \le k \le K \tag{1}$$

where $b_m(k) \in \{\pm 1 \pm j\}$ are QPSK symbols. Source 1 is assumed to be the desired user and the rest of the sources are the interfering users. This data can be interpreted to be a frequency-domain data and subsequently converted into a time-domain signal by an IFFT operation. This process can be written as:

$$\bar{y}_m = \frac{1}{K} F^H \bar{x}_m \qquad 1 \le m \le M \tag{2}$$

where,

$$\bar{y}_m = \left[y_m(1), y_m(2), \ldots, y_m(K)\right]^T \tag{3}$$

$$F = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & e^{\frac{-j2\pi(1)(1)}{K}} & \cdots & e^{-\frac{j2\pi(1)(K-1)}{K}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{\frac{-j2\pi(K-1)(1)}{K}} & \cdots & e^{-\frac{j2\pi(K-1)(K-1)}{K}} \end{bmatrix} \tag{4}$$

$$\bar{x}_m = \left[x_m(1), x_m(2), \ldots, x_m(K)\right]^T \tag{5}$$

F is representing the FFT operation matrix and *H* denotes the Hermitian transpose of a matrix. To add the CP, $\bar{y}_m$ is cyclically extended to generate $\tilde{y}_m$ by inserting the last *v* element of y_m at its beginning; i.e.,

$$\tilde{y}_m = \begin{bmatrix} J_v \\ I_K \end{bmatrix} \bar{y}_m \tag{6}$$

where J_v contains the last *v* rows of a size *K* identity matrix I_K .

Finally, the OFDM time signals are transformed to the analog format by a D/A converter prior to transmission. A multipath channel with a maximum of *L* paths exists between the m^{th} source (desired or interference) and the array in the form of:

$$h_m(k) = \sum_{l=0}^{L-1} \alpha_{m,l} \delta(k-l) \qquad m = 1, \cdots, M \tag{7}$$

where $\alpha_{m,l}$ denotes a complex random number representing the l^{th} channel coefficient for the m^{th} source and δ(.) is the delta function.

Figure 2 illustrates the architecture of Pre-FFT beamforming at the receiver of an OFDM system, where *CP* is assumed to be longer than the channel length (*v* > *L*); thus, received signal on the p^{th} antenna of a Uniform Linear Array (ULA) for one OFDM symbol can be written as:

$$r_p(k) = \sum_{m=1}^{M} \sum_{l=0}^{L-1} \alpha_{m,l} \tilde{y}_m(k+v-l) e^{-j\frac{2\pi}{\lambda}(p-1)d\cos(\theta_{m,l})} + \eta_p(k) \tag{8}$$
$$1 \le p \le P \quad , \qquad 1 \le k \le K$$

where, $\eta_p(k)$ represents the channel noise entering the p^{th} antenna. $\theta_{m,l}$ denotes the direction of arrival (DOA) of the l^{th} path and m^{th} source. Without loss of generality, we have assumed that the channels of all sources have the same length *L*. At the receiver, the converted digital signal with a spatial phase for each array element is multiplied by the weight (w_p) of adaptive beamformer and then transformed back into frequency-domain (data and pilot) symbols by the FFT. This process can be written as:

$$Z(k) = W^H \cdot \bar{R} \tag{9}$$

$$W = [w_1 \, w_2 \quad \cdots w_P]^T \tag{10}$$

$$\bar{R}(k) = [\bar{r}_1(k) \bar{r}_2(k) \quad \cdots \bar{r}_P(k)]^T \tag{11}$$

$$\bar{Z} = [z(1) \, z(2) \quad \cdots z(K)] \tag{12}$$

$$\hat{Z} = \bar{Z} \cdot F \tag{13}$$

where $\hat{Z}$ is the frequency-domain data, which is given by:

$$\hat{Z} = \left[\hat{z}(1), \hat{z}(2), \ldots, \hat{z}(K)\right]^T \tag{14}$$

and $\hat{z}(k)$ denotes the corresponding received sample at the k^{th} subcarrier .

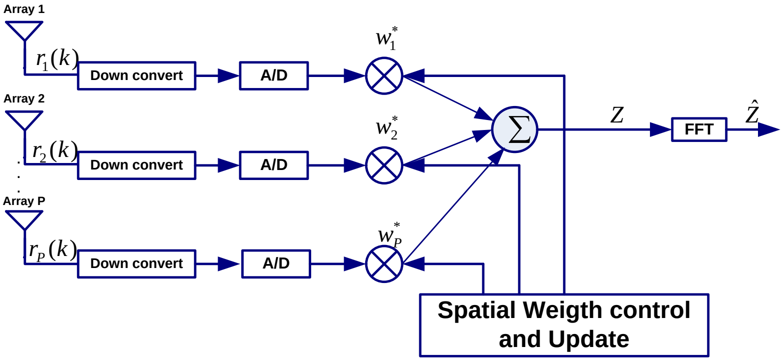


Figure 2. Block diagram of the Pre-FFT OFDM adaptive receiver [1].

The estimate of the transmitted bit $b_i(k)$ is given by:

$$\hat{b}_i(k) = \begin{cases} b^{[1]} = +1 + j, & (\hat{z}_{\text{Re}}(k)) \geq \text{and } (\hat{z}_{\text{Im}}(k)) \geq 0, \\ b^{[2]} = -1 + j, & (\hat{z}_{\text{Re}}(k)) < 0 \text{ and } (\hat{z}_{\text{Im}}(k)) \geq 0, \\ b^{[3]} = -1 - j, & (\hat{z}_{\text{Re}}(k)) < 0 \text{ and } (\hat{z}_{\text{Im}}(k)) < 0, \\ b^{[4]} = +1 - j, & (\hat{z}_{\text{Re}}(k)) \geq 0 \text{ and } (\hat{z}_{\text{Im}}(k)) < 0, \end{cases} \tag{15}$$

where $\hat{b}_i(k)$ is the k$^{\text{th}}$ symbol of user i, which takes values from a QPSK symbol set shown in equation (15), $(\hat{z}_{\text{Re}}(k))$ denotes the real part of $\hat{z}(k)$ and $(\hat{z}_{\text{Im}}(k))$ denotes the imaginary part of $\hat{z}(k)$.

3. Adaptive Beamforming for Pre-FFT OFDM System

Implementing MMSE Beamforming using LMS adaptive algorithm is done by comparing the received pilot symbols with their known values, so that an error signal is generated at the receiver [1]-[2]. Since this error signal is in frequency domain while Pre-FFT weights are updated in time domain, the frequency-domain error signal must be converted into time domain.

If there are a total of *Q* pilot symbols in every OFDM symbol, then we define two $K \times 1$ vectors d_q and Z_q, such that the k^{th} element of d_q is zero if *k* is a data subcarrier and is the known pilot value if *k* is a pilot subcarrier.

Similarly, the k^{th} element of Z_q is zero if the *k* is a data subcarrier and is the received pilot value if *k* is a pilot subcarrier. Therefore, the error signal in frequency domain is given by:

$$E_q = d_q - Z_q\,. \tag{16}$$

This error signal must be converted into time domain for the Pre-FFT weight adjustment algorithm. Therefore,

$$\bar{e} = \frac{1}{K} F^H E_q \tag{17}$$

where *e* is the vector of error samples in time domain.

$$\bar{e} = [e(1)\, e(2) \quad \cdots e(K)]^T \ . \tag{18}$$

Consequently, the Pre-FFT weights are updated using the following Least Mean Squares (LMS) algorithm [1]-[2].

$$\begin{gathered} W(k) = W(k-1) + 2\mu \cdot r(k) \cdot e^*(k) \\ 1 \le k \le K \end{gathered} \tag{19}$$

where μ is the step size parameter, and $*$ represents the complex conjugate. The last update at the end of each OFDM block (W (K)) is used as the initial value of the next block.

4. MBER-Based Beamforming Algorithms

In this section, Pre-FFT adaptive beamforming based on MBER criteria is introduced to obtain the optimum weight set. The theoretical MBER solution for the Pre-FFT OFDM beamformer is obtained in [1]-[2], [17] where the channel is assumed to be non-dispersive with additive white Gaussian noise. The error probability (BER cost function) of the frequency domain signal of the beamformer is given by:

$$P_E(W) = \text{Prob}\{\text{sgn}(b_1(k)\text{Real}(\hat{z}(k)) < 0\} \tag{20}$$

where $\text{sgn}(\cdot)$ is the sign function.

From equation (20), define the signed decision variable

$$\begin{aligned} \hat{z}_s(k) &= \text{sgn}(b_1(k))\text{Re}(\hat{z}(k)) \\ &= \text{sgn}(b_1(k))\text{Re}(\hat{z}'(k)) + \eta'(k) \end{aligned} \tag{21}$$

where,

$$\hat{z}'(k) = W^H[\bar{r}(k) - \eta(k)]F(k) \tag{22}$$

and

$$\eta'(k) = \text{sgn}(b_1(k))\text{Re}(W^H)\eta(k)F(k)) \tag{23}$$

$\hat{z}_s(k)$ is a very good error indicator for the binary decision; i.e., if it is positive, then the decision is correct, else if it is negative, then an error occurred. *F*(*k*) is the $\eta'(k)$ k^{th} column of *F*. Notice that *F* is a unitary matrix, so is still Gaussian with zero mean and variance $\sigma_n^2 \cdot W^H W$.

Obviously, the two marginal conditional p.d.f.s for $z_{\text{Re}}(k)$ and $z_{\text{Im}}(k)$ are Gaussian mixtures. Define

$P_{E,\text{Im}}(W\ ) = \text{prob}(\hat{b}_{\text{Im},1}(k)) \ne b_{\text{Im},1}(k))$, $\hat{b}_1(k) = \hat{b}_{\text{Re},1}(k) + j\hat{b}_{\text{Im},1}(k)$ and

$b_1(k) = b_{\text{Re},1}(k) + jb_{\text{Im},1}(k)$. Obviously, the two marginal conditional p.d.f.s are for $z_{\text{Re}}(k)$ and $z_{\text{Im}}(k)$.

The BER of the beanformer for equation (20) is:

$$P_E(W) = \frac{1}{2}(P_{E,\text{Re}}(W\ ) + P_{E,\text{Im}}(W)) \ . \tag{24}$$

The MBER beamforming solution is defined as:

$$W = \arg\min_W P_E(W) \ . \tag{25}$$

The gradient of P_E(w) with respect to w is:

$$\nabla P_E \ (W) = \frac{1}{2}(\nabla P_{E,\text{Re}}(W\ ) + \nabla P_{E,\text{Im}}(W)) \ . \tag{26}$$

Given the gradient, the optimization problem (26) can be solved for interactively using the simplified conjugated gradient algorithm, which is detailed in [4], [14].

The conditional probability density function (pdf), given the channel coefficients $\alpha_{m,l}$ of the error indicator, $\hat{z}_s(k)$, is a mixed of Gaussian distributions [3]; i.e.,

$$\begin{gathered} p_z(\hat{z}_s) = \frac{1}{K\sqrt{2\pi}\sigma_n\sqrt{W^H W}} \cdot \\ \sum_{k=1}^{K} \exp(-\frac{(\hat{z}_s - \text{sgn}(b_1(k))\text{Re}(\hat{z}'_s(k)))^2}{2\sigma_n^2 W^H \cdot W}) \end{gathered} \tag{27}$$

And it is the best indicator of a beamformer's BER performance deriving a closed form for the average error probability is not easy. Therefore, we use the gradient conditional error probability to update the weight vector. The conditional error probability, given the channel coefficients $\alpha_{m,l}$ of the beamformer $P_E(W\ )$, is given in [3].

$$\begin{gathered} p_E(w) = \frac{1}{K\sqrt{2\pi}\sigma_n\sqrt{W^H W}} \cdot \\ \sum_{k=1}^{K} \int_{q_k(w)}^{\infty} \exp(-\frac{u^2}{2})du \end{gathered} \tag{28}$$

$$\hat{P}_E(W) = \frac{1}{2K}\sum_{k=1}^{K}(Q(q_{\text{Re}}^{(k)}(W\ )) + Q(q_{\text{Im}}^{(k)}(W\ )) \tag{29}$$

where $Q(\cdot)$ is the Gaussain error function and is given by:

$$Q(u) = \frac{1}{\sqrt{2\pi}}\int_u^{\infty} \exp(-\frac{y^2}{2})dy \tag{30}$$

$$q_{\text{Re}}^{(k)}(W\ ) = \frac{\text{sgn}(b_1(k))(\hat{z}'_{\text{Re}}(k))}{\sigma_\eta\sqrt{W^H W}} \tag{31}$$

$$q_{\text{Im}}^{(k)}(W\ ) = \frac{\text{sgn}(b_1(k))(\hat{z}'_{\text{Im}}(k))}{\sigma_\eta\sqrt{W^H W}} \ . \tag{32}$$

Based on the definition, the gradient of $P_E(W\ )$ with respect to W is:

$$\begin{gathered} \nabla P_E(W\ ) = \frac{1}{K_p\sqrt{2\pi}\sigma_n\sqrt{W^H W}} \cdot \sum_{n=1}^{N_b} \exp(-\frac{\hat{z}'^2(k)}{2\sigma_n^2 W^H \cdot W}) \\ \cdot \text{sgn}(b_1(k))(\frac{\text{Re}(\hat{z}'(k))W}{W^H W} - A\bar{r}(k)) \end{gathered} \tag{33}$$

In an OFDM system, it is assumed that there are pilot signals in every symbol to do channel estimation [1]-[2]. The pilot signals are also used to adaptive update of the weight vector of the beamformer. The transmitted pilot signal vector of desired user $\bar{x}_{1p}$ and the received pilot signal vector $\bar{\hat{z}}_p$ in frequency domain can be written as follows:

$$\bar{x}_{1p} = [x_1(1), 0.., x_1(\Delta p + 1), 0, .., x_1((K_p - 1)\Delta p + 1)), 0, ..] \tag{34}$$

$$\begin{aligned}\bar{\hat{z}}_p &= [\hat{z}(1),0..,\hat{z}(\Delta p+1),0,..,\hat{z}((K_p-1)\Delta p+1),0,..] \\ &= W^H \bar{R} F_p\end{aligned}$$

where

$$Fp=\begin{bmatrix} 1 & 0 & \cdots & 1 & \cdots & 1 & 0\cdots \\ 1 & 0 & \cdots & e^{-j2\pi(1)(\Delta p)/K} & \cdots & e^{-j2\pi(1)(K_p-1)\Delta p/K} & 0\cdots \\ \vdots & 0 & & \vdots & \ddots & \vdots & 0\cdots \\ 1 & 0 & \cdots & e^{-j2\pi(K-1)(\Delta p)/K} & \ldots & e^{-j2\pi(K-1)(K_p-1)\Delta p/K} & 0\cdots \end{bmatrix} \tag{35}$$

$\bar{R} = [\bar{r}(1)\bar{r}(2) \quad \cdots \bar{r}(K)]$, Δp and K_p represents the frequency spacing between consecutive pilot symbols and the number of pilot symbols inserted in OFDM symbol, respectively. We assume that the first pilot symbol is positioned at the first sub-channel.

