

## Energy Optimal and High Performance Clustering for Non-Uniform, Heterogeneous Wireless Sensor Networks

C.B. Vinutha<sup>1,\*</sup>, N. Nalini<sup>2</sup>, M. Nagaraja<sup>1</sup> M. Z. Kurian<sup>1</sup>

<sup>1</sup>Dept. of Electronics and Communication, Sri Siddhartha Institute of Technology, Tumakuru-572105, India

<sup>2</sup>Dept. of Computer Science and Engineering, Nitte Meenakshi Institute of Technology, Bangalore-560064, India

Corresponding Author: C.B Vinutha

---

**Abstract:** Limiting the sensor node's energy consumption has continually been the point of concern in Wireless Sensor networks (WSN). Major portion of node's battery energy is taken away in communicating the sensed data from sensor nodes to Base Station owing, to large distances between nodes and Base Station. Recently the technique of clustering the sensor nodes into different groups have been proved to be an important factor for minimizing energy dissipation and extending the lifetime of WSNs. Selecting the most appropriate Cluster Heads(CH), in a more practical node distribution assumptions such as, non-uniform and non-circular node pattern with minimum or no node overlapping is of vital importance in improving routing efficiencies to maximize network lifetime and throughput. In the work presented here, we employ the probabilistic Expectation Maximization technique to determine the best CHs with location which are at an average mean distances from their CMs to achieve optimal communication cost, by modeling non-uniform node distribution as a Mixture of Gaussians. Besides, the presented work takes the advantage of a more relevant scenario of node heterogeneity, present in many applications to extend the lifetime of sensor network. In particular, we analyze and compare the network energy consumption, lifetime and throughput typically, for random deterministic node distribution and non-uniform or non-circular node density scenarios. Simulation results using MatLab software shows that our algorithm out performs the other three Clustering algorithms namely, LEECH, EECP and SEP in terms of energy consumption, network lifetime and throughput in both Homogeneous and heterogeneous energy conditions.

**Keywords:** Clustering; Energy Optimal; Non-Uniform; Heterogeneous; Wireless Sensor Networks.

---

Date of Submission: 26-01-2019

Date of acceptance: 09-02-2019

---

### I. Introduction

The recent advancements in the field of VLSI leading to miniaturization and low power methods have consequently led to the realization of battery operated devices called sensors which are able to detect the physical parameters from the environment and send the sensed data towards the sink [1]. Owing to size of the battery, energy capabilities of the sensor nodes are very stringent which should be carefully utilized for sensing, processing and communication mechanism [2][3]. In particular, communication process depletes the major chunk of battery energy and necessitates the design of effective network topology, efficient communication protocols, best aggregation and routing methods [4]. Recently, the techniques of clustering the sensor nodes into different groups have been proved to be an important factor for decreasing energy dissipation and extending the life time of WSNs. In each cluster a node is selected as leader named as Cluster Head (CH), while all other nodes are called Cluster members (CMs) [5]. The sensed data from CMs are forwarded to CH which in turn aggregates them and sends to the Base Station. Many research works have been proposed on clustering methods based on cluster formation criterion and CH election for both homogeneous and heterogeneous sensor networks, specifically to result in a best energy balance, scalable, and increased network lifetime. Picking the most appropriate CH, which should be energy robust and proximal to all CMs in clustering, is a significant issue. This is because the cost of communicating sensed data from sensor members to CH depends on communication range between them and has a direct impact on nodes lifetime [6].

Majority of the research works up till now assumes random deterministic node distributions in the feature space, following a circular non overlapping cluster formation. And Cluster Head selection is performed by identifying the most central, energy intensive sensor nodes which consequently will be at the mean position of all cluster members. However, the practical node deployment in the phenomenon [7] might not follow uniform, deterministic nodes distribution and the subsequent assumption of circular systematic clustering may not be correct. In the presented research work we assume more practical, spatially non- uniform node density wherein, CH selection is not as simple as determining the centroid or mean for the nodes' distribution.

In addition to giving optimal solution for CH selection in practical network conditions, we also consider heterogeneous responsive scenario with sensor nodes to prolong the time interval before the death of first node, referred to as stability period [8][9]. The behavior of random homogeneous sensor networks turn out to be very erratic and unreliable once the first node dies [9]. This is a crucial issue in many applications where response from sensor network must be reliable. As mentioned in [8][10] it is reasonable to assume heterogeneity in terms of node energy due to different issues such as,

- i. Re-energizing the sensor network by adding more nodes which certainly have more energy supply than nodes already in use.
- ii. Spatial density in some applications may be a restriction in cost wise.
- iii. Lastly, as a result of simple network operation as it evolves over a period of time, nodes could consume different chunk of energy owing to arbitrary events or radio communication channel characteristics etc. Thus, it is worth studying the performance of network lifetime which may be improved by simply considering the condition of node heterogeneity, without introducing new nodes.

### **1.1 GMM Based Cluster head Selection for Non-uniform node density**

Classical Clustering methods and the CH selection criteria are an effective solution for random, uniform or orderly distributed node pattern scenarios which can be modeled as simple Gaussian. On the other hand, Clustering of non-uniformly scattered node distribution (or instance) wherein, predominant fashion of node spread does not fall under Gaussian by its very nature, can be modeled as Mixture of Gaussian or Gaussian mixture Model (GMM) [11]. Such node distribution cannot be modeled with single Gaussian to result with mean or centroid calculations for selecting CH at optimal distances. However, we can group them into different components of separate Gaussian functions referred to as GMM. Hence, we meticulously model the cluster design for non uniform node spread based on Mixture of Gaussians and the solution to determine the best CH node with respect to the position is by using Expectation Maximization Algorithm. GMM is an effective method of Probabilistic clustering Algorithm which can be thought of as a generalized form of K-Means clustering [12]. GMM adopts soft clustering analysis wherein elements belonging to different clusters are weighted with different Probabilistic values, ranging from 0 through 1. In contrast to this, in K-means algorithm [12][13] an element or an instance either belongs to a cluster or it does not, with a probability of 0 or 1. Effective and robust procedure of estimating the parameters of probabilistic GMM models is offered by Expectation and Maximization (EM) steps [11][14][15]. For a Gaussian Mixture Model, EM algorithm results in Maximum Likelihood (ML) estimation of parameters like mean, covariance and mixing coefficients to map from many to one [11]. EM algorithm involves two basic steps: an Expectation step followed by Maximization step [15]. EM algorithm is an iterative optimization technique operated locally until condition for convergence is achieved.

- a) An Estimation step or Expectation step involves estimation of Latent variables using the current estimate of the parameters based on the observations.
- b) The Maximization step then updates the parameters of the model using ML method, based on the latent variables calculated, in expectation step.

## **II. Related Works**

There have been several research works proposed earlier with the aim of improving network lifetime and most of them have the approaches and strategies based on energy efficient network topology, communication protocols, routing protocols and best aggregation methods. In particular, there are some popular Clustering Algorithms and CH selection criteria which are briefed out and analyzed as follows;

Low Energy Adaptive Clustering Hierarchy (LEACH) [16] is one of the most important distributed dynamic and Probabilistic clustering protocols, typically applied for stationary, homogeneous Wireless Sensor Networks. In LEACH algorithm, energy balance in the network is achieved by periodically rotating the CH designation amongst all the nodes in the network. The cluster members join to a particular cluster based on least communication cost spent. Although LEACH offers a good load balancing, it has some drawback owing to probabilistic election of CH wherein, there is every chance of selecting a node with low energy as CH.

Hybrid Energy Efficient Distributed Clustering (HEED) [17], is a distributed clustering algorithm where in, selection of CHs is based on two parameters; namely residual energy and intra cluster communication cost. This algorithm results with better distributed clustering with optimal number of elected CHs and offering improved connectivity. Despite all this advantages, there is synchronization problem and not suitable for practical networks.

In [13] the authors have developed a non hierarchical clustering method using K-Medoids algorithm which focuses on optimal number of cluster formation resulting with scalable, energy efficient and balanced Wireless Sensor Networks.

EADUC [18] is an efficient distributed energy aware unequal clustering protocol which elects CHs based on ratio between average residual energy of neighbor nodes and the residual energy of the node itself.

Isolated points are eliminated in EADUC. More over small sized clusters are formed near base station to preserve energy for inter-cluster data forwarding to result with prolonged life time of the network.

ECDC [7] takes up combined issues of effective energy utilization and coverage preservation to improve QOS of networks. Two basic coverage importance metrics namely point coverage and area coverage are used to solve the coverage issues for different practical application. The algorithm results in extending network lifetime while improving network coverage with less control signal information.

EECP [19] is one of the other variants of dynamic distributed clustering protocol based on three way message exchange between every sensor and its one hop neighbors to results with clustered topologies with networked CHs whose transmission range capability is three times the transmission range of non cluster head sensors. The protocol out performs other similar protocols in terms of network life time and ratio of number of CHs to total number of sensors.

Stable election protocol (SEP) [9] considers the influence of heterogeneity of nodes' energy to prove an improved network stability. That is, the algorithm proposed in SEP results in increased network life time in which there is an enhancement of time interval before the first node's death, leading to more stable network. Here no prior assumption of initial node energy is considered and works equally good for all sized networks.

An optimized fuzzy clustering for wireless sensor networks contributes in minimizing the energy expended in data transmission between sensor node and base station by selecting powerful CH using three different fuzzy conditions [20]. The basic three parameter influencing the lifetime of wireless sensor networks are energy, centrality and node density. The proposed algorithm considers all the combination of the above three parameters and implements using fuzzy logic systems which outperforms LEACH algorithm. The author also proves that the best performance is with energy centrality based fuzzy clustering method.

### **III. Problem description**

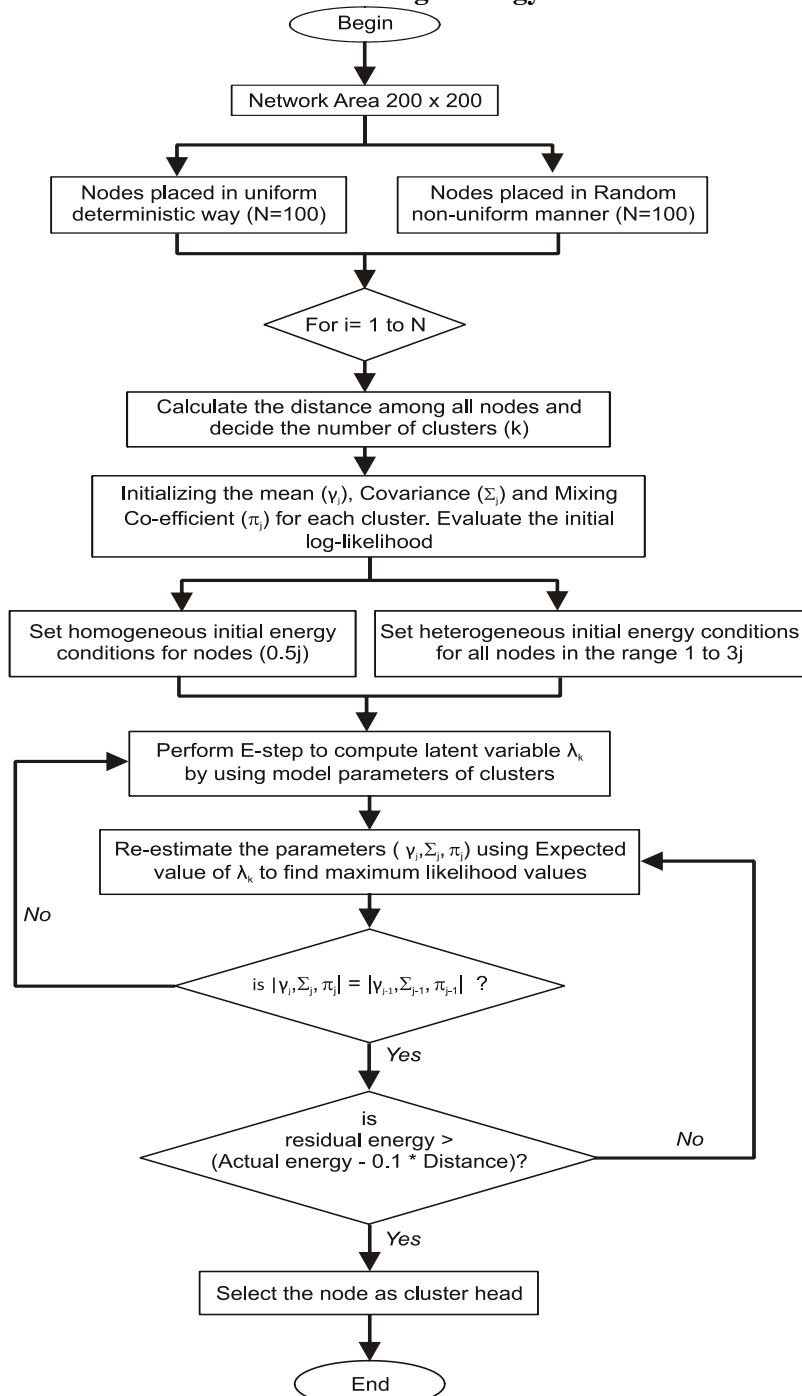
Despite wide range of research works proposed to address the energy problems at all layers of WSNs, limiting the communication cost between sensor nodes and organizing the network structures as two tier systems has highly contributed to energy saving of the network systems. Most of the previous clustering algorithms proposed, are based on the assumption that the optimal probability of a node being elected as a CH is a function of spatial density following circular shaped cluster formation, when nodes are spread uniformly over the sensor field. Such clustering mechanism is optimal in the sense that energy consumption is minimum with the constraint of uniform node density assumption, which is impractical in most of the cases.