The method of approximating a conditional pdf, known as a kernel density or Parzen window-based estimate [7]-[8], is used to estimate the conditional error probability given that the channel coefficients $\alpha_{m,l}$ is used on OFDM systems. Given a symbol of K_p training samples $\{\bar{r}(k), b_1(k)\}$, a kernel density estimate of the conditional pdf given the channel coefficients $\alpha_{m,l}$ at pilot locations, is given by:

$$\hat{p}(\hat{z}) = \frac{1}{2\pi K_p \rho_\eta^2 W^H W} \sum_{K_p=1}^{K_p} \exp\left(\frac{|\hat{z}-\hat{z}(k)|^2}{2\rho_\eta^2 W^H W} \right) \tag{36}$$

where the kernel width ρ_η is related to the noise standard deviation σ_η .From this estimated p.d.f., the estimated BER is given by:

$$\hat{P}_E(W) = \frac{1}{2K_p} \sum_{k=1}^{K_p} (Q(\hat{g}_{\mathrm{Re}}^{(k)}(W)) + Q(\hat{g}_{\mathrm{Im}}^{(k)}(W)) \tag{37}$$

where

$$\hat{g}_{\mathrm{Re}}^{(k)}(W) = \frac{\mathrm{sgn}(b_{\mathrm{Re},1}(k\times\Delta p+1))\mathrm{Re}(W^H \bar{R} F_p(k\times\Delta p+1))}{\rho_\eta \sqrt{W^H W}} \tag{38}$$

and

$$\hat{g}_{\mathrm{Im}}^{(k)}(W) = \frac{\mathrm{sgn}(b_{\mathrm{Im},1}(k\times\Delta p+1))\mathrm{Im}(W^H \bar{R} F_p(k\times\Delta p+1))}{\rho_\eta \sqrt{W^H W}} \tag{39}$$

$F_p(k\times\Delta p+1)$ is the $(k\times\Delta p+1)^{th}$ column of F_p . From this estimated conditional pdf, given the channel coefficients $\alpha_{m,l}$, the gradient of the estimated BER is given by [3]:

$$\begin{aligned}\nabla\hat{P}_{E,\mathrm{Re}}(W) &= \frac{1}{2K_p\sqrt{2\pi}\rho_\eta\sqrt{W^H W}} \sum_{k=1}^{K_p} \exp(\frac{-z_{\mathrm{Re}}^2(k)}{2\rho_\eta^2 W^H W}) \\ &\times \mathrm{sgn}(b_{\mathrm{Re},1}(k))(\frac{z_{\mathrm{Re}}^2 W}{W^H W} - \bar{R}F(k))\end{aligned} \tag{40}$$

ρ_n is related to the standard deviation σ_n of the channel noise. From this estimated pdf, the gradient of the estimated BER is given by [1]:

$$\nabla \widehat{P}_E(W\ ) = -\frac{1}{\sqrt{2\pi}\rho_n\sqrt{W^H W}} \sum_{k=0}^{K_p-1} \exp(-\frac{(\mathrm{Re}(\hat{z}(k\times\Delta p+1)))^2}{2\rho_n^2 W^H W}) \\ \times \mathrm{sgn}(b_1(k\times\Delta p+1))\bar{R}F_p(k\times\Delta p+1) \tag{41}$$

$$\nabla \hat{P}_{E,\mathrm{Im}}(W) = \frac{1}{2K_p\sqrt{2\pi}\rho_\eta\sqrt{W^H W}} \sum_{k=1}^{K_p} \exp(\frac{-z_{\mathrm{Im}}^2(k)}{2\rho_\eta^2 W^H W}) \\ \times \mathrm{sgn}(b_{\mathrm{Im},1}(k))(\frac{z_{\mathrm{Im}}^2 W}{W^H W} - \bar{R}F(k)) \tag{42}$$

For each OFDM symbol, we can find the optimum weight vector W by the steepest-descent gradient algorithm [3]:

$$\nabla \widetilde{P}_E(W,k) = \frac{1}{4\sqrt{2\pi}\rho_\eta}[-\mathrm{sgn}(b_{\mathrm{Re},1}(k\times\Delta p))\exp(\frac{-(\hat{z}_{\mathrm{Re}}(k\times\Delta p+1))^2}{2\rho_\eta^2}) \\ +j\,\mathrm{sgn}(b_{\mathrm{Im},1}(k\times\Delta p))\exp(\frac{-(\hat{z}_{\mathrm{Im}}(k\times\Delta p+1))^2}{2\rho_\eta^2})]\times\bar{R}F_p(k\times\Delta p+1)) \tag{43}$$

That is to say, W weight vector can be updated K$_P$ times in one OFDM symbol. Thus, complexity is reduced and consequently, the update equation is given by:

$$W(k+1) = W + \mu(-\nabla \widetilde{P}_E(W(k),k) \tag{44}$$

$$c_1 = W^H(k+1)P_1 \tag{45}$$

$$P_1 = A_1 s_1 \tag{46}$$

where s_1 is the steering vector and A_1 is power of desired user.

$$W\ (k+1) = \frac{c_1}{|c_1|}W(k+1)\,. \tag{47}$$

We assume a perfect P1 at receiver and known arrival direction of the desired user. It is shown in ref. [15] that the steering vector s1 should be known at the receiver, because the beamformer's output consists of: (the desired signal + residual interference) , the steering vector s1 determines the desired signal. Also it is indicated that QPSK case is different from BPSK case in which the receiver does not require the steering vector of the desired user to make a decision.

The proposed MBER algorithm is summarized in Table 2, this algorithm is composed of two main loops. The outer loop is for each symbol of data and the inner loop is repeated over the same symbol of data until certain number of iterations is reached. In the main loop, we formulate a symbol of data (300 bits) from the output of the antenna array. In the inner loop the gradient vector is determine from (43) at pilot locations (Np =50). Then, we compute the weight update vector from (44). After the end of the inner loop, we determine the detected signal by multiplying the computed optimized weight vector with the received signal in order to use it in calculating the BER the last update at the end of each OFDM block W (k) which used as an initial value in the next symbol. Then, we get back to the main loop and form another symbol of data and so on. These processes iterate until we finish all the incoming data.

5. Computational Complexity

In this section, we compare the two algorithms in terms of computational complexity [4]. Table 3 illustrates the computational complexity of pre-weight update to complete a single iteration; i.e., detecting one bit. The proposed MBER maintains the linearity in complexity.

Table 2. MBER algorithm summary.

Initialization
$i=1, \mu=.01$, size $K=300$, $K_p=50$. • Calculate variance of noise σ_η, ρ_η. • Initial weight vector $W=.01\times ones(N_t,1)$.
Outer loop (1: floor (all bits/symbol))
• Form a symbol of data from the received signals.
Inner loop (while $k<K_p$)
• Calculate the gradient matrix over the symbol in eqn. (43). • Update the weight matrix as: $W(k+1)=W(k)+\mu(-\nabla\hat{P}_E(W(k),w))$. • $P_1=A_1 s_1$ $c_1=W^H(k+1)P_1$, where c_1 is real and positive. $W\ (k+1)=\frac{c_1}{\lvert c_1\rvert}W(k+1)$, the rotating operation. • Normalize the solution $W\ (k+1)=W(k+1)/\lVert W\ (k+1)\rVert$.
• ***end of inner loop*** • Determine the detected signals in order to be used for calculating the BER. • Increment the symbol number.
• ***end of outer loop***

Table 3. Comparison of computational complexity pre-weight update.

	Multiplications	additions	exp(•) evaluation
MBER	$4\times P+4$	$4\times P-1$	1
MMSE	$8\times P+2$	$8\times P-1$	–

6. Convergence Rate

In this section, we run the algorithm of the MBER for 400 samples limited to 1 and 11 iterations. The results are shown in Figure 11, where we can see that the proposed algorithm converges very fast to the optimal solution (after one iteration only).

Figure 9 and Figure 10 illustrate the convergence performance of LMS based on MMSE Pre-FFT beamformer. The optimum weights and steady state MSE performance, respectively, can be obtained under the same conditions after about 200 OFDM symbols.

The figures resulting from the simulation show that shorter training symbols and less computational complexity OFDM symbols are required in the MBER Pre-FFT beamformer algorithm and the MBER algorithm converges faster than MMSE.

7. System Specification

In this paper, we follow the specification of 3GPP-LTE Release 8 [10]-[11] to perform the analysis, simulation and implementation. The 3GPP-LTE supports a maximum 512-point FFT size at 10-MHz bandwidth and has six antenna elements and half-wavelength spacing. The general shape of the diamond shape pilot signal was shown in Figure 1.

8. Simulation Results

In this section, the simulation is performed to illustrate and compare the performances of Pre-FFT beamformer using different MBER-based algorithms. 300 subcarriers (*50 + 250*) are used. The OFDM system is perfectly synchronized with a CP length larger than the channel length (*v=16*). QPSK modulation is used in the system with six antenna elements and half-wavelength spacing. The example used in our computer simulation study considers one desired user with DOA at 90° and two interferers with SIR= -3dB and 0dB, respectively and DOA 50° and 140°. We further assumed channels with different lengths and with real coefficients *0.8935,0.0957,0.0107* and *0* for all sources and with an angle spread of $\pm 15^o$ (for all sources) [12].

Figure 3 and Figure 4 compare the BER performance of the MBER beamformer with that of the MMSE beamformer for SIR= -3dB and 0dB, respectively with AWGN. Figure 5 and Figure 6 compare the BER performance against SNR for the LMS and MBER beamformers for the case that the number of elements is 6 under selective fading channel. It is observed that the BER performance of the MBER beamformer is better than that of the LMS. Figure 7 and Figure 8 for different channel real coefficients illustrates the beam pattern of the LMS, MBER and beamformers for Pre-FFT OFDM adaptive antenna array when the number of antenna elements is 6, respectively. It shows that the MBER Pre-FFT beamformer has lower sidelobe levels and deeper nulls.

It is observed that the beam pattern performance of the MBER beamformer is better than that of the LMS beamformer. Note that the LMS beamformer appears to have a better amplitude response than the MBER beamformer. If the amplitude response alone would constitute the ultimate performance criterion of a beamformer, the MMSE beamformer would appear to be more beneficial. However, considering the magnitude alone can be misleading. It is shown in [3] in detail that the MBER solution has a better ability to cancel interfering signals.

Figure 9 shows that regarding the convergence performance of LMS based on MMSE Pre-FFT beamformer, we can see that after about 400 OFDM symbols we still don’t have convergence. Figure 10 shows that we still have mean square errors even after 400 OFDM symbols. As we note from Figure 11, the optimum weights and steady state MSE performance, respectively, can be obtained under the same conditions after about 5 OFDM symbols which results in much fewer calculations to reach convergence. For the simulation, shorter training OFDM symbols are required in the MBER Pre-FFT beamformer algorithm.

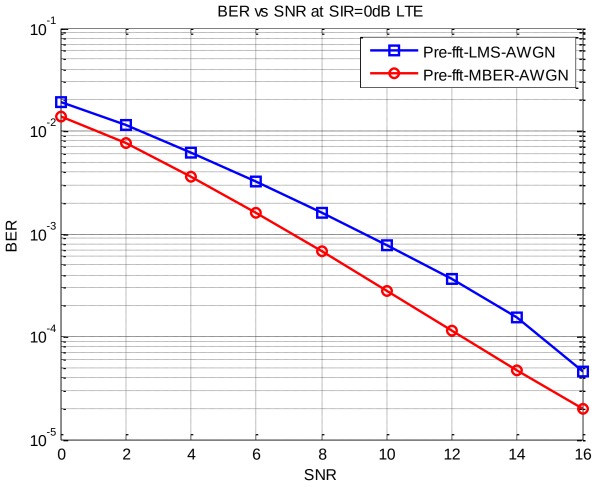


Figure 3. Comparison of the bit error performance.

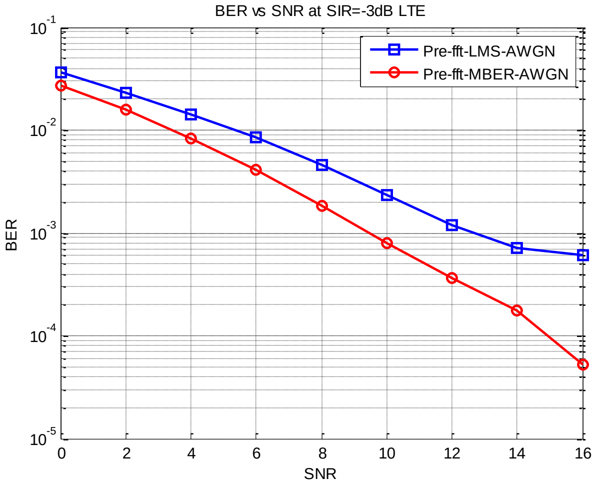


Figure 4. Comparison of the bit error performance.

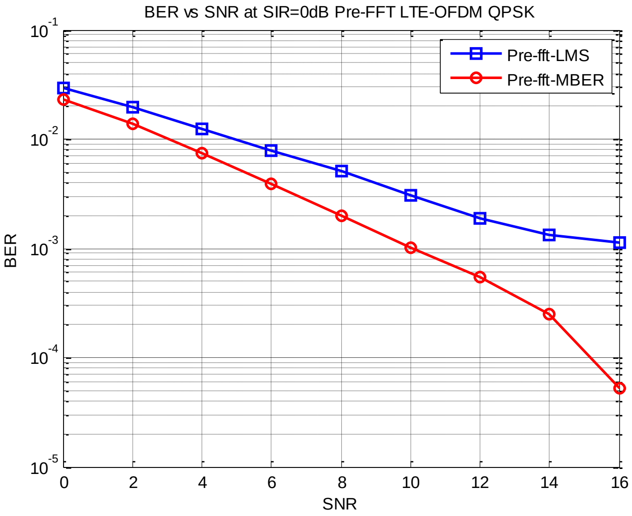


Figure 5. Comparison of the bit error performance.