We identify two basic flaws with such assumptions of clustering methods which are as follows:

- i) Firstly, grouping of nodes using such concepts with non uniform node scatter might result in clusters that overlap spatially leading to ambiguity in assigning cluster members to CHs appropriately.
- ii) Secondly, if the node density is non-uniform, grouping the nodes into circular clusters is not optimal which further follows that distance calculations using simple Euclidean distance between CMs and CH cannot be a valid assumption.

In other words, uniform node density modeled as single Gaussian function may not hold for non-uniform node density and CH selection will not have a closed form of solution resulting in an inefficient and unbalanced network performance. The second fold advantage of the presented work is that it considers another important practical assumption of node heterogeneity to enhance the network lifetime. By node heterogeneity we mean that all nodes in the field have heterogeneous settings of battery energy

**IV. Working Strategy**

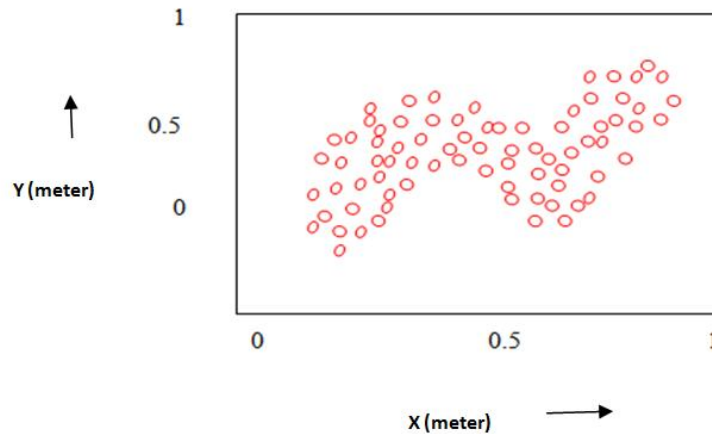


**Fig 1.** Work flow diagram of the proposed algorithm

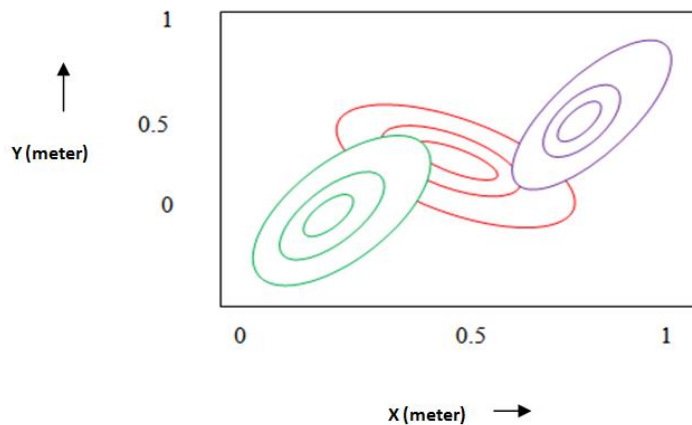
The work flow diagram of the proposed GMM clustering method for non-uniform heterogeneous sensor network with energy optimality and high performance is as shown in figure 1. Our algorithm is validated considering two cases of node distributions namely, uniform and non-uniform node deployments. Before running the algorithm cluster parameters like mean, co-variance and mixing coefficient are initialized for k-clusters. The initial node energy conditions are set as 0.5j homogeneously in one of the scenarios and are set in the range of 1 to 3j heterogeneously in other scenarios. This is followed by iterative Expectation and Maximization steps until the parameter values do not change in successive cycles. The Cluster Head selection is further performed by combining minimum node energy conditions.

**1.3.1 Non-Uniform Node Density and Gaussian Mixture Model**

Clustering the Sensor nodes in WSN is one of the effective methods in harnessing the nodes’ energy to prolong network lifetime. Efficiency of clustering method in turn depends on optimal number of clusters selection [17], Cluster head selection criteria [20] and so on, especially when practical node distribution patterns are considered. Generally, grouping of sensor nodes into different clusters for uniform node density can be characterized as simple Gaussian distribution, so that node which is closest to the mean position can be elected as CH.



**Fig 2.** Example showing nodes scattered in three distinct Hidden Groups [11]



**Fig 3.** Three different Iso contour, asymmetric Gaussian Distribution in 2D[14]

However, if the nodes are assumed to be randomly scattered as a sort of non-uniform density as shown in Figure 2, where the overall fashion of node distribution does not result with circular shaped clusters falling under single Gaussian distribution by itself, then we can cluster the nodes into three different components which insists that each of this individual components is a cluster following Gaussian distribution. As a matter of fact, such node distribution can be represented as three different Gaussian distributions as given by three different iso contours as shown in Figure 3. That is to say that, we cannot model suchlike node pattern into single Gaussian and hence, such example demands the need of multiple Gaussians or mixture of Gaussians [11].

Further, this presents certain challenges in clustering mechanism which is not looked into in classical clustering protocols. Firstly, fixing the CH at the most central positions in the feature space becomes very difficult in non circular node spread pattern. This may add to transmission cost for sensor nodes. Secondly, different clusters overlap spatially which makes the grouping of nodes to specific cluster very difficult. For example, in Figure 3 the nodes can be seen to form three groups where in, conventional clustering method fails to discover the assignment of nodes to single cluster.

Non uniform node density offer extremely difficult analysis and modeling such node pattern can, on the other hand be fitted in a more statistical fashion by assuming Mixture of Gaussians. Mixture model is probabilistically a sound way of doing soft clustering which may overlap. However, the degree of association of

different nodes to different clusters may vary. Each cluster is a generative model typically Gaussian or multinomial where, not only mean calculation is enough, but we have to compute co-variance also to characterize the energy balanced clustering model. EM algorithm permits an automatic discovery of these parameters for ‘k’ clusters where nodes will be assigned to different clusters with some soft probability. Owing to this, a generative probabilistic model is created after grouping for each of the nodes and each cluster is constituted by different Gaussian components [11][14].

**V. Expectation Maximization Algorithm**

Expectation maximization algorithm are methods which deal with multiple Gaussians as like clustering the nodes in several parts which leads to analysis based on Gaussian mixture model [11]. The univariate Gaussian distribution in one dimension is given by equation 1.

$$G(x|v, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-v)^2}{2\sigma^2}} \tag{1}$$

Where, v is the mean, σ is Standard deviation, σ<sup>2</sup> is the variance and x is the reference node with location (a, b). Multivariate Gaussian distribution can be represented by equation 2, where Σ is covariance matrix.

$$N(x|v, \Sigma) = \frac{1}{(2\pi|\Sigma|)^{1/2}} \left\{ -\frac{1}{2}(x - v)^T \Sigma^{-1}(x - v) \right\} \tag{2}$$

To extend the multivariate Gaussian distribution one of them in higher dimension to multiple of these spread over the nodes, it is necessary to estimate the parameters: v and Σ of a distribution. One of the methods to estimate v and Σ is Maximum Likelihood (ML) method.