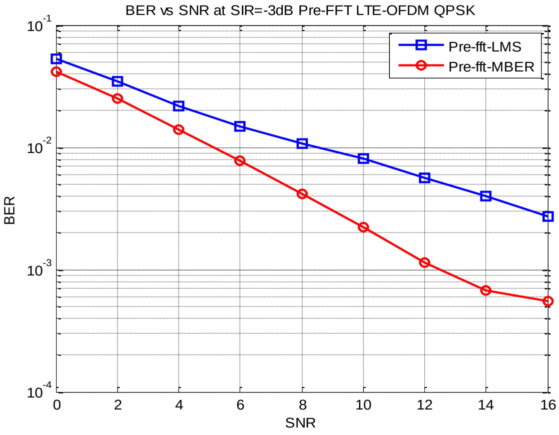


Figure 6. Comparison of the bit error performance.

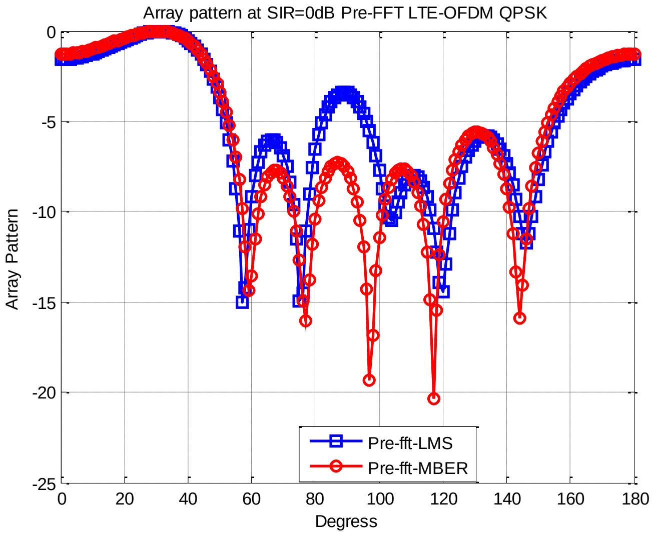


Figure 7. Beam pattern of the MBER and LMS.

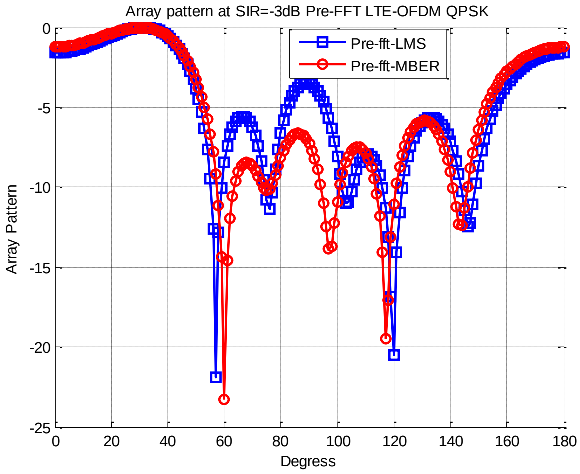


Figure 8. Beam pattern of the MBER and LMS.

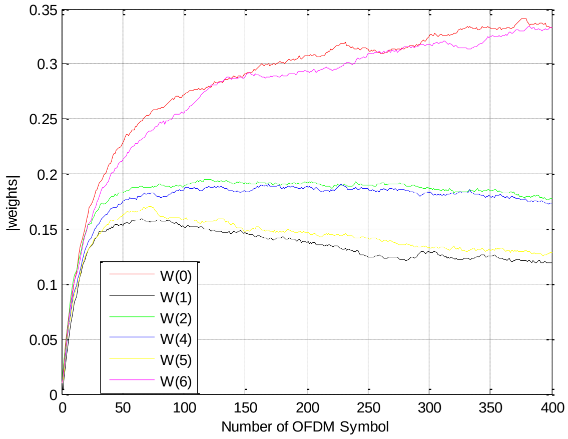


Figure 9. Convergence of the MMSE beamforming to obtain the optimum weights on the Pre-FFT performance.

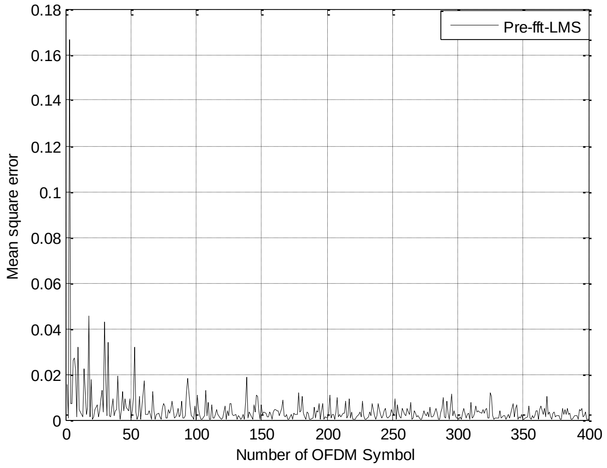


Figure 10. Mean square error of the LMS beamformer.

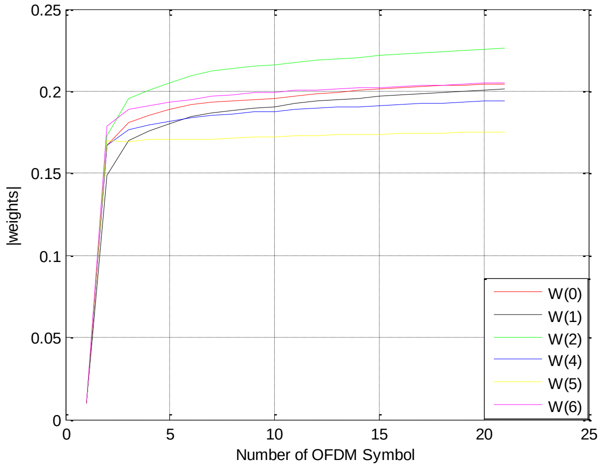


Figure 11. Convergence of the MBER beamforming to obtain the optimum weights on the Pre-FFT performance.

9. Conclusions and Future Work

In this paper, BER performance is compared against the iteration index during adaptive implementation, SNR, beam pattern. When applying on LTE, simulation results show that better efficiency is achieved, because fewer pilots are used and as a result more data is transmitted. Also, the use of MBER resulted in quick conversion compared to MMSE. The proposed MBER yields a much better convergence speed while maintaining quadratic complexity. A proposed extension to this work would be to investigate the use of other modulation techniques, such as 16 QAM or 32 QAM, with a diamond shape pilot for different types of communication channels.

REFERENCES

[1] L. Fan, H. Zhang and C. He," Minimum Bit Error Beamforming for Pre-FFT OFDM Adaptive Antenna Array," IEEE International Conference, 25-28 Sept. 2005.

[2] S. Seydnejad and S. Akhzari, "A Combined Time-Frequency Domain Beamforming Method for OFDM Systems," 2010 International ITG Workshop on Smart Antennas, IEEE, 23-24 Feb. 2010.

[3] S. Chen, N. N. Ahmad and L. Hanzo, "Adaptive Minimum Bit-Error Rate Beamforming," IEEE Trans. on Wireless Commun., vol. 4, no. 2, pp.341-348, March 2005.

[4] Z. Lei and P. S. Chin, “Post and Pre-FFT Beamforming in an OFDM System," Proc. of IEEE Vehicular Tech. Conf., vol. 1, pp. 39-43, 17-19 May 2004.

[5] J. Via´ , V. Elvira, I. Santamari´a and R. Eickhoff, "Minimum BER Beamforming in the RF Domain for OFDM Transmissions and Linear Receivers," IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP 2009), Taipei, Taiwan, Apr. 2009.

[6] S. Hara, M. Budsabathon and Y. Hara, "A Pre-FFT OFDM Adaptive Antenna Array with Eigenvector Combining", IEEE International Conference on Commun., vol. 4, pp. 2412 - 2416, June 2004.

[7] T. A. Samir, S. Elnoubi and A. Elnashar, "Block Shanno MBER Beamformimg," International Symposium on Signal Processing and Its Applications (ISSPA 2007), Sharjah, UAE, 12-15 Feb. 2007.

[8] S. Chen, A. K. Samingan and B. Mulgrew, "Adaptive Minimum-BER Linear Multiuser Detection for DS-CDMA Signals in Multipath Channels," IEEE Trans. on Signal Process., vol. 49, no. 6, pp. 1240-1247, June 2001.

[9] T. A. Samir, S. Elnoubi and A. Elnashar, "Block Shanno Minimum Bit Error Rate Beamforming," IEEE Trans. on Vehicular Technology, vol. 57, no. 5, September 2008.

[10] 3GPP, Release 8 V0.0.3, "Overview of 3GPP Release 8: Summary of all Release 8 Features," November 2008.

[11] M. Hsieh and C. We, "Channel Estimation for OFDM Systems Based on Comb-Type Pilot Arrangement in Frequency Selective Fading Channels," IEEE Transactions on Wireless Communication, vol. 2, no. 1, pp. 217-225, May 2009.

[12] 3GPP TS 36.211, Physical Channels and Modulation, 3GPP Technical Specification, Rev. 8.9.0, 2009.

[13] M. Morelli and U. Mengali, "A Comparison of Pilot-aided Channel Estimation Methods for OFDM Systems," IEEE Transactions on Signal Processing, vol. 49, pp.3065–3073, December 2001.

[14] W. Abdallah, "Adaptive Minimum Bit Error Rate Beamforming Assisted QPSK for Pre-FFT OFDM Communication System," NNGT Int. Journal on Networking and Communications, vol. 1, July 2014.

[15] S. Chen, L. Hanzo, N. N. Ahmad and A. Wolfgang, "Adaptive Minimum Bit Error Rate Beamforming Assisted QPSK Receiver," IEEE Communications Society, 2004.

[16] S. Chen, A. Livingstone, H.Q. Du and L. Hanzo, "Adaptive Minimum Symbol Error Rate Beamforming Assisted Detection for Quadrature Amplitude Modulation," IEEE Trans. on Wireless Communications, vol. 7 , no. 4, pp. 1140-1145, April 2008.

[17] S. R. Seydnejad and S. Akhzari, "Performance Evaluation of Pre-FFT Beamforming Methods in Pilot -Assisted SIMO-OFDM Systems," Telecommunication Systems, Springer Science and Business Media, March 2015.

209

"Minimum Bit Error Rate Assisted QPSK for Pre-FET Beamforming in LTE OFDM Communication Systems", Waleed Abdallah, Mohamad Khdair and Mos'ab Ayyash.

ملخص العمل:

في هذا البحث، تمت دراسة أثر استخدام إشارة مرجعية على شكل معين (Diamond Shape Pilot Signal) لتقدير الإشارات المرسلة عبر قناة تستخدم نظام تقسيم الترددات المتعامدة (OFDM). ومن شأن هذا الترتيب أن يقلل من عدد الإشارات المرجعية المرسلة عبر قناة الاتصال ، وهو ما يؤدي بدوره إلى زيادة تدفق البيانات عبر القناة مع المحافظة على دقة مقبولة في تخمين القناة. وهنا يتم الجمع بين مصفوفة الهوائيات المتكيفة (Adaptive Antenna Array) ونظام تقسيم الترددات المتعامدة (OFDM) بهدف التغلب على التداخل بين الرموز (ISI) والتداخلات الاتجاهية.

في هذا البحث يتم الحصول على مجموعة الأوزان المثالية لتشكيل الشعاع بناءً على خوارزمية أدنى معدل الخطأ (MBER) مع استخدام إشارة مرجعية على شكل معين في أنظمة إرسال تقسيم الترددات المتعامدة (OFDM) لشبكات الاتصال للجيل الرابع (LTE)، حيث تم افتراض أن قناة الإرسال هي قناة خبو متعددة المسارات (Multipath Fading Channel). تبين نتائج المحاكاة أن إرسال الإشارات باستخدام الترميز التربيعي المعتمد على إزاحة الطور (QPSK) بناءً على خوارزمية أدنى معدل الخطأ (MBER) يستفيد من عناصر مصفوفة الهوائي على نحو أكثر فعالية من التقنية المعيارية القائمة على الحد الأدنى للخطأ في المتوسط التربيعي(MMSE).

A RULE-BASED APPROACH TO UNDERSTAND QUESTIONS IN ARABIC QUESTION ANSWERING

Emad Al-Shawakfa

CIS Department, Faculty of Information Technology, Yarmouk University,
Irbid 63-211, Kingdom of Jordan
shawakfa@yu.edu.jo

(Received: 30-Jun.-2016, Revised: 08-Sept.-2016, Accepted: 09-Nov.-2016)

ABSTRACT

Research on Arabic Natural Language Processing (NLP) is facing a lot of problems due to language complexity, lack of machine readable resources and lack of interest among Arab researchers. One of the fields that research has started to appear in is the field of Question Answering. Although some research has been done in this area, few have proved to be effective in producing exact relevant answers. One of the issues that affected the accuracy of producing correct answers is proper tagging of entities and proper analysis of a user's question. In this research, a set of 60+ tagging rules, 15+ Question Analysis rules and 20+ Question Patterns were built to enhance the answer generation of Natural Language Questions posed over some corpora collected from different sources. A QA system was built and experiments showed good results with an accuracy of 78%, a recall of 97% and an F-Measure of 87%.

KEYWORDS

Arabic; Question Answering; Question Analysis; Tagging, NLP.

1. INTRODUCTION

After the computer revolution and the rapid spread of computer usage around the world, many researches were conducted in the field of Natural Language Processing (NLP) toward providing a better and easier way of interaction and usage of computers and their applications by different users; especially, naïve ones. For this, we started to see different applications covering many language computing fields like Stemming, Information Retrieval (IR), Information Extraction (IE), Question Answering (QA), Text Classification and Categorization (TC), Machine Translation (MT), among many others. Such research, at the beginning of the era, was mainly conducted using Latin-based languages; especially English, which has formulated the foundation for interfaces with computers.

With the increased volume of information stored in computers; especially databases and recently the Web, people started to look for different ways to help in extracting needed information from different sources of data through expressing their requests in their own daily used natural languages without any technical experience or knowledge. For this, different search engines have started to appear, enabling users to look for information on the web using different Information Retrieval tools.

When one poses a question to a search engine, then based on query words, one might get either too little or too much of documents as a result to his/her query and thus need to dig more into the results to obtain proper answers he/she is looking for; which might become more time consuming. Obtained results are even sometimes far away from whatever actually needed and hence disappointing. Such disappointment is related mainly to the IR algorithm used by the search engine(s) in retrieving only possible relevant documents; rather than exact answers we are looking for. The concept of Question Answering came as a rescue to solve such a problem.

As defined by [1], Question Answering (QA) is: “the task to automatically providing an answer for a question posed by a human in natural languages”. Unlike IR systems, QA systems seek to obtain an Exact Answer to a given question from a list of documents rather than to have an answer being represented as a list of relevant documents; as is the case with search engines. For most of existing QA systems today, results have not always been satisfactory to the users; especially, for the case of Arabic. However, to obtain an exact result of a query, a more detailed process has to be carried on from the point of posing a query to the point of obtaining a requested answer.

Many Question Answering Systems have started to appear in the early 1960s by introducing systems that can be used to extract information from databases using English [2]. Since then, more systems have started to appear, like ([3]-[5]). A recent survey of versatile question answering systems was given by [6]. Research on enhancing the results of QA systems was also developed ([7]-[10]).

Research in the field of Arabic Natural Language Processing (ANLP); especially the field of Question Answering, has lagged behind research in its counterpart Latin-based languages due to many reasons. The complexity of Arabic language itself, the lack of support for Arabic by computers in the early history of computers and the fact that most of the Arabic content on the web at early stages of the digital era was in non-searchable image format are some of such reasons. Furthermore, the lack of standardized machine readable resources, as well as the lack of researchers and/or users who are willing and interested in working with Arabic have made it more difficult to conduct research in the field of Arabic NLP. Regardless of this, many researches in the field of Arabic Question Answering (AQA) have started to appear. Some of the developed systems were AQAS [11], QARAB [12], ArabiQA [13], QArabPro [14], AQuASys [15], QA4MRE@ CLEF [16], among many more.

As the public would say: “Understanding a question is almost half the answer” and since computer systems did not reach that level of intelligence to be able to understand questions properly by themselves, more research in the field of question analysis and understanding are still needed to construct proper answers to posed question(s) by the user(s).

To conduct this research, the researcher has heavily searched the literature for Question Analysis and understanding rules and found out that only very few researchers built a very little number of such rules; a maximum of five rules were found in [14]. In this research and to enhance the accuracy of answer generation, the researcher built more than 15+ detailed Question Analysis rules that are completely different from those available in the literature, combined with 60+ tagging rules and 20+ question patterns. This, as well as the lack of research in Arabic Question Answering, constituted the main contribution and objective behind this research. Since the majority of Arabic QA systems are rule-based, the researcher has adopted this approach in this research as well. Further enhancement of the existing rules as well as building new other rules constitute a future work.

2. RELATED STUDIES

There are different categorizations of QA systems in existence today, depending mainly on what, how and where from the QA system is trying to answer the question(s). Question Answering systems that are looking only for facts or definitions are called Factoid or Definitional QA Systems [1]. Systems that are looking for a more detailed answer beyond a fact or a definition are called Non-Factoid or Passage Retrieval Systems as also indicated by [17], or even sometimes called List QA Systems [18]. Systems that are looking for a reason or an explanation to some happening, are called Why-QA Systems.

As for the domain being searched, QA systems that deal with limited domains are said to be of Restricted Domain in comparison to Open Domain QA systems that deal with a non-restricted open domain text; like the Internet. Non-restricted open domain QA systems are sometimes referred to as Web-based QA systems. Furthermore, systems might deal with either structured or unstructured sources of data. For structured data, the system would be an interface to a database; as the data would be stored into a database. On the other hand, unstructured data does not have a uniform structure and thus cannot be stored into a database ([18]-[19]). Such systems constitute the majority of today's developed QA systems due to advances in NLP tools that would enable better Information Retrieval capabilities. In addition, there are some types of QA systems that look only for Yes/No answers, like [20]. Other QA systems could also be categorized as either Shallow or Deep QA systems, depending on how much semantic and/or syntactic analysis is being performed to obtain the answer [21].

Another categorization of QA systems is given by [19]. This categorization is based on the methods used for the extraction of the answer. Information Retrieval/Extraction methods would give the first category of QA systems to refer to Web-based and IR/IE-based QA systems. The second category refers to systems that use reasoning in the extraction of the answer. This type of category would refer to Domain-oriented QA systems and Rule-based QA systems. Table 1 gives such characterization of QA System types.

Table 1. Characterization of QA systems [19].

Dimensions	QA Systems based on NLP and IR	QA Systems Reasoning with NLP
Technique	Syntax processing, Named Entity tagging and Information Retrieval	Semantic analysis or high reasoning
Data Resource	Free text documents	Knowledge base
Domain	Domain Independent	Domain-oriented
Responses	Extracted Snippets	Synthesized responses
Questions Dealt with	Mostly Wh- type of questions	Beyond the Wh- type of questions
Evaluations	Use existing Information Retrieval	N/A

Most of Arabic QA systems are of either Factoid or definitional type looking for short answers. Very few have managed to deal with Why and How (much/many/to) types of questions. Examples of such Arabic QA Systems are that of [14], [22] and [18]. Table 2 gives a very good comparison between some of the existing Arabic Question Answering Systems given by [18].

The first Arabic Question Answering System was introduced by [11] and was called AQAS. The AQAS system is a knowledge-based QA system that extracts answers from structured data stored into a Database. The authors of AQAS did not report any testing or evaluation results for their system, so no one could give any advantages or disadvantages of such system.

According to [1], there was no work performed on Arabic Question Answering from 1993 until 2002, when [12] introduced a rule-based Factoid QA system for Arabic called QARAB which deals with unstructured data from documents collected from Al-Raya Newspaper with 113 Factoid questions fed into the system. However, QARAB did not handle the two types of questions كيف، "لماذا” (How and Why), because they require long and complex processing.

In 2006, [23] introduced an ongoing implementation of a Factoid Arabic QA system that was used for the purposes of QA tracks in the Cross Language Evaluation Forum (CLEF) and Text Retrieval Conference (TREC) competitions. In their paper, the authors have only introduced partially implemented modules of the system in which the Named Entity Recognition (NER) module as well as a Java Information Retrieval System (JIRS) module were embedded. However, this system was completed and introduced later on by [24], in which the effect of correctly identifying a Named Entity (NE) to produce correct answers is emphasized.

Table 2. A comparison between some existing Arabic QA systems ([18]).

Main Features	**QARAB**	**ArabiQA**	**ArQA**	**QASAL**	**AQuASys**	**JAWEB**
Web-based system	×	×	×	×	×	√
Retrieves answers from a corpus	√	√	√	√	√	√
Retrieves answers from the web	×	×	×	×	×	×
Natural language processing tools	√	×	√	√	√	√
Named entity recognition	×	√	√	√	×	×
Answers factoid questions	×	√	√	√	√	√
Answers open domain questions	×	√	×	√	√	√
Supports multiple languages	×	×	×	×	×	×
Supports Arabic language	√	√	√	√	√	√
Provides short answers block	√	√	√	√	√	√
Measures answer precision	-	√	√	-	√	√
Measures answer recall	-	√	√	√	√	√

A Factoid and Definitional Arabic Question Answering system called QASAL was introduced by [25]. The authors have used NooJ platform and local grammars to help in obtaining the right answers. For the Factoid questions, authors have used the collection of Tunisian books as a corpus. However, for the definitional type of questions, the authors used the Arabic version of Google search engine as a web resource to look for Arabic documents with 43 definition questions. According to the authors, 94% accuracy for definitional questions was obtained.

An Arabic QA system (QAS) to answer short Factoid questions in Arabic was described by [26]. The authors based their testing on a collection consisting of 25 manually collected documents that were gathered from the web in addition to some relevant documents that were provided by the authors applying 12 questions to the set. Authors reported different recall levels of {0, 10 and 20%}, where the interpolated precision was equal to 100% and at recall levels 90 and 100% to be equal to 43%. As is the case with QARAB, QAS did not handle the “كيف، لماذا"” (How and Why) types of questions due to the complex processing needed.

An Arabic Definition Question Answering system named DefArabicQA was introduced by [27]. This system answers questions of the form “What is X?” with the web as the data source. The authors claim that their system provides effective and exact answers to definition questions expressed in Arabic using little linguistic analysis and language understanding capabilities. To evaluate their system, two experiments were conducted with Google only as a web source in the first experiment and Google coupled with Wikipedia as the web source for the second experiment. The experiments reported a Mean Reciprocal Rank (MRR) score of 0.7 and a question rate of 0.54 for the first experiment and an MRR score of 0.81 and a question rate of 0.64 for the second.

A rule-based Question Answering System for Arabic called QArabPro was developed by [14]. The system performs reading comprehension texts and tries to answer questions posed upon such texts. QArabPro assumes that the answer must exist within one of the documents that were used as a corpus. The authors claim that they have answered all types of questions including the “كم، لماذا” How (much/many) and Why types in contrast to other existing QA systems that avoided such types of questions due to their complexity. To test their system, they used a set of documents that were collected from Wikipedia with 75 documents and 335 questions. According to the authors, the claimed results were of 93% Precision, an 86% Recall and an F-measure of 89%. Obtained test results showed an overall accuracy for “كم” (How much/many) of 69% and 62% for “لماذا” (Why) questions that were handled. However, QArabPro did not handle “كيف” (How) type of questions.

In [15], a Factoid Arabic QA system named AQuASys was developed. A Factoid natural extension to AQuASys with a web interface and an extended corpus was built by [18] which the authors called JAWEB.

A Question Answering System called IDRAAQ was developed in [28] in the framework of the main task of Question Answering for Machine Reading Evaluation (QA4MRE@CLEF2012).This system was based on keywords and structure levels through query expansion and Distance Density N-gram model-based passage retrieval to improve the results of the system. According to the authors, IDRAAQ has obtained promising results with the QA4MRE framework; especially with Factoid type of questions.

In the QA4MRE@CLEF2012 framework, a work on Arabic Question Answering was given by [16]. According to the authors, the work of [16] has obtained an accuracy of 0.19 with very little of reasoning and inference; an issue requested in analyzing and understanding documents for this framework.

An Arabic Language Question Answering Selection In Machines called ALQASIM was introduced in [29]. ALQASIM was used to answer Multiple Choice questions of the QA4MRE. According to the authors, a novel technique was used in understanding and analyzing test documents which led to an accuracy of (0.31) in comparison to accuracies of (0.13) obtained by IDRAAQ [28] and (0.19) by the approach used in [16].

An Entailment-based Why Arabic Question Answering (EWAQ) system was introduced by [30]. According to the author, EWAQ enhanced the accuracy of “Why” questions by improving the re-ranking of passages that are relevant and retrieved by many search engines as possible answers. She claimed that the accuracy of her system has improved over that of search engines; Google, Yahoo and Ask.com.

In most of existing types of QA systems; especially Factoid QA systems, the search will be for a Named Entity as part of an answer; or even the answer itself. In English and other Latin-based languages, it is very easy to locate a noun with all of its categories in a given text due to the capitalization feature that exists in the language itself. However, since Arabic does not support any capitalization of letters and is a highly inflected and derived language with rich morphology and complex syntax, the identification is not straight forward; a process that would be more difficult to carry on for Arabic. For this, a research on handling Named Entities in Arabic was conducted.

Once introduced into research during the sixth Message Understanding Conference (MUC-6), the concept of Named Entity (NE) did not just cover only proper nouns, but also included other types. The types, or classes, that were introduced by MUC-6 for NE were ENAMEX (referring to person names, locations and organizations), NUMEX (referring to money and percentage [numerical] expressions) and TIMEX (referring to time and date expressions).

There are two approaches to build Name Entity Recognition (NER) systems in Arabic; a rule-based approach like that of NERA system was introduced by [31] and a Machine Learning (ML) approach like ANERSys system was built by [24]. Some of the systems that adopted the rule-based approach are: TAGARAB [32], PERA [33], ARNE [34] among many others. As for the ML approach, many researches have been conducted using this approach like those of ([13],[31] and [35]-[37]), among many others. For more information, one can refer to a very good and recent survey of Arabic Named Entity Recognition systems given by [38].

3. THE QUESTION ANSWERING PROCESS

The generic Question Answering System consists of three major modules; namely, the Question Analysis and Understanding Module, the Document/Passage Retrieval Module and the Answer Extraction and Response Generation Module. This research is part of an ongoing research on Arabic Question Answering and is concerned with the first module. Work on other modules of the Arabic Question Answering; IR and Response generation, is currently being carried on.

3.1 Question Analysis and Understanding Module

One of the most important steps in the Question Answering process is the issue of question understanding; a question must be properly analyzed to clarify what is meant by such question thus enabling us to be directed in the proper path of finding the right and exact answer to our query. In this module, a correct understanding of what a question might be looking for; or what is known as the Scope (or Focus) of the Question and Question Type constitute a crucial step toward providing the right answer to a given question.

Regardless of the natural language being used, a question type would fall into one of the following categories:

1) Who/Whose: such type of question will be looking for animate objects such as a person.
2) What/Which: such type of question will be looking for inanimate objects, like an entity or a thing.
3) Where: such type of question usually looks for a place or location.
4) When: such type of question will be looking for a time or time-related information.
5) Why: such questions are usually not easy to answer, but they will be looking for a reason or a cause for the happening of some action.
6) How: depending mainly on what follows; for instance, in the case of How much or How many, such questions will be looking for numbers or quantity. However, if How is not followed by much or many, these questions will be looking for a process or procedure on doing some action.

In Arabic, there are more ways to ask questions than in English. Arabic uses the same WH interrogative nouns of English in addition to more than one way of representation of some of the interrogatives. For instance, interrogatives like Who “من”, Whose “لمن”, What “ما” “ماذا” “مما”, Which “أي”, Where “أين”, When”متى” “أيّان”, Why “لماذا”, How (much/many) “كم” and How “كيف”, are some ways of asking questions in Arabic. In addition, Arabic Interrogative nouns include other particles that are related to the expected question types and/or scopes, like:

1) To be type "هل": This is an interrogative tool where such type of question usually would be looking for a Yes/No answer. Examples on such: "هل تمطر في الخارج؟" (Is it raining outside?) and "هل اتى محمد؟" (Did Mohammad come yet?).
2) Hamza "ء": This interrogative tool in Arabic is used as a disapprovingly tool "أداة استنكارية". As an example: in the Holy Quran "ءإله مع الله؟" (Is there a God with Allah?), the answer to such questions would be either Yes or No, but in our example the answer must definitely be No.