ML method for estimating the parameters is by considering a logarithm of Gaussian distribution function, as given by equation (3)

$$\ln P(x|v, \Sigma) = -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln|\Sigma| - \frac{1}{2} (x - v)^T \Sigma^{-1} (x - v) \tag{3}$$

The mean and covariance can be determined by taking the derivative of equation 3, with respect to v and Σ and equating it to zero

$$v_{ML} = \frac{1}{N} \sum_{n=1}^N X_n \tag{4}$$

$$\Sigma_{ML} = \frac{1}{N} \sum_{n=1}^N (X_n - v_{ML})(X_n - v_{ML})^T \tag{5}$$

Where, N is number of sensor nodes.

If multiple of Gaussian exists, which is also called as linear super position set of ‘k’ number of Gaussians then, probability for the distribution for the samples ‘x’ is given by equation 6.

$$P(x) = \sum_{k=1}^K \pi_k N(x|v_k, \Sigma_k) \tag{6}$$

Where, k = Total number of Gaussians, π<sub>k</sub> = Mixing coefficients for the k<sup>th</sup> Gaussian and

N(x|v<sub>k</sub>, Σ<sub>k</sub>) = Normal multivariate Gaussian distribution for the class K

The log likelihood of equation 6 is given by equation 7:

$$\ln P(x|v, \Sigma, \pi) = \sum_{n=1}^N \ln P(x_n) = \sum_{n=1}^N \ln \left\{ \sum_{k=1}^K \pi_k N(x_n | v_k, \Sigma_k) \right\} \tag{7}$$

With Maximum likelihood method there is no closed form of solution and therefore, the parameters Σ, v are calculated using Expectation maximization technique.

**Expectation Maximization (EM) algorithm:** EM algorithm is an iterative optimization technique which is operated locally to find out the values of parameters; mean (v<sub>k</sub>), covariance (Σ<sub>k</sub>) and mixing coefficients which is a scalar quantity (π<sub>k</sub>) of ‘k’ clusters. It involves two main steps: Expectation step and Maximization step [14,15].

**Step 1:** Initialize the mean (v<sub>i</sub>), covariance (Σ<sub>i</sub>) and mixing coefficients (π<sub>i</sub>), with ‘j’ varying from 1 to k number of clusters. Evaluate the initial value of log likelihood which is given by equation (8).

$$\ln P(X|v, \Sigma, \pi) = \sum_{n=1}^N \ln P(X_n) = \sum_{n=1}^N \ln \left\{ \sum_{j=1}^K \pi_j N(X_n | v_j, \Sigma_j) \right\} \tag{8}$$

Where, N is the total number of samples or nodes, P is the probability of distribution for the samples X, N(X<sub>n</sub> | v<sub>j</sub>, Σ<sub>j</sub>) is the Normal multivariate Gaussian distribution for the class j.

Computing the overall mean of the entire sensor node distance data set and considering the individual mean of cluster to be node’s mean. Node distance is calculated by Euclidian distance formula.

**Step 2: E-Step:** Compute the expected value of the latent variable (responsibility), defined as posterior probability, using the current parameters obtained in step 1. Hence, this step is called as expectation step. Latent variable (λ<sub>k</sub>) is given by the equation (9).

$$\lambda_k(X) = \frac{\pi_k N(X|v_k, \Sigma_k)}{\sum_{j=1}^K \pi_j N(X|v_j, \Sigma_j)} \tag{9}$$

**Step 3: M Step:** Re-estimate the parameters using current responsibilities.

$$V_j = \frac{\sum_{n=1}^N \lambda_k(X_n) X_n}{\sum_{n=1}^N \lambda_j(X_n)} \quad (10)$$

$$\Sigma_j = \frac{\sum_{n=1}^N \lambda_k(X_n) (X_n - v_k) (X_n - v_k)^T}{\sum_{n=1}^N \lambda_j(X_n)} \quad (11)$$

$$\pi_j = \frac{1}{N} \sum_{n=1}^N \lambda_k(X_n) \quad (12)$$

**Step 4:** Evaluate the log likelihood. Log likelihood is estimated by the equation (13).

$$\ln P(X|v, \Sigma, \pi) = \sum_{n=1}^K \ln \left\{ \sum_{k=1}^K \pi_k N(X_n | v_k, \Sigma_k) \right\} \quad (13)$$

Select the cluster head depending on the log likelihood value and residual energy of the sensor nodes. Residual energy is calculated by (actual energy) - 0.1\*(distance covered by the nodes). Node which is nearer to the entire sensor nodes in the network and the node which have more residual energy will be selected as a cluster head. All sensor nodes which belong to a cluster are nearer to CH and hence, CH balances the load among each sensor nodes. Therefore energy consumed for the transfer of data from cluster members to CH is reduced.

Repeat this process from step 2 till the convergence criteria is reached.

Convergence criteria can be set as either of the following conditions:

1. In the successive iteration or in the last few iteration parameters ( $v, \Sigma, \pi$ ) do not change.
2. Log likelihood value change is negligible in the last few iteration.

### VI. Proposed EM Algorithm

**Begin (Expectation Maximization algorithm)**

{Initialize the mean ( $v_i$ ), co-variance ( $\Sigma_j$ ) & mixing coefficients ( $\pi_j$ )}

$i = 1$

Convergence = FALSE

While convergence == false do

$i = i+1$

{E- Step}

Compute the latent variable ( $\lambda_k$ )

$$\lambda_k(X) = \frac{\pi_k N(X|v_k, \Sigma_k)}{\sum_{j=1}^K \pi_j N(X|v_j, \Sigma_j)}$$

{M- step} ; {update the parameter based on  $\lambda_k$ }

$$V_j = \frac{\sum_{n=1}^N \lambda_k(X_n) X_n}{\sum_{n=1}^N \lambda_j(X_n)}$$

$$\Sigma_j = \frac{\sum_{n=1}^N \lambda_k(X_n) (X_n - v_k) (X_n - v_k)^T}{\sum_{n=1}^N \lambda_j(X_n)}$$