The Arabic interrogative word "مما" is actually a combination of the two words "من" and "ماذا" (From What), but when combined in Arabic it is reduced to "مما".

In addition to the above, Arabic has more indirect ways to ask questions. For instance, in Arabic, using the phrase "في أي" (In What/Which), we could say "في أي عام ولد ابن خلدون؟" In What year Ibn Khaldoun was born? Or "في أي بلد ولد ابن خلدون؟" (In which country was Ibn Khaldoun born?). Such questions are formulated using the phrase: "في أي" (In What/Which); a phrase used to formulate question types related to Time (In What) and Place (In Which) concepts. Also, one can use the phrase "من أين" (From Where) to usually refer to a source (Location, Method, …etc.) like asking"من اين اكتسبت هذا المال؟" (From where did you earn this money?); referring to the source of the money, or ask "من أين أتى الغزاة؟" (From where did the attackers come?); referring to the location the attackers came from.

On the other hand, the phrase"فيـمَ" (In Where) refers to questions that ask about a target Named Entity like "فيما انفقت مالك؟" (Where did you spend your money?). This type of question might require some extra semantic analysis and search capabilities to reach a proper answer.

In Arabic, once the type of a question is identified, from the interrogative noun we can identify the Gender and the expected scope (Target Answer type); which would be used later in the Response Generation module; another part of the ongoing research on Arabic QA. This can be illustrated in formulating what is known as Question Patterns. For instance, if one asks the following question:

"من هو محمد الفاتح؟"(Who is Mohammad Al-Fatih?)

then using the following Question Pattern for Definitional type of question Who+be +< topic> (من هو|هي + <الموضوع>), the question type and scope would be identified from the usage of the interrogative noun "من" as being a Definitional type of question looking for an NE→ Enamex. From the word "هو", referring to the question pattern would indicate that gender is identified as masculine. The Target Answer (topic) would be referring to the Named Entity "محمد الفاتح" Mohammad Al-Fatih. Furthermore, a question like:

"متى حدثت معركة الكرامة؟" (When did Al-Karamah Battle occur?)

then using the following Question Pattern for Temporal type of question: When+Verb+<topic> (متى+ فعل + <الموضوع>), the question type and scope would be identified from the usage of the pattern of "متى" as being a Factoid type looking for an NE → Timex and the gender would be identified as feminine from both the Taa marboutah "ـة" in "معركة الكرامة" and the connected pronoun "ت" in the verb "حدثت". The Target Answer (topic) will be looking for a Time or Date value related to the occurrence of the scope "معركة الكرامة"; i.e., a year or a specific date. Table 3 gives a sample of Question Patterns that were identified and used for this purpose.

Table 3. Sample of Question Patterns used in the approach.

Question Pattern	**Type**	**Scope**
من + هو | هي + <الموضوع>	Definitional	Definition of NE→ Enamex
لمن + إسم اشارة + <الموضوع>	Factoid	NE→ Enamex
لمن + <الموضوع> + اسم إشارة	Factoid	NE→ Enamex
ما + هو|هي + <الموضوع>	Definitional	Definition of NE
كم [تبلغ | يبلغ] + Countable-Qty.-Term + < تكملة السؤال >	Factoid	NE→ Numex
متى + فعل + <الموضوع>	Factoid	NE→ Timex
كيف + فعل + <بقية السؤال>	Method	List of steps

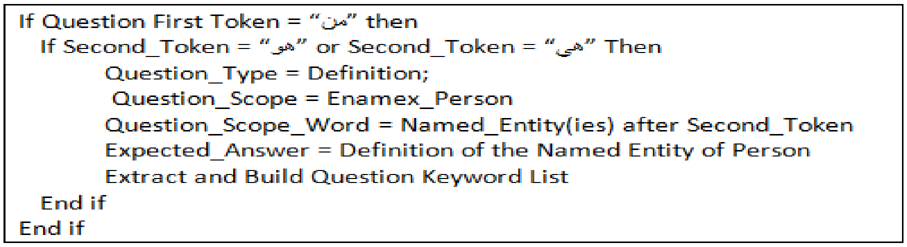
3.1.1 Question Processing

To process a question, a set of rules and patterns were developed for each question type. Table 4 gives Question types and scope. The current version of the implemented system can answer some of the rules for (How and Why) “كيف، لماذا”. Further assessment to produce more accurate answers is needed; which constituted part of the future research as well. The rules for the interrogative noun (List) “اذكر” were not tested in our approach, since they require more semantic analysis. Different rules for interrogative nouns in Arabic were built and implemented within the system. Figure 1 and Figure 2 give the rules used for “من” and “متى، إيان، أيان”, respectively.

So far, the rule in Figure 1 deals with the question pattern (من هو|هي + <الموضوع>), in which the second token of the given question must be either “هو” or “هي”. The rule for the case where the interrogative noun “من” followed by a verb was not built and implemented, since it requires more analysis. So, if a question like من جاء مع محمد إلى الجامعة ؟ (who came with Mohammed to the University) was asked, then it will not be answered, as it is not handled properly in the current approach due to more analysis. Figure 3 describes the question processing applied in this approach.

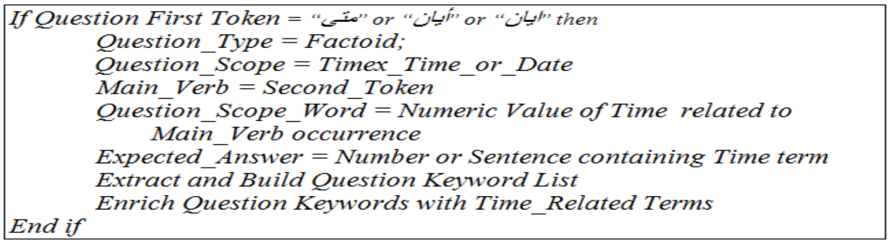
Table 4. Question types and scope in Question Understanding.

Question Type	**Question Words**	**Scope**
Factoid	متى ، إيان ، أيان	Timex, looking for a time
Factoid	لمن	Enamex, looking for Named Entity → Person
Factoid	في اي + Time_Term	Timex, looking for a time
Factoid	في اي + Location_Term	Enamex, looking for a Location
Factoid	أين	Enamex, looking for a Location
Factoid	كم	Looking for Numeric Value
Definition	من	Enamex, looking for definition of Named Entity → Person
Definition	ما	Enamex, looking for definition of Named Entity → Location or Organization
Causal	ماذا	Looking for result and/or cause
Method	كيف	Looking for a process to do something
Purpose	لماذا	Looking for a reason for doing something
List	اذكر، ما هي	Looking for a list of steps. In case ما هي, we need further analysis like next word is طريقة، مقادير، خطوات،
Yes/No	هل	Looking for either Yes or No as a reflection of action



```
If Question First Token = “من” then
  If Second_Token = “هو” or Second_Token = “هي” Then
        Question_Type = Definition;
         Question_Scope = Enamex_Person
        Question_Scope_Word = Named_Entity(ies) after Second_Token
        Expected_Answer = Definition of the Named Entity of Person
        Extract and Build Question Keyword List
  End if
End if
```

Figure 1. Rule for interrogative noun “من”.



```
If Question First Token = “متى” or “أيان” or “إيان” then
        Question_Type = Factoid;
        Question_Scope = Timex_Time_or_Date
        Main_Verb = Second_Token
        Question_Scope_Word = Numeric Value of Time  related to
            Main_Verb occurrence
        Expected_Answer = Number or Sentence containing Time term
        Extract and Build Question Keyword List
        Enrich Question Keywords with Time_Related Terms
End if
```

Figure 2. Rule for interrogative nouns “متى”, “إيان” or “أيان”.

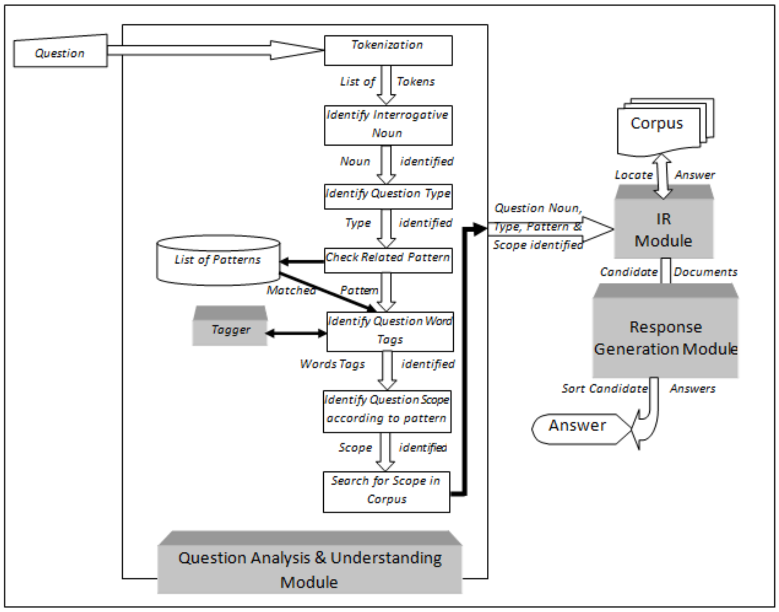


Figure 3. Question processing of the research.

3.1.2 The Tagger System

Arabic is a highly inflectional and derivational language, which exists in three different forms today. The first is the Classical Arabic (Fus’ha), which is the language of the Holy Quran and poetry of ancient pre-Islam era before about 1500 years. This form is completely diacriticized and is used today in the teachings of the Holy Quran and any Islam-related issues as well as Arabic Language educational books. The second form is the Modern Standard Arabic (MSA), which is the language of the media (Newspapers, TV, Internet, …etc.) and is used daily for official purposes by

Arab countries and their organizations as well as the United Nations as one of the official languages. Finally, the third form is the Colloquial Language (CL), which refers to different dialects that are used in different Arab countries amongst people on the streets [38].

The majority of research in Arabic NLP in existence today has targeted Modern Standard Arabic; a vowel free form of the language that introduces high ambiguity. For instance, if we look at the MSA word “علم”, without diacritic symbols shown on the word, it could be interpreted to mean ‘Science’ “عِلْم” (Ilm), ‘Flag’ “ عَلَمْ” (Alam), ‘been known’ (عُلِمْ) (Olim), …etc. An example of similar derivations is given in Table 5 for the root ق ب ل [39]. This ambiguity is one of the main reasons why research on Arabic NLP did not reach the levels of its counterpart Latin-based languages.

Table 5. Different derivations of the root ق ب ل [39].

English Concept	Arabic word
Tribe	قبيلة
to meet	تقابل
Before	قبل
Future	مستقبل
To receive	استقبل
To come to	أقبل
Kiblah	قبلة

Arabic words are either native Arabic words or Arabized; brought from other languages. According to [40], the following ١٢ letters (ض ط ظ ق ع ح ة ء ؤ ئـ ى ص) are restricted to Arabic native words where none of them is used for either Transliterated and/or Arabized words. An Arabic native word is mainly classified into one of three types; Noun, Verb and Particle. When performing NLP using Modern Standard Arabic form; especially in Question Answering systems, it is very important to identify the type of the word we are dealing with; especially nouns. Figure 4 gives the Part-of-Speech categorization of Arabic words [41].

Of course, before we could identify Target Answer types, we need to tag question words properly; and this is the task of the Tagger. In this research, we have built a tagger that is a combination of both rule-based and word weights “أوزان” to identify the proper tag value of a question word. The used rules and weights, and hence the built tagger, are from two unpublished research works by the author.

In this research, 35+ tagging rules were identified from literature that could be used for the proper identification of Nouns. For instance, the following types of words are all identified as Nouns: Proper nouns, Action nouns, Genus nouns, Agent nouns, Patient nouns, Adjectives, Time, Place, Instrument, Adverbs and Demonstrative Nouns. In addition, any words that start with the definite article “ال” (the) or end with “ـة” are considered Nouns. In most cases, words that end with “ـاء” are considered nouns. In this research, we used question patterns which enforce us to use nouns and not verbs that end with “ـاء”. If tagging rules failed in identifying Nouns properly, then a set of Named Entities that was compiled by [31] are used to help in properly classifying Named Entities. In addition, 15+ rules are identified to tag verbs and 12+ rules are identified and used to identify particles. Such tagging rules are needed to help in proper identification of answers. In addition to the rules, 72+ word weights were used to help in the proper identification of word tags and are implemented within the system.

The tagging algorithm is given as follows:

1) Get the word from the Tokenized list of words.
2) Assign the value “Unknown” to Tag_value .
3) Check if word is a Particle. If so, assign Tag_value = “Particle” and go to step 11.
4) Check if word is a Noun. If so, assign Tag_value = “Noun” and go to step 7.
5) Check if word is a Verb. If so, assign Tag_value = “Verb” and go to step 11.
6) If Tag_value equals “Unknown”, Check the NE database for a match.
7) If the word is found in the Location NE table, assign Tag_value = “NE_Loc” and go to step 11.
8) If the word is found in the Person NE table, assign Tag_value = “NE_Hum” and go to step 11.
9) If the word is found in the Organization NE table, assign Tag_value = “NE_Org” and go to step 11.
10) If Tag_value is still “Unknown”, assign “Failed or Foreign” to Tag_value.
11) Return Tag_value and Exit.