$$\pi_j = \frac{1}{N} \sum_{n=1}^N \lambda_k(X_n)$$

If  $|v_j, \Sigma_j, \pi_j, v_{j-1}, \Sigma_{j-1}, \pi_{j-1}| < \Sigma$  then

Convergence = TRUE

End if

End while

Evaluate log likelihood

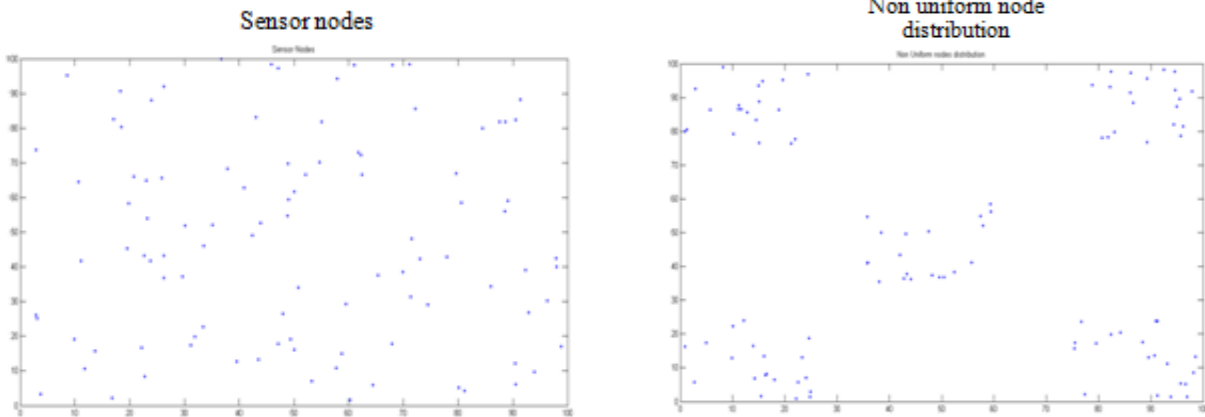
$$\ln P(X|v, \Sigma, \pi) = \sum_{n=1}^K \ln \left\{ \sum_{k=1}^K \pi_k N(X_n | v_k, \Sigma_k) \right\}$$

**Return**  $v_j, \Sigma_j, \pi_j$

**VII.Simulation Results**

**Case 1 uniform node pattern**

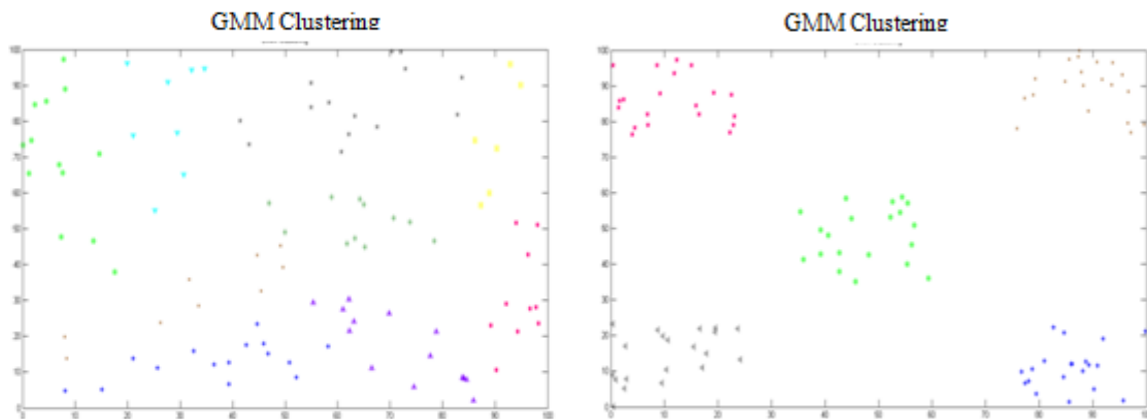
**Case 2 non uniform pattern**



**Fig 4.** Simulation Scenarios

**Case 1. GMM Clustering**

**Case 2. GMM Clustering**



**Fig 5.** GMM clustering Scenarios

In order to select best CH for more realistic and practical network structure, two types of simulation scenarios are chosen as shown in Figure 4. Scenario 1 involves a field of 200 m X 200 m in which 100 nodes are deployed in uniform manner and in the other scenario 100 nodes are non-uniformly deployed over the same field dimension and Figure 5 depicts the GMM clustering for the two cases.

**7.1 Network Model**

**Table 1:** Simulation Parameters

Network Area	200*200
Number of Nodes	100
Base Station Location	(50,175)
Optimal Election Probability of a node to become CH	0.05
Packet Length	6400
Control Packet length	200
Initial Energy	0.5 J
Transmission Energy	5.0000e-08 J
Receiving Energy	5.0000e-08 J
Free Space Energy	1.0000e-11 J
Transmit Amplifier	1.3000e-15 J
Maximum No. of Rounds	999
No. of Clusters (k)	9



The parameters of simulation are listed as in Table 1. We study our GMM CH selection algorithm with the above mentioned scenarios by setting nodes' initial energy conditions as homogeneous as well as heterogeneous.

Homogeneous Scenario of network is defined as condition where initial settings of node energy are same with each node in the network which is, 0.5J in our case. And Heterogeneous Scenario is defined as network conditions where initial settings of node energy vary uniformly over a specific range, typically 1-3J in our case.

We run our algorithm with uniform-homogeneous, uniform-heterogeneous and non uniform-heterogeneous network criteria to compare with three other different CH selection approaches namely LEACH, DEEC and SEP, to prove that our approach works efficient in all the scenarios. In specific, the proposed GMM algorithm is studied to analyze and compare the network lifetime, residual battery energy and throughput with different CH selection methods in the more practical network scenarios.

We run EM algorithm for clustering the nodes scattered in uniform and non-uniform manner as in the figure (4) and (5). Cluster Head is selected by initializing mean ( $\mu$ ), co-variance ( $\Sigma$ ) and mixing coefficient ( $\pi$ ) as follows:

- i) Initial Mean is chosen as a random number lying between 1 to N for k number of clusters, where N is the network area of 200x200 and k is number of clusters equal to 9.
- ii) Initial co-variance is selected as an overall co-variance of the number of nodes (100) as the initial variance for each k cluster.
- iii) Initial mixing coefficients or weights are assumed to be equal Prior Probabilities to each cluster (i.e 1/k)

We evaluate the basic network parameters namely network lifetime, residual energy and the throughput of our GMM clustering scheme is by limiting the algorithm iterations to 1000 rounds.

## **7.2 Performance measures**

In this subsection, we define the different measures that are used to assess the performance of our protocol.

**Network Lifetime:** Here we define the network life time as the period from start of network operation until 70% of nodes are alive.

**Residual Energy:** Net amount of sensor nodes' battery energy left after data transmission for a given communication cycle.

**Throughput:** It is defined as the total data rate over the network, i.e. the rate at which data is sent from CHs to sink and the data rate from cluster members to CH.

In Figures 6, 7 and 8 the network lifetime of our GMM clustering method are compared to study and analyze with other protocols for different communication rounds in three distinct practical scenarios namely uniform-homogeneous, uniform heterogeneous and non-uniform heterogeneous network scenarios. Next, we run GMM, SEP, LEACH and DEEC in all the three scenarios to demonstrate the energy consumption performance and the associated readings are graphically represented in Figures 9, 10 and 11. The results here reflects the number of communication cycles they can support when the average network residual battery energy are in the range between 5mj to 50mj. In general, the energy performance of GMM in different scenarios are much alike; in the sense that GMM can fit for uniformly or non-uniformly deployed networks plus for the case of homogeneous and heterogeneous network conditions. Lastly, we compare network throughput w.r.t. different cluster head selection methods in figures 12, 13 and 14 to show that our proposed work achieves significant throughput count in both node deployment scenarios and network energy conditions.