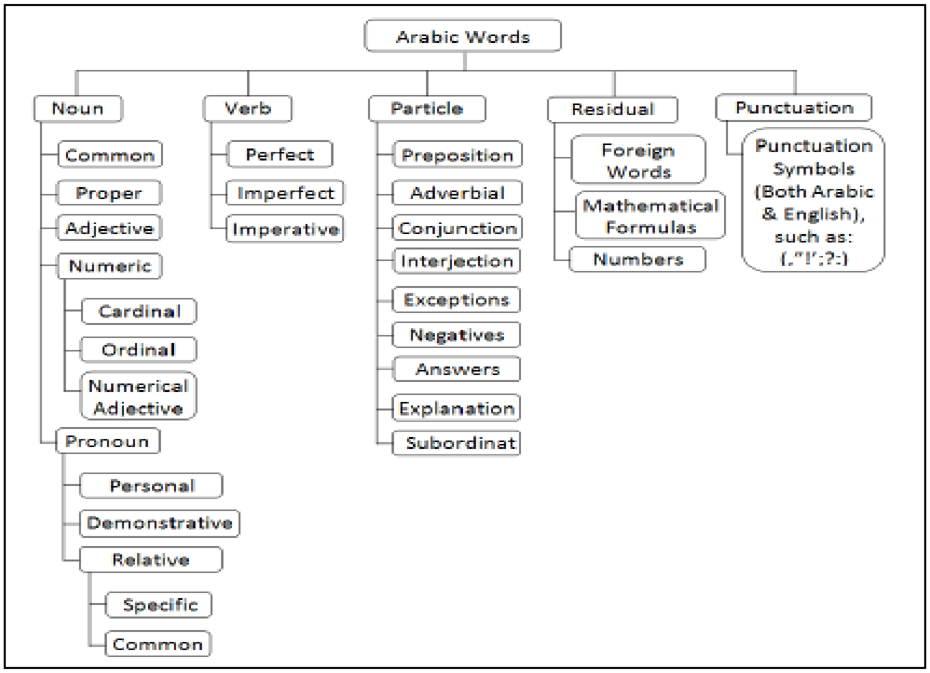
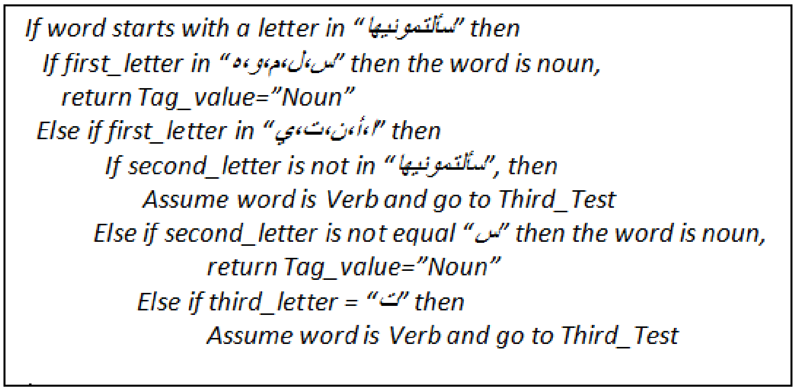


Figure 4. Arabic Part-of-Speech [41].

In this algorithm, the Tag_value is needed for proper identification of question words to better match question types and scope with Question Patterns. If the Tag_value was identified as either Particle or Verb, then the algorithm just returns the Tag_value. If, however, the Noun was identified as a Tag_value, or the Tag_value was not identified (Unknown Tag_value), then a search is needed into the Named Entity tables to better identify the Question Scope; especially with Factoid and Definitional types of questions. Figure 5 gives a snapshot of Noun tagging rules.



```
If word starts with a letter in "سألتمونيها" then
  If first_letter in "س،ل،م،و،ه" then the word is noun,
    return Tag_value="Noun"
 Else if first_letter in "ا،أ،ن،ت،ي" then
      If second_letter is not in "سألتمونيها", then
          Assume word is Verb and go to Third_Test
     Else if second_letter is not equal "س" then the word is noun,
               return Tag_value="Noun"
          Else if third_letter = "ت" then
               Assume word is Verb and go to Third_Test
```

Figure 5. Snapshot of Noun tagging rules.

3.2 The Information Retrieval/Extraction Module

The focus of this research did not involve building a state-of-art IR engine. For this, we have adopted the approach followed by many researchers like that of QARAB, QArabPro, …etc. that was based on Salton Vector Space Model (VSM) to search and retrieve relevant documents using a relational database system. In this system, we keep data in tables where the major tables are:

1) A Document table, where we store different corpora files into the database.
2) A Stop words table to keep a list of Arabic Stop words. Here, we use a list of Arabic Stop words that was downloaded from [42] and contains 13000+ Stop words.
3) A Named Entity table, which stores a list of Named Entities. Here, we use the list compiled by [31] that contains 5000+ person, location and organization names.

In addition to the above mentioned tables, the approach includes some other supporting tables that are needed during the Question Analysis and Information Retrieval process.

The Vector Space Model (VSM) defines a vector that represents each document and a vector that represents the query. The model works by assigning weights to index terms in both the queries and the documents, which are then used to calculate the degree of similarity between each document and the query. In this research, the Cosine similarity is used as a measure of similarity between the query and the retrieved relevant documents to find the most appropriate answers.

After the proper formulation of the query posed by the question through enriching it with extra keywords, the IR model retrieves the most relative documents to the query that need to further choose among. The choice is made after sorting documents using some scoring mechanism. Given a document dj and a query q, the Cosine Similarity value that could be used to provide such score is calculated as:

$$CosSim\,(d_j, q) = \frac{d_j . q}{\left|\overrightarrow{d_j}\right| . \left|\overrightarrow{q}\right|} = \frac{\sum_{i=1}^{t}(W_{ij}.W_{iq})}{\sqrt{\sum_{i=1}^{t} W_{ij}^2 . \sum_{i=1}^{t} W_{iq}^2}};$$

where w_{ij} is the weight of the term i in the document j and w_{iq} is the weight of the term i in the query. The term weight is the Normalized term weight and is calculated as:

$f_{i,j} = tf_{i,j} / \max tf_{i,j,}$ where:
$f_{i,j}$ =Normalized frequency,
$tf_{i,j}$ = Frequency of term i in document j and
max $tf_{i,j}$ = Maximum frequency of term i in document j.

Since we will be looking for the most appropriate answer, then the document with the highest score will be selected as the candidate document containing the answer. Once the document is identified, a pattern matching is performed to find the best match between the different sentences in the document and the pattern scope and shape to retrieve the best answer.

4. The Data Set

For the purposes of this research, we have started looking for different sources of Data Sets that could be used. Although many corpora were collected, only the Data Set built for QArabPro by [14] was used, due to an access that was thankfully obtained to both the questions and the documents from the lead author. The Data Set consists of 335 questions posed over 74 documents from different categories; which has formulated the basis for testing and evaluating our approach. In addition, ANERSys Named Entity corpus built by [31] was used to help in the tagging process. The testing results are reported in the experimental section of this paper.

After extensive search of the Internet and repositories, other corpora were also collected like that of the Holy Quran [43], List of Stop words from [42] and a corpus of 1256 documents collected by the author from both Al-Rai (www.alrai.com) and Addustour (www.addustour.com) newspaper sites. However, such corpora were not used in this research due to the lack of experimental results to compare with and to make sure that the approach of this research really works. As a future research, the author is planning to use such corpora and to apply the approach upon them.

5. Research Approach

To achieve the objectives of this research, the following steps were followed as given in Figure 6.

Step 1) Collecting and organizing information on different Arabic Question Answering systems in existence with their related problems.

After extensive search of existing Arabic QA systems, the main problems found to be faced by such Arabic QA systems could be summarized as follows:

(1) The lack of standardized Arabic resources, like an Arabic corpora, Grammar, IR tools, …etc., that could be used as a bench mark to compare with and judge the effectiveness of an approach.
(2) Some of the discussed QA systems did not report their testing results or even did not mention anything about their results, like AQAS and QARAB.
(3) The majority of systems are of ***Factoid*** and/or ***Definitional*** type of QA systems. Very few of them managed to handle the “Why and How” type of question like QArabPro and none were found to deal with ***List*** type of question. This is because more processing and semantic analysis is requested, which requires more elaborative work.
(4) The majority of the systems assume that the answer exists in the set of documents being searched; a limitation of the QA system. But, what if the answer does not exist? None of the researchers said anything about that.

Step 2) Identifying different rules in existence for Analyzing Arabic Questions.

To perform this process, the literature was searched for existing rules that could be helpful and useful for the purposes of Analyzing Arabic Questions. Most of the rules had to deal with the tagging process of Arabic words and a very limited number of rules was clearly mentioned in the papers (5 rules only by [14]). Some authors did not even mention the way in which they have analyzed Arabic questions in their systems. For this, we had to build our own set of rules that matches the logic of dealing with different forms of Questions.

Step 3) Enhancing and/or building rules for Question Analysis.

By analyzing different ways of asking questions in Arabic and from different sources in literature, we have identified six different categories of question types; those are: Factoid, Definitional, Causal, Method, Purpose and List. Although in English type of questions there are six different Question words; Who, When, Where, Which, Why and How, their Arabic counterparts constitute around 15 different variations. The Arabic question words are: من، متى، أين، أي، لماذا، كيف، ممَّ، في اي، ماذا، هل، اذكر، أيان، كم with two variations for each of كم and في أي. Rules were built for each type of Arabic questions.

Step 4) Building and/or collecting proper corpora for testing purposes.

Many corpora were collected from different resources of the Internet and from different scholar sites. The total number of documents in these corpora amounted to more than 50000 text documents in addition to the Holy Quran in Arabic textual format. However, to test the validity of our approach, only the corpus provided by [14] is used.

Step 5) Building an Arabic QA system equipped with the new rules to test the approach.

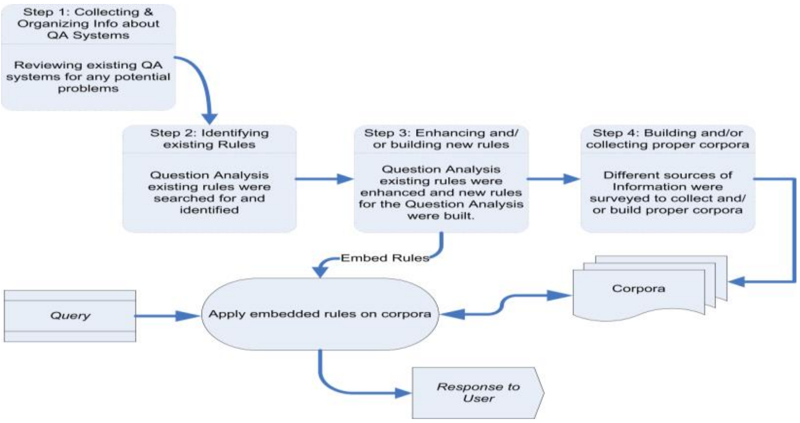

Step 1: Collecting & Organizing Info about QA Systems
Reviewing existing QA systems for any potential problems
Step 2: Identifying existing Rules
Question Analysis existing rules were searched for and identified
Step 3: Enhancing and/ or building new rules
Question Analysis existing rules were enhanced and new rules for the Question Analysis were built.
Step 4: Building and/or collecting proper corpora
Different sources of Information were surveyed to collect and/ or build proper corpora
Embed Rules
Corpora
Query
Apply embedded rules on corpora
Response to User

Figure 6. Generic steps of research approach.

6. IMPLEMENTATION

In this research, a rule-based QA system using Visual Studio.Net 2015 with SQL Server 2014 Express Edition within a DotNet framework version 4.6 under Windows 7 environment was implemented for testing purposes. The system was implemented using the identified and constructed rules. The main screen of the system is given in Figure 7. From this figure, we can notice that the user has the choice to either ask a question directly; by selecting Question option, or load a set of questions from a question file; by selecting the Question File option. A user must identify the corpus to be used for the search purposes. The main reason behind selecting the proper corpus is to work in the same manner as that of the QArabPro for bench marking purposes. Furthermore, to select a file containing a set of questions was a choice to compare the results with those of QArabPro.

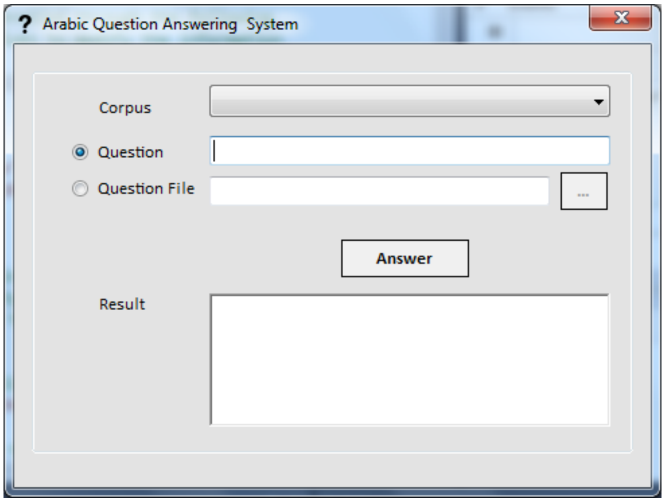


Figure 7. Main screen of the implemented system.

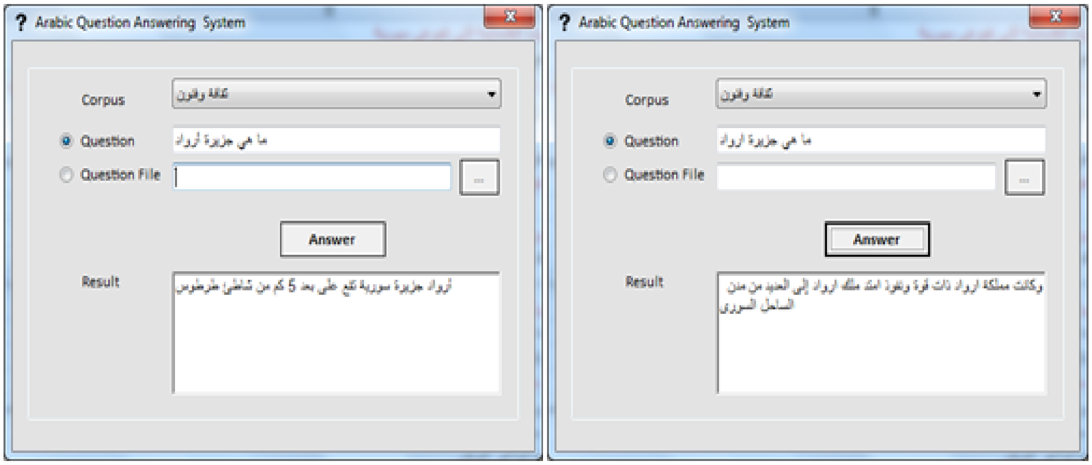
The system makes use of the documents and the Named Entities stored into the database as of well as other tables mentioned previously in the section about the IR module. No document pre-processing or letter normalization is performed. Instead, when storing the document into the database, we store a document ID, a category for which the document belongs and the document text as is.

When conducting research on Arabic NLP, many researchers tend to perform normalization of some letters by converting different versions of alif “ا إ آ أ”into a single version "ا" as well as the case for haa ”ـة ـه” and alif maqsourah “ي ى”. In the author’s opinion, such process will lead to invalid answers in a question answering system and Figure 8 gives the proof. This opinion matches with a finding by [44], in which is proved that the removal of Stop words and Normalization of letters had no significant effect on the retrieval process and might not justify the cost of carrying pre-processing.

In Figure 8, if one asks about Arwad island using alif with hamzah “ما هي جزيرة أرواد؟” and alif without hamzah “ما هي جزيرة ارواد؟”, the results will be completely different; with the one containing the hamzah as the correct answer. If the corpus was written properly without typos, then the question without hamzah will not obtain any answer; since Arwad is usually written with a hamzah.

If normalization was performed on both the documents and queries, incorrect answers might be obtained (as in Figure 8). By eliminating the normalization process and using the question patterns, better results are obtained.

The IR module of the approach was implemented based on the Vector Space Module (VSM) to search for answers among the set of documents and then rank them using the Cosine Similarity measure, in which the document with the highest earned value is selected as a candidate document. To extract the answer, the system performs pattern matching between the question pattern and different sentences in the document that contain the question scope. The sentence that contains the scope and best matches the pattern is returned as the answer.


? Arabic Question Answering System
Answer
? Arabic Question Answering System
Answer

Corpus: ثقافة وفنون
● Question: ما هي جزيرة أرواد
○ Question File:
Result: أرواد جزيرة سورية تقع على بعد 5 كم من شاطئ طرطوس

Corpus: ثقافة وفنون
● Question: ما هي جزيرة ارواد
○ Question File:
Result: وكانت مملكة ارواد ذات قوة ونفوذ امتد ملك ارواد إلى العديد من مدن الساحل السوري

Figure 8. Reason for not normalizing Arabic characters.

7. Experimental Results

To test this approach, the data set used for QArabPro in [14] was used as a bench mark. In this data set, a set of 335 questions and 74 documents were used. However, before we could use this data set, we had to convert all text documents from Windows 1256 code page into UTF-8 encoding. This was a necessity for the success of dealing with this corpus under Windows 7 environment. In addition, the set of documents were stored into the database in their original format where a document ID, category and text were stored. Table 6 gives more information on the number of questions of each question type adopted in this approach from [14].

With IR systems, both Precision (P) and Recall (R) are used as measures to show the efficiency of such systems in retrieving relative documents. When calculating both P and R, it is clearly noted that both have an inverse relationship; if P increases, R should decrease. However, since Question Answering is usually interested in finding an exact answer; not a document (*or list of documents*), then using P; in the author’s opinion, will not be an accurate measure, as it will be difficult with such case to identify True Negative and False Positive answers needed to calculate the Precision. When posing a question, only one of three outcomes will be noticed: a correct answer, an incorrect answer or no answer. As is the case with many researches on QA, we used Accuracy (Acc) instead of Precision; which refers here

to the ratio of correctly answered questions over the total number of questions posed, as a measure. The following formulae show how we measured the efficiency of the approach.

Table 6. Question distribution per question type.

Question Type	Total # Questions	Total # Answered	Correctly Answered	Incorrectly Answered	Not Answered	Accuracy
Definitional	141	136	118	23	5	0.867647
Factoid	128	125	86	36	3	0.688
Causal	40	39	32	6	1	0.820513
Purpose	26	25	18	6	1	0.72
Total	335	325	254	71	10	**0.781538**

$$Recall = \frac{\text{Number of answered questions}}{\text{Total number of asked questions}} = \frac{325}{335} = 97\%$$

$$Accuracy = \frac{\text{Number of correctly answered questions}}{\text{Total number of answered questions}}$$

$$Definitional\ Accuracy = \frac{118}{136}$$

$$Definitional\ Accuracy = 0.867647$$

$$Factoid\ Accuracy = \frac{86}{125}$$

$$Factoid\ Accuracy = 0.688$$

$$Causal\ Accuracy = \frac{32}{39}$$

$$Causal\ Accuracy = 0.820513$$

$$Purpose\ Accuracy = \frac{18}{25}$$

$$Purpose\ Accuracy = 0.72$$

$$Overall\ Accuracy = \frac{254}{325}$$

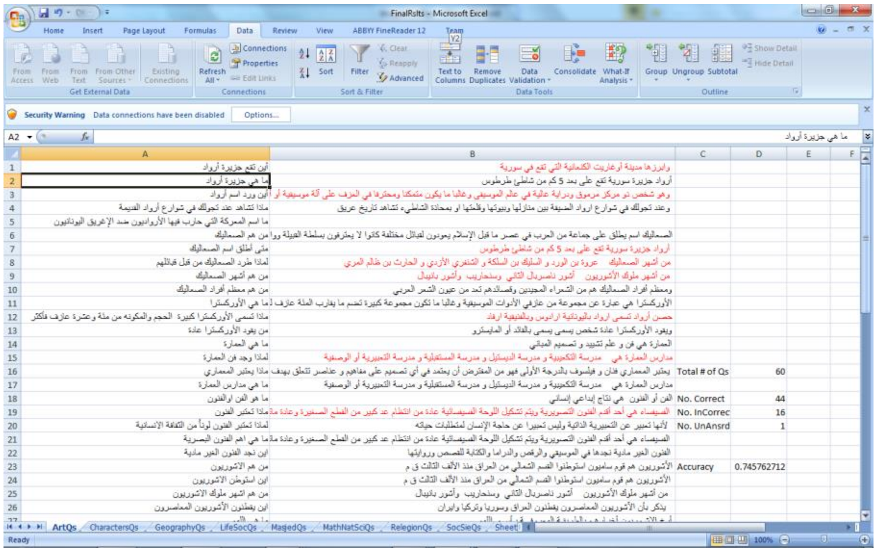
$$Overall\ Accuracy = 0.781538$$

$$Fmeasure = 2 * \frac{\text{Accuracy} * \text{Recall}}{\text{Accuracy} + \text{Recall}} = 2 * \frac{0.781538 * 0.970149}{0.781538 + 0.970149}$$

$$Fmeasure = 0.8656869642.$$

To reach the obtained results, we have isolated different questions per each category into files and then used these files as input to the system. The results were stored in an Excel file showing each question with its corresponding answer. A manual process was then performed by a human expert to validate the answers given by the system. Finally, all results of categories were combined into one file. The number of correctly answered questions, as well as the number of incorrectly answered questions and the number of unanswered questions were manually calculated. Figure 9 shows a snapshot of the combined results in which answers in light colour are incorrect, empty cells indicate unanswered questions, while answers in dark colour refer to correctly answered questions.

Figure 10 shows the number of questions that were answered from each type of question posed and Figure 11 gives the percentage of correctly answered questions for each question type. As can be noticed from Figure 11, the system managed to answer around 72% of Purpose type of questions, 82% of Causal type of question, 87% of Definition type of question and 68.8% of Factoid type of question. Only 2.9% of questions were not answered and 21.2% were incorrectly answered.



Security Warning Data connections have been disabled

A2 ما هي جزيرة أرواد

	A	B	C	D
1	أين تقع جزيرة أرواد	وابرزها مدينة أوغاريت الكنعانية التي تقع في سورية		
2	ما هي جزيرة أرواد	أرواد جزيرة سورية تقع على بعد 5 كم من شاطئ طرطوس		
3	أين ورد اسم أرواد	وهو شخص ذو مركز مرموق ودراية عالية في علم الموسيقى وغالبا ما يكون متمكنا ومحترفا في العزف على آلة موسيقية أو		
4	ماذا تشاهد عند تجولك في شوارع أرواد القديمة	وعند تجولك في شوارع ارواد الضيقة بين منازلها وبيوتها وقلعتها او بمحاذاة الشاطيء تشاهد تاريخ عريق		
5	ما اسم المعركة التي حارب فيها الأرواديون ضد الإغريق اليونانيون			
6	من هم الصعاليك	الصعاليك اسم يطلق على جماعة من العرب في عصر ما قبل الإسلام يعودون لقبائل مختلفة كانوا لا يعترفون بسلطة القبيلة ووا		
7	متى أطلق اسم الصعاليك	أرواد جزيرة سورية تقع على بعد 5 كم من شاطئ طرطوس		
8	لماذا طرد الصعاليك من قبل قبائلهم	من أشهر الصعاليك عروة بن الورد و السليك بن السلكة و الشنفري الأزدي و الحارث بن ظالم المري		
9	من هم أشهر الصعاليك	من أشهر ملوك الأشوريون أشور ناصربال الثاني وسنحاريب وأشور بانيبال		
10	من هم معظم أفراد الصعاليك	ومعظم أفراد الصعاليك هم من الشعراء المجيدين وقصائدهم تعد من عيون الشعر العربي		
11	ما هي الأوركسترا	الأوركسترا هي عبارة عن مجموعة من عازفي الأدوات الموسيقية وغالبا ما تكون مجموعة كبيرة تضم ما يقارب المئة عازف ا		
12	ماذا تسمى الأوركسترا كبيرة الحجم والمكونه من مئة وعشرة عازف فأكثر	حصن أرواد تسمى ارواد باليونانية ارادوس وبالفنيقية ارفاد		
13	من يقود الأوركسترا عادة	ويقود الأوركسترا عادة شخص يسمى يسمى بالقائد أو المايسترو		
14	ما هي العمارة	العمارة هي فن و علم تشييد و تصميم المباني		
15	لماذا وجد فن العمارة	مدارس العمارة هي مدرسة التكعيبية و مدرسة الديستيل و مدرسة المستقبلية و مدرسة التعبيرية أو الوصفية		
16	ماذا يعتبر المعماري	يعتبر المعماري فنان و فيلسوف بالدرجة الأولى فهو من المفترض أن يعتمد في أي تصميم على مفاهيم و عناصر تتعلق بهدف	Total # of Qs	60
17	ما هي مدارس العمارة	مدارس العمارة هي مدرسة التكعيبية و مدرسة الديستيل و مدرسة المستقبلية و مدرسة التعبيرية أو الوصفية		
18	ما هو الفن اوالفنون	الفن أو الفنون هي نتاج إبداعي إنساني	No. Correct	44
19	ماذا تعتبر الفنون	الفسيفساء هي أحد أقدم الفنون التصويرية ويتم تشكيل اللوحة الفسيفسائية عادة من انتظام عد كبير من القطع الصغيرة وعادة ما	No. InCorrec	16
20	لماذا تعتبر الفنون لوناً من الثقافة الانسانية	لأنها تعبير عن التعبيرية الذاتية وليس تعبيرا عن حاجة الإنسان لمتطلبات حياته	No. UnAnsrd	1
21	ما هي اهم الفنون البصرية	الفسيفساء هي أحد أقدم الفنون التصويرية ويتم تشكيل اللوحة الفسيفسائية عادة من انتظام عد كبير من القطع الصغيرة وعادة ما		
22	اين نجد الفنون الغير مادية	الفنون الغير مادية نجدها في الموسيقي والرقص والدراما والكتابة للقصص ورواياتها		
23	من هم الاشوريون	الأشوريون هم قوم ساميون استوطنوا القسم الشمالي من العراق منذ الألف الثالث ق م	Accuracy	0.745762712
24	اين استوطن الاشوريون	الأشوريون هم قوم ساميون استوطنوا القسم الشمالي من العراق منذ الألف الثالث ق م		
25	من هم اشهر ملوك الاشوريون	من أشهر ملوك الأشوريون أشور ناصربال الثاني وسنحاريب وأشور بانيبال		
26	اين يقطنون الأشوريون المعاصرون	يذكر بأن الأشوريون المعاصرون يقطنون العراق وسوريا وتركيا وايران		

Figure 9. Snapshot of the combined results.

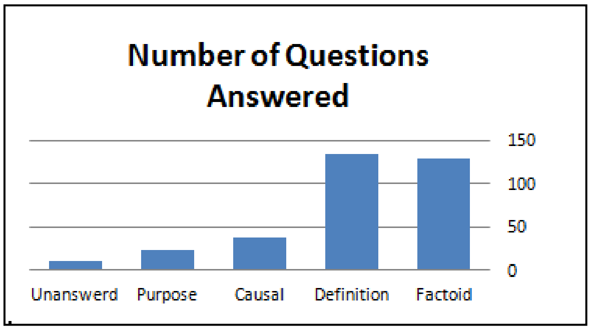


Figure 10. Number of questions answered by the QA system.

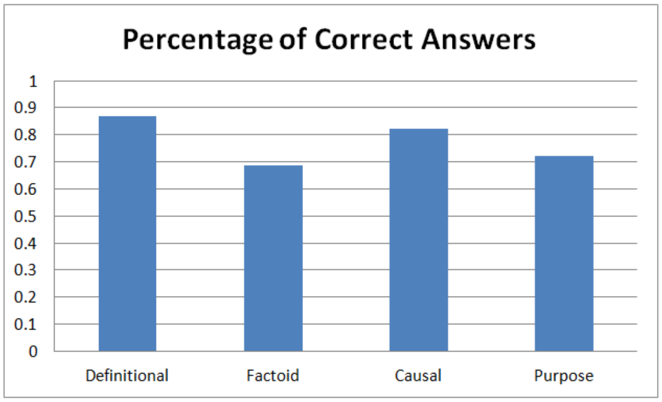


Figure 11. Percentages of correct answers among different categories.

8. SUMMARY AND CONCLUSION

In this research, a set of rules for the Analysis and Understanding of Arabic Questions in an Arabic Question Answering environment was built. To achieve the purpose of this research, different tagging rules as well as question patterns that could help in locating a more accurate answer were also built.

In comparison to the work of [14] which was used for benchmarking purposes, our approach has obtained better recall (97% vs 86%). In addition, our approach has managed to obtain better accuracy for some types of questions. For instance, our approach has obtained an accuracy of 75% for "كم" type and 72% for "لماذا" type of questions in comparison to 69% and 62% respectively in [14]. Accuracy of other types of questions was not mentioned in [14], so we could not compare our results with theirs. In our opinion, lower overall accuracy of 78% obtained by our approach, as well as low accuracy results obtained for some type of questions; especially for Factoid types (86 out of 125), can be referred to the content of the data set where many typos were found in both the text documents and the formulation of the questions. So, to obtain better results, the documents and questions need to be revised.

The scope of this part of research has concentrated on the Question Analysis and Understanding module. The IR module, however, was built using the approach used by other authors; i.e., the Relational Database approach. The Cosine similarity over the Salton VSM module was used to rank different candidate answer documents. Testing results showed an overall accuracy of 78% with a recall of 97% and an F-Measure of about 87%.

9. FUTURE RESEARCH

It can be noted that not all rules were fully constructed and implemented in this approach. For instance, rules matching a pattern like (<الموضوع> + فعل + من) are not handled in this approach. For this, we are planning to expand and extensively review and enhance all built rules to obtain better answers and performance. We are also planning to prepare and double check the collected corpora for any typos to be used for testing purposes and to make sure that our approach is performing well by manually double checking the expected answers from each of the questions. This step will be used toward applying and generalizing our approach on other collected corpora.