### **7.2.1 Network Lifetime and iterations**

There is no absolute definition of network lifetime of wireless sensor network and as given in [25] it can be represented as time interval when first node dies or certain percentage of nodes alive or the last node dies. Here, we define the network lifetime as the time interval when 70% of the nodes are alive. The performance metrics are studied and analyzed in three basic practical network scenarios namely uniform homogeneous, non uniform homogeneous and non uniform heterogeneous conditions. In each of the above scenarios our protocol is compared with LEACH, DEEC and SEP. Firstly, the network lifetime versus iterations is evaluated as in Figures 6, 7 and 8. Corresponding to the three network conditions here, we spot three basic regions to assess our protocol namely; time when (i) first node dies (ii) 70% of nodes are alive and (iii) all the nodes in the network dies.

As seen from Figure 6 the network life time of our proposed algorithm is stable for longer time compared to LEACH, DEEC and SEP. The detailed observations can be briefed out as follows:

The first node death with GMM algorithm occurs at 460<sup>th</sup> round as against 300<sup>th</sup> round for DEEC and 410<sup>th</sup> round for SEP and LEACH.

The second region of concern is when 70% of nodes are alive, which is also observed to be outperformed by our proposed GMM which corresponds to 580<sup>th</sup> round. On the other hand, LEACH and SEP loses their nodes much earlier as 450<sup>th</sup> round and DEEC at 380<sup>th</sup> round.

In general, at the end of 1180 rounds, all the network nodes are dead with DEEC, LEACH and SEP. However, with GMM network nodes are alive until 1280 iterations are elapsed.

Figure 7 shows the result for the case of lifetime versus iterations in the case of non uniform-homogeneous scenarios. It is very much evident that stable region of GMM is longer by maximum of 56% as compared to DEEC, SEP and LEACH respectively. This is because GMM protocol takes the advantage of electing CH with weighted probability of centrality and residual energy conditions. Also with the same reason, it follows that our proposed GMM protocol supports maximum rounds of 630 at 70% of network lifetime. The last node death in the network for existing protocols occurs at maximum of 1080 rounds as compared to our GMM at 1180.

The proposed GMM algorithm fits best with non-uniform and heterogeneous network conditions as demonstrated in Figure 8. It is very clear that stable region of GMM is increased significantly compared to DEEC. The stable region of LEACH is close to that of SEP which is 50% lesser as compared to other protocols. It is also noticeable that the proposed GMM algorithm offers measurable number of communication cycles in comparison to the other three protocols. This is because our protocol takes the advantage of electing the CH with Energy-Centrality characteristics plus energy heterogeneity of nodes, to extend the death of first node and hence improving the network stability for non uniform node spread. In other words, with node heterogeneity, nodes equipped with the same energy would die as in homogeneous case whereas nodes with higher initial energy are still alive to result with optimal CH selection and Cluster formation. This extends the network activity for more number of iterations for the proposed GMM algorithm.

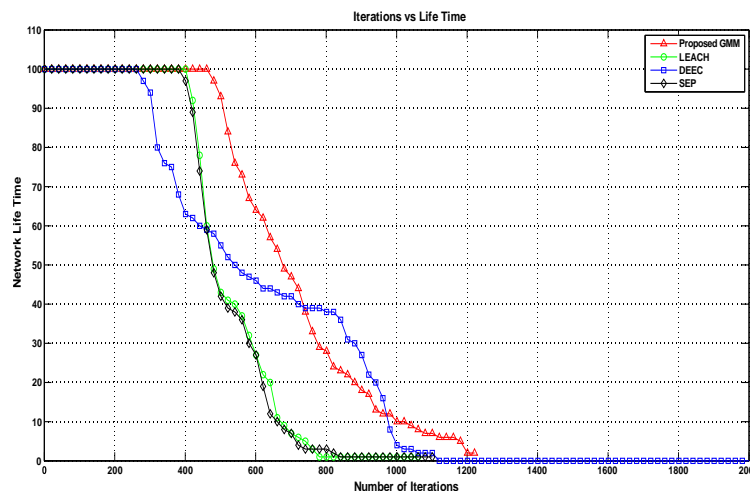


Fig 6. Iterations v/s Life time in uniform- homogeneous Scenario

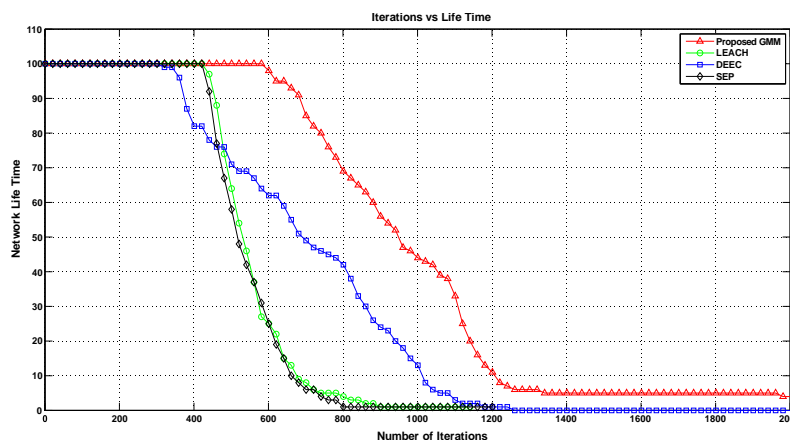


Fig 7. Iterations v/s Life time in non uniform- homogeneous Scenario

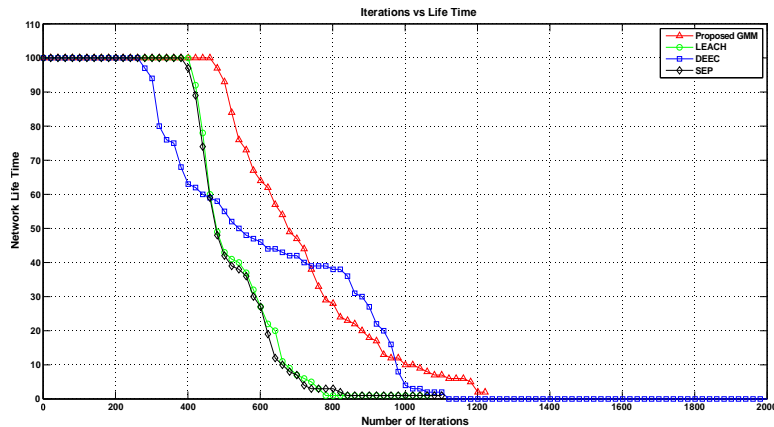


Fig 8. Iterations v/s Life time in non uniform-heterogeneous Scenario

7.2.2 Residual Energy versus Iterations

Figures 9, 10 and 11 shows the average sum of energy of sensor nodes in WSNs versus rounds, when different energy efficient clustering approaches are used in different scenarios. The graph in Figure 9 shows that network with GMM method lasts its average battery energy longer than the other three protocols, corresponding to uniform homogeneous network conditions. The remaining battery energy of the network using GMM with non uniform homogeneous conditions as in Figure 10, is almost zero at 22<sup>nd</sup> round which behaves closely same as DEEC. However in non uniform heterogeneous environment, the residual battery energy depletes after larger number of rounds as shown in Figure 11.