With some variations to the approach and its rules, work on obtaining answers from the Holy Quran [43], would constitute another part of future research. In addition, we are planning on extending the approach to deal with the sayings of the Prophet Mohammad (Peace be Upon Him).

The current research is based on syntactic analysis of words. As part of future research, we are planning on using the semantic analysis as well to help in locating proper answers.

Since this research constitutes part of an ongoing research, we are currently working on building proper response generation rules that can be used to give better answers in a dialog like context with the user.

REFERENCES

[1] A. Ezzeldin and M. Shaheen, "A Survey of Arabic Question Answering: Challenges, Tasks, Approaches, Tools and Future Trends," Proc.13th International Arab Conference on Information Technology (ACIT 2012), Paper ID 13106, Zarqa University, Jordan.

[2] C. L. Paris, "Towards More Graceful Interaction: A Survey of Question-Answering Programs," Technical Report, Columbia University, Report no. CUCS-209-85, 1985.

[3] M. R. Kangavari, S. Ghandchi and M. Golpour, "Information Retrieval: Improving Question Answering Systems by Query Reformulation and Answer Validation," Journal of World Academy of Science: Engineering & Technology, pp. 303-310, 2008.

[4] N. Kuchmann-Beauger and M. A. Aufaure, "A Natural Language Interface for Data Warehouse Question Answering," Natural Language Processing and Information Systems, pp. 201-208, 2011.

[5] D. Tufiş, "Natural Language Question Answering in Open Domains," Computer Science Journal of Moldova, vol. 19, no. 2, 2011.

[6] S. Mittal and A. Mittal, "Versatile Question Answering Systems: Seeing in Synthesis," International Journal of Intelligent Information and Database Systems, vol. 5, no. 2, pp. 119-142, 2011.

[7] S. Blair-Goldensohn, K. R. McKeown and A. H. Schlaikjer, "Defscriber: A Hybrid System for Definitional QA," Proc. 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 462-462, 2003.

[8] M. R. Kangavari, S. Ghandchi and M. Golpour, "A New Model for Question Answering Systems," Journal of World Academy of Science: Engineering & Technology, vol. 2, no. 6, pp. 506-513, 2008.

[9] A. Monroy, H. Calvo and A. Gelbukh, "NLP for Shallow Question Answering of Legal Documents Using Graphs," Computational Linguistics and Intelligent Text Processing, pp. 498-508, 2009.

[10] C. Unger and P. Cimiano, "Pythia: Compositional Meaning Construction for Ontology-based Question Answering on the Semantic Web," Natural Language Processing and Information Systems, pp. 153-160, 2011.

[11] F. A. Mohammed, K. Nasse and H. M. Harb, "A Knowledge Based Arabic Question Answering System (AQAS)," ACM SIGART Bulletin, vol. 4, no.4, pp. 21-30, 1993.

[12] B. Hammo, H. Abu-Salem and S. Lytinen, "QARAB: A Question Answering System to Support the Arabic Language," Proc. ACL-02 Workshop on Computational Approaches to Semitic Languages, pp. 1-11, 2002.

[13] Y. Benajiba, P. Rosso and A. Lyhyaoui, "Implementation of the ArabiQA Question Answering System's Components," Proc. Workshop on Arabic Natural Language Processing, 2nd Information Communication Technologies Int. Symposium, ICTIS-2007, Fez, Morroco, 3-5 April 2007.

[14] M. Akour, S. Abufardeh, K. Magel and Q.Al-Radaideh, "QArabPro: A Rule Based Question Answering System for Reading Comprehension Tests in Arabic," American Journal of Applied Sciences, vol. 8, no. 6, pp. 652-661, 2011.

[15] S. Bekhti, A. Rahman, M. Al-Harbi and T. Saba, "AQUASYS: An Arabic Question-Answering System Based on Extensive Question Analysis and Answer Relevance Scoring," International Journal of Academic Research, vol. 3, pp. 45-54, 2011.

[16] O. Trigui, L. H. Belguith, P. Rosso, H. B. Amor and B. Gafsaoui, "Arabic QA4MRE at CLEF 2012: Arabic Question Answering for Machine Reading Evaluation," Proc. CLEF (Online Working Notes/ Labs/ Workshop), 2012.

[17] N. Fareed, H. Mousa and A. Elsisi, "Enhanced Semantic Arabic Question Answering System Based on Khoja Stemmer and AWN," Proc. 9th International Conference on Computer Engineering, (ICENCO-2013), pp. 85-91, 2013.

[18] H. Kurdi, S. Alkhaider and N. Alfaifi, "Development and Evaluation of a Web Based Question Answering System for Arabic Language," International Journal on Natural Language Computing (IJNLC), vol. 3, no. 2, 2014.

[19] V. Guda, S. Sanampudi and L. Manikyamba, "Approaches for Question Answering," International Journal of Engineering Science and Technology (IJEST), vol. 3, no. 2, 2011.

[20] W. Bdour and N. Gharaibeh, "Development of Yes/No Arabic Question Answering System," International Journal of Artificial Intelligence & Applications (IJAIA), vol. 4, no. 1, 2013.

[21] O. Al-Harbi, S. Jusoh and N. Norwaw, "Handling Ambiguity Problems of Natural Language Interfaces for Question Answering," International Journal of Computer Science Issues, vol. 9, no. 3, pp. 17-25, 2012.

[22] A. Azmi and N. Al-Shenaifi, "Handling 'Why' Questions in Arabic," Proc. 5th International Conference on Arabic Language Processing, (CITALA 2014), pp. 206-209, 2014.

[23] P. Rosso, Y. Benajiba and A. Lyhyaoui, "Towards an Arabic Question Answering System," Proc. 4th Conference on Scientific Research Outlook & Technology Development in the Arab World (SROIV), Damascus, Syria, 2006.

[24] Y. Benajiba, P. Rosso and J. M. Benedíruiz, "ANERsys: An Arabic Named Entity Recognition System Based on Maximum Entropy," Computational Linguistics and Intelligent Text Processing, pp. 143-153, 2007.

[25] W. Brini, M. Ellouze, S. Mesfar and L. H. Belguith, "An Arabic Question-Answering System for Factoid Questions," Proc. International Conference on Natural Language Processing and Knowledge Engineering (NLP-KE), pp. 1-7, 2009.

[26] G. Kanaan, A. Hammouri, R. Al-Shalabi and M. Swalha, "A New Question Answering System for the Arabic Language," American Journal of Applied Sciences, vol. 6, no. 4, pp. 797-805, 2009.

[27] O. Trigui, L. H. Belguith and P. Rosso, "DefArabicQA: Arabic Definition Question Answering System," Proc. 7th Workshop on Language Resources and Human Language Technologies for Semitic Languages (LREC), Valletta, Malta, pp. 40-45, 2010.

[28] L. Abouenour, K. Bouzoubaa and P. Rosso, "IDRAAQ: New Arabic Question Answering System Based on Query Expansion and Passage Retrieval," Proc. CLEF 2012 Workshop on Question Answering for Machine Reading Evaluation (QA4MRE), 2012.

[29] A. Ezzeldin, M. Kholief and Y. El-Sonbaty, "ALQASIM: Arabic Language Question Answer Selection in Machines, Information Access Evaluation, Multilinguality, Multimodality and Visualization," Lecture Notes in Computer Science, vol. 8138, pp. 100-103, 2013.

[30] F. Al-Khawaldeh, "Answer Extraction for Why Arabic Question Answering Systems: EWAQ," Proc. World of Computer Science and Information Technology Journal (WCSIT), vol. 5, no. 5, pp. 82-86, 2015.

[31] K. Shaalan and H. Raza, "NERA: Named Entity Recognition for Arabic," Journal of the American Society for Information Science and Technology, vol. 60, no. 8, pp. 1652-1663, 2009.

[32] J. Maloney and M. Niv, "TAGARAB: A Fast, Accurate Arabic Name Recognizer Using High-precision Morphological Analysis," Proc. Workshop on Computational Approaches to Semitic Languages, pp. 8-15, 1998.

[33] K. Shaalan and H. Raza, "Person Name Entity Recognition for Arabic," Proc. 2007 Workshop on Computational Approaches to Semitic Languages: Common Issues and Resources, pp. 17-24, 2007.

[34] C. Shihadeh and G. Neumann, "ARNE: A Tool for Named Entity Recognition from Arabic Text," Proc. 4th Workshop on Computational Approaches to Arabic Script-based Languages (CAASL4), Located at the 10th Biennial Conference of the Association for Machine Translation in the Americas (AMTA), pp. 24-31, 2012.

[35] Y. Benajiba, M. Diab and P. Rosso, "Using Language Independent and Language Specific Features to Enhance Arabic Named Entity Recognition," Proc. International Arab Journal of Information Technology (IAJIT), vol. 6, no. 5, 2009.

[36] Y. Benajiba, I. Zitouni, M. Diab and P. Rosso, "Arabic Named Entity Recognition: Using Features Extracted from Noisy Data," Proc. ACL 2010 Conference Short Papers, ACLShort, Stroudsburg, PA., pp. 281–285, 2010.

[37] Y. Benajiba and P. Rosso, Arabic Question Answering, Diploma of Advanced Studies, Technical University of Valencia, Spain, 2007.

[38] K. Shaalan, "A Survey of Arabic Named Entity Recognition and Classification," Proc. Computational Linguistics, vol. 40, no. 2, 2014.

[39] Microsoft Arabic Word-Breaker, white paper, http://www.microsoft.com/en-ph/download/confirmation.aspx?id=32828, (accessed June 23rd, 2015).

[40] M. Attia, A. Toral, L. Tounsi, M. Monachini and J. Van Genabith, "An Automatically Built Named Entity Lexicon for Arabic," Proc. 7th Conference on International Language Resources and Evaluation (LREC'10), Valletta, Malta, European Language Resources Association (ELRA), May 2010.

[41] I. Gharaibeh and N. Gharaibeh, "Towards Arabic Noun Phrase Extractor (ANPE) Using Information Retrieval Techniques," Software Engineering, vol. 2, no. 2, pp. 36-42, 2012.

[42] Tenzil.net web site, Quran Corpus from http://tanzil.net/download/ (accessed on June 16th, 2015).

[43] Arabic Stop words https://sourceforge.net/projects/arabicStop words/ (accessed on July 26th, 2015).

[44] M. Al-Nabhan, An Investigation of the Impact of Stop Words Removal and Word Normalization on the Performance of Stem-based Arabic Information Retrieval, Unpublished MSc Thesis, Computer Information Systems Department, Faculty of IT, Yarmouk University, December 2015.

ملخص البحث:

يواجه البحث في معالجة اللغة العربية العديد من المشاكل التي تسببها صعوبة اللغة، وقلة الموارد المقروءة آلياً، وقلة الاهتمام من الباحثين العرب. يعد حقل السؤال والجواب أحد الحقول التي بدأت عمليات البحث الظهور فيه. وعلى الرغم من وجود بعض البحوث في هذا المجال، فإن القليل منها فقط أثبتت فعاليتها في إيجاد الإجابة الصحيحة. وتعد عمليات تحديد نوع العناصر وتحليل السؤال وفهمه من الأمور التي أدت إلى التأثير في دقة الإجابات المستخرجة. تم في هذا البحث بناء مجموعة من ٦٠+ من القواعد المناسبة لتحديد نوع العناصر ومجموعة من ١٥+ من القواعد لتحليل الأسئلة بطريقة صحيحة، و٢٠+ من قوالب الأسئلة لتحسين عملية الوصول إلى النتائج المطلوبة بشكل صحيح من وثائق معلومات تم جمعها من مصادر مختلفة. ولإثبات فعالية القواعد، تم بناء نظام للأسئلة والأجوبة وتم الوصول إلى دقة إجمالية بنسبة تبلغ حوالي ٧٨% واسترجاع بنسبة ٩٧% و مقياس (F) بنسبة تقرب من ٨٧%.

EDITORIAL BOARD SUPPORT TEAM

JJCIT ADDRESS

WEBSITE: www.jjcit.org

EMAIL: jjcit@psut.edu.jo

ADDRESS: Princess Sumaya University for Technology, Khalil Saket Street, Al-Jubaiha.

B.O. BOX: 1438 Amman 11941 Jordan.

TELEPHONE: +962-6-5359949.

FAX: +962-6-7295534.

Princess Sumaya University for Technology
Since 1991 تأسست
PSUT
جامعة الأميرة سمية للتكنولوجيا

صندوق دعم البحث العلمي
Scientific Research Support Fund

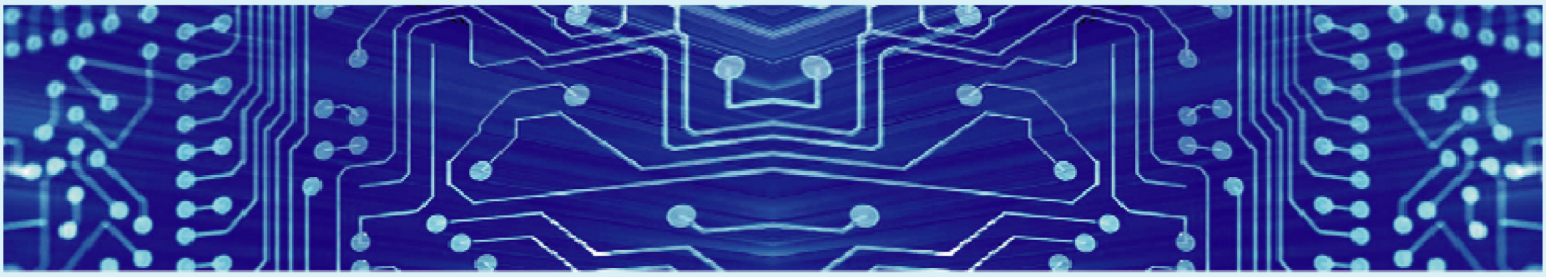
المجلة الأردنية للحاسوب و تكنولوجيا المعلومات

ISSN 2415 - 1076 (Online)
ISSN 2413 - 9351 (Print)

العدد ٣ المجلد ٢ كانون الأول ٢٠١٦

JJCIT

www.jjcit.org jjcit@psut.edu.jo



مجلة علمية عالمية متخصصة محكمة

تصدر بدعم من صندوق دعم البحث العلمي