Besides, it can be seen that for non uniform heterogeneous environment the proposed GMM supports maximum of 120 rounds when compared to other two scenarios. This clearly indicates that well before in time, almost all the sensor nodes in the network would have spent their energy, as a result of which network stops its functions in the case of other clustering methods. Also, it can be observed by comparing the Figures 9, 10 and 11 that network energy lasts for larger number of iterations in the case of non uniform heterogeneous network conditions. The result is obtained in such a way because, CHs are located at optimal distances and hence the aggregated data transmission cost from sensor nodes to CH are essentially controlled. In addition to this, heterogeneity with node’s initial energy enables energy robust nodes to add up to accommodate larger average sum of network energy. As a consequence of which, the network functions are prolonged for more number of rounds.

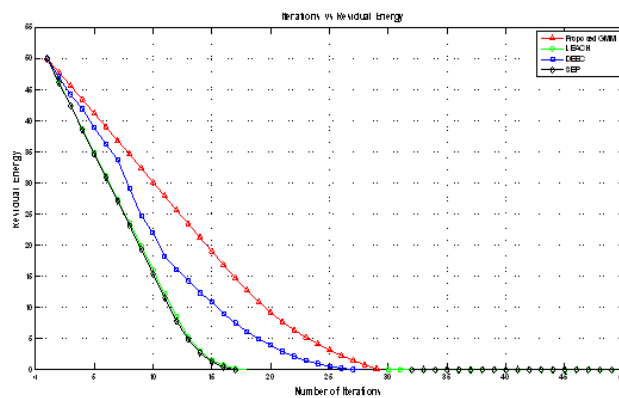


Fig 9. Iterations v/s Residual energy in uniform homogeneous scenario

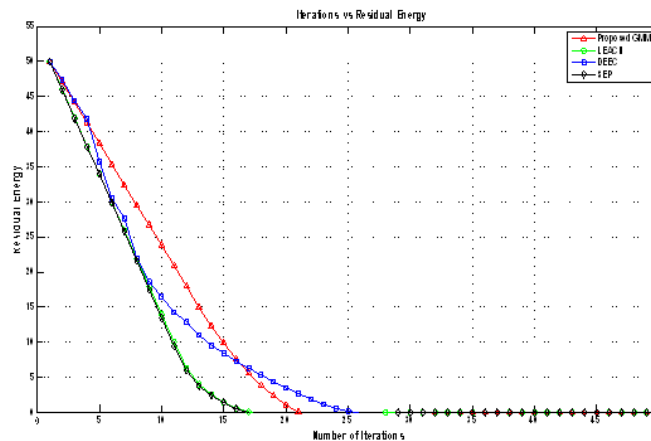


Fig 10. Iterations v/s Residual energy in non-uniform homogeneous scenario

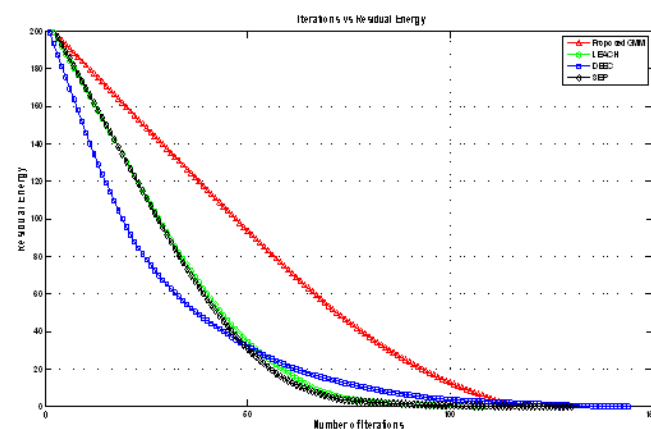


Fig 11. Iterations v/s Residual energy in non- uniform heterogeneous scenario

7.2.3 Throughput v/s Methods

We analyze the performance of different clustering methods in terms of average packets delivered to sink per unit time as shown in Figures 12, 13 and 14.

The results in Figure 13 shows that the network with GMM sends 3.5 times and 1.75 times more data compared to the network those implements DEEC and LEACH respectively, in homogeneous non uniform network conditions.

Similarly in the other two network conditions also, the proposed GMM outperforms in data transmission compared to DEEC and LEACH. This result is possible because it is apparent that GMM can minimize the energy load among the sensor nodes, which leads to maximum alive nodes over the communications cycle. This excess number of sensor nodes out turns in more data being sent to base station. The ability to gather more data in a given time is a valuable concern in the design of wireless sensor networks. Here, it is clearly demonstrated that by applying GMM algorithm for clustering provides a better throughput than using other protocols.

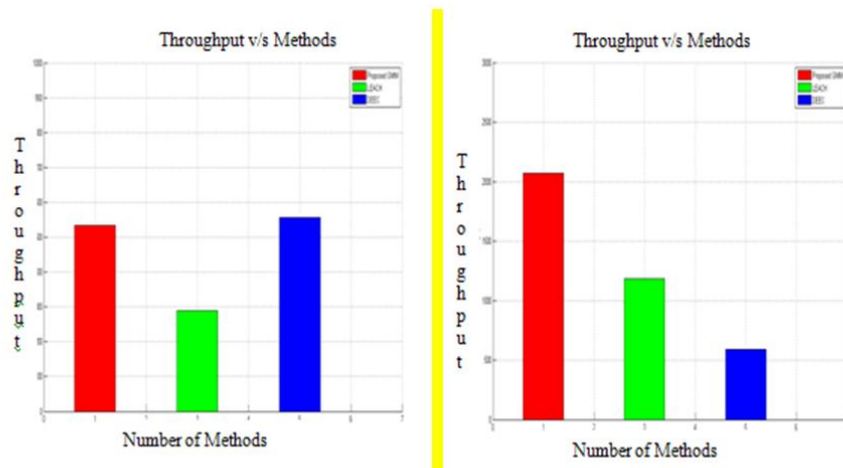


Fig12. Throughput v/s Methods in Uniform-homogeneous Scenario

Fig13. Throughput v/s Methods in Nonuniform-Homogeneous Scenario

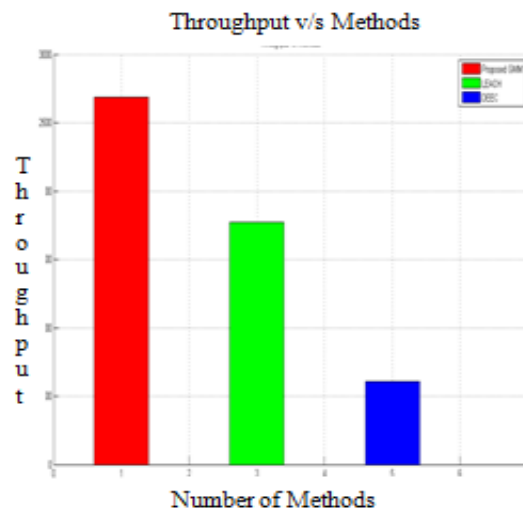


Fig14. Throughput v/s Methods in Nonuniform Heterogeneous Scenario

### VIII. Conclusion

This paper focuses on preserving the wireless sensor nodes' energy expended on data communication by designing a novel clustering method with CH selection condition. In contrast to most of the classical clustering approaches, where uniform node distribution is assumed to follow circular cluster formation, we make a tighter and practical assumption of node pattern which is non-uniform and non circular. Further, such node scatter is appropriately characterized with a more flexible class of probability distributions called as Gaussian Mixture Model. The model parameters are estimated using EM Algorithm to result in optimal selection of CHs which in turn saves the communication cost of the network. Besides, the research work extends the network lifetime by considering node heterogeneity with its initial energy settings. The combined strategy to minimize the entire network energy is effectively proved by theoretical analysis and simulation results in more realistic scenarios.

### References

- [1]. Akyildiz, Su W, Sankarasubramaniam, Cayirci E (2002) A survey on sensor networks. *IEEE Communications Magazine*, Volume: 40 Issue: 8, August 2002, pp.102-114.
- [2]. Holger Karl, Andreas Wiling Hasso (2005) Protocols and architectures for wireless sensor network. Copyright 2005 *Jhon Wiley and sons Ltd. ISBN: 0-470-09510-5.*
- [3]. Kazem Sohraby, Daniel Minoli, Taieb Znati (2006) Wireless Sensor Networks Technology, Protocols & Applications. *A Jhon Wiley & Sons Inc., Publications. ISBN:978-0-471-74300-2. Doi:10.1002/047011276X.*
- [4]. Chiara Buratti, Andrea Conti, Davide Dardari, and Roberto Verdone (2009) An Overview on Wireless Sensor Networks Technology and Evolution. *Sensors*, 6869-6896. *doi:10.3390/s90906869.*
- [5]. Ameer Ahmed Abbasi A, Mohamed Younis (2007) A survey on clustering algorithms for wireless sensor networks. *Elsevier Computer Communications* 30 2826-2841. *doi:10.1016/j.ComCom2007.05.024.*

- [6]. Devasvaran V, Latiff NM, Lik Abdul Malik NN, (2014) Energy Efficient Protocol in Wireless Sensor Networks using Mobile Base Station. *IEEE 2nd International Symposium on Telecommunication Technologies (ISTT) 978-1-4799-5982-2*.
- [7]. Xin Gu, Jiguo Yu, Dongxiao Yu, Yuhua Lv (2014) ECDC: An energy and coverage-aware distributed clustering protocol for wireless sensor networks. *Elsevier, Computer and Electrical Engineering 40:384-398*. <http://dx.doi.org/10.1016/j.compeleceng.2013.08.003>.
- [8]. Mhatre V, Rosenberg C (2004) Homogeneous vs heterogeneous clustered sensor networks: A comparative study. *Proceedings of IEEE International Conference on Communications (ICC 2004) Paris France, June 2004*.
- [9]. Georgios Smaragdakis, Ibrahim Mata, Azer Bestavros, (2004) SEP: A Stable Election Protocol for clustered heterogeneous wireless sensor networks. [www.cs.bu.edu/techreports/pdf/2004-022-sep](http://www.cs.bu.edu/techreports/pdf/2004-022-sep).
- [10]. Jiguo Yu, Li Feng, Lili Jia, Xin Gu, Dongxiao Yu (2014) A Local Energy Consumption Prediction-Based Clustering Protocol for Wireless Sensor Networks. *Sensors, 23017-23040; doi: 10.3390/s141223017*.
- [11]. [www.youtube.com](http://www.youtube.com), Mod-02 Lec-23 Gaussian Mixture Model (GMM). <http://youtu.be/ZBLyXgiBx3Q>.
- [12]. Hae-Sang Park, Jong-Seok Lee and Chi-Hyuck Jun, (2006) A K-means-like Algorithm for K-medoids Clustering and its Performance. Work supported by KOSEF System Bio-Dynamics Research center at Post Tech. <https://www.researchgate.net/publication/249891613>.
- [13]. Vinutha CB, Nalini N, (2016) Energy Aware Optimal Clustering and Reliable Routing based on Markov Model in Wireless sensor Networks. *IEEE WiSPNET 978-1-4673-9338-6/16*.
- [14]. [www.youtube.com](http://www.youtube.com), EM algorithm: how it works. [http://youtu.be/REypj2sy\\_5U](http://youtu.be/REypj2sy_5U).
- [15]. Todd K Moon, The Expectation-Maximization Algorithm. (1996) *IEEE Signal Processing Magazine 1996*, [www.eecs.yorku.ca/course\\_archive/2007-08/W/6328/Reading/EM\\_tutorial.pdf](http://www.eecs.yorku.ca/course_archive/2007-08/W/6328/Reading/EM_tutorial.pdf).
- [16]. Heinzelman WR, Chandrakasan AP, Balakrishnan H, (2000) Energy-efficient communication protocol for wireless microsensor networks. *Proceedings of the 33rd Hawaii International Conference on System Sciences (HICSS-33) , January 2000*. doi:10.1.1.720.3487.
- [17]. Younis O, Fahmy S Heed, (2004) A hybrid, energy-efficient, distributed clustering approach for ad-hoc sensor networks. *IEEE Trans. Mob. Comput. 2004, Vol-3 Issue 4, Page 366-379*. Doi: 10.1109/TMC.2004.4.1
- [18]. Jiguo Yu, Yingying Qi, Gang Wang, Qiang Guo, and Xin Gu (2011) An Energy-Aware Distributed Unequal Clustering Protocol for Wireless Sensor Networks. *Hindawi Publishing Corporation International Journal of Distributed Sensor Networks, Volume 2011, Article ID 202145, doi:10.1155/2011/202145*.
- [19]. Ali Chamam, Samuel Pierre, (2009) A distributed energy-efficient clustering protocol for wireless sensor networks. *Elsevier, Computers and Electrical Engineering 36 (2010) 303-312*. doi:10.1016/j.compeleceng.2009.03.008.
- [20]. Ashutosh Kumar Singh, Neetesh Purohit (2013) An optimised fuzzy clustering for wireless sensor Networks. *International Journal of Electronics. 101:8, 1027-1041, DOI:10.1080/00207217.2013.805387*.

IOSR Journal of Electronics and Communication Engineering (IOSR-JECE) is UGC approved Journal with SI. No. 5016, Journal no. 49082.

C.B Vinutha. "Energy Optimal and High Performance Clustering for Non-Uniform, Heterogeneous Wireless Sensor Networks." *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE) 14.1 (2019): 41-54*.