CrossMark

Energy efficient clustering algorithms for wireless sensor networks: novel chemical reaction optimization approach

P. C. Srinivasa Rao[1] · Haider Banka[1]

Abstract Clustering has been accepted as one of the most efficient techniques for conserving energy of wireless sensor networks (WSNs). However, in a two-tiered cluster based WSN, cluster heads (CHs) consume more energy due to extra overload for receiving data from their member sensor nodes, aggregating them and transmitting that data to the base station (BS). Therefore, proper selection of CHs and optimal formation of clusters play a crucial role to conserve the energy of sensor nodes for prolonging the lifetime of WSNs. In this paper, we propose an energy efficient CH selection and energy balanced cluster formation algorithms, which are based on novel chemical reaction optimization technique (nCRO), we jointly called these algorithms as novel CRO based energy efficient clustering algorithms (nCRO-ECA). These algorithms are developed with efficient schemes of molecular structure encoding and potential energy functions. For the energy efficiency, we consider various parameters such as intra-cluster distance, sink distance and residual energy of sensor nodes in the CH selection phase. In the cluster formation phase, we consider various distance and energy parameters. The algorithm is tested extensively on various scenarios of WSNs by varying number of sensor nodes and CHs. The results are compared with original CRO based algorithm, namely CRO-ECA and some existing algorithms to demonstrate the superiority of the proposed algorithm in terms of energy consumption, network lifetime, packets received by the BS and convergence rate.

1 Introduction

Clustering sensor nodes is one of the most effective techniques for conserving the energy of wireless sensor networks (WSNs) [1, 2]. In the process of clustering, the network is divided into several groups, called clusters. Each cluster has a leader referred as cluster head (CH). CHs are responsible to collect the local data from their member sensor nodes within the clusters, aggregate them and send it to the base station (BS) directly. As an example, the functionality of a cluster based WSN is shown in Fig. 1.

CH-selection [3] is an NP-hard problem as the selection of m CHs among n sensor nodes gives $^{n}c_{m}$ possibilities and also computational complexity varies exponentially when the size of a network increases. Note that, for a given n sensor nodes and m CHs, if each sensor node has an average of p CHs in the communication range, then the valid number of assignments are p^n. Therefore, the computational complexity of assignment of n sensor nodes to m CHs is also varies exponentially [4]. Brute force approaches are inefficient to solve such kind of problems.

Chemical Reaction Optimization (CRO) [5] paradigm is one of the efficient variable population based meta-heuristics inspired from chemical reactions. It emerges various engineering domains to tackle diverse hard optimization problems and shows its superiority or competitiveness over the existing popular meta-heuristics such as

✉ P. C. Srinivasa Rao
pasupuleti.vasu@yahoo.com

Haider Banka
hbanka2002@yahoo.com

[1] Department of Computer Science and Engineering, Indian School of Mines, Dhanbad 826004, India

Fig. 1 A cluster based WSN model
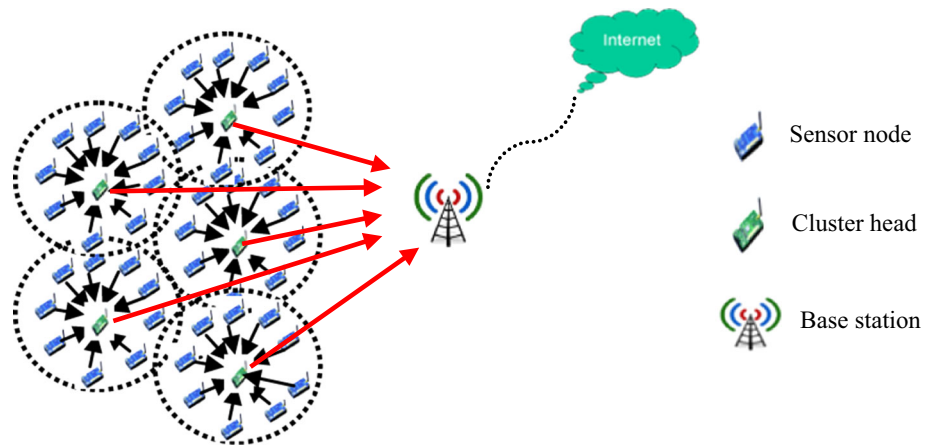
GA, ACO, PSO, etc. [6, 7]. This technique captures magnus attention and strong directions in wireless communication [8]. In WSNs, very few problems were solved using CRO, however, to the best of our knowledge, this is the first attempt to address the one of the most promising problems in WSN, i.e., clustering. Some nature inspired techniques focus on either CH selection or cluster formation phase. Also, as far as we know, no nature inspired technique was existed by considering these two essential phases of clustering as we have taken care of both the phases of clustering and propose two algorithms based on nCRO based technique. A novel chemical reaction optimization which is an improved technique of CRO by following the theories and experimental chemical kinetics like collision theory, kinetic molecular theory and Avogadro's law [9, 10], however, we assume that the collisions among the molecules in a container are inelastic. Therefore, a novel CRO can be a better choice for such NP-hard problem due to its ease of implementation, high quality of solution, ability to escape from the local optima and quick convergence.

In this paper, we deal with the CH selection and cluster formation phases with the help of novel CRO technique. We first present a Linear programming (LP) formulation for the CH selection problem and then propose an nCRO based algorithm for the same. We also present the cluster formation phase, which is also based on nCRO paradigm before that we present an LP formulation. The proposed nCRO-ECA algorithm is developed with efficient molecule structure encoding schemes. The potential energy functions are derived by considering various distance parameters and residual energy to make the nCRO-ECA based approach energy efficient. The algorithm is tested extensively to demonstrate its superiority over other existing algorithms. Therefore, our main contributions are summarized as follows:

- LP formulations for the CH selection and cluster formation problems.

- Introducing a novel mechanism by replacing the random decision making during inter or uni-molecular collisions with determined condition by following the theories, experimental chemical kinetics and Avogadro's principle for faster convergence and better quality of solution.
- nCRO based energy efficient CH selection algorithm with an efficient scheme of molecular structure encoding and a novel potential energy function.
- nCRO based energy balanced cluster formation algorithm with an efficient scheme of molecular structure encoding and a new potential energy function.
- Simulation of the proposed algorithm to demonstrate the efficiency over existing algorithms.

The rest of the paper is organized as follows. Section 2 presents the review of related works. The preliminaries of CRO, network model, energy model and terminologies are provided in Sect. 3. The LP formulations, the proposed approach for CH-selection and cluster formation algorithms which are based on nCRO paradigm are provided in Sect. 4. The simulation results are explained in Sect. 5 followed by the conclusion in Sect. 6.

2 Review of related literature

A large number of clustering algorithms have been developed for WSNs. We present the review of related work based on both brute force and nature inspired approaches. However, our main emphasis is on recent nature inspired approaches as our proposed algorithm is based on it.

2.1 Brute force approaches

A large number of clustering algorithms [11–14] based on brute force approaches have been developed for WSNs. However, LEACH [15] is a well-known distributed

clustering algorithm in which the sensor nodes elect themselves as a CH with some probability. However, the main disadvantage of this algorithm is that it may select a CH with very low energy, which can die quickly and thus degrades the performance of the network. Therefore, the more number of algorithms [16–22] have been developed to improve LEACH. PEGASIS [16] and HEED [17] are popular among them. PEGASIS organizes the nodes in the form of chains so that each node transmits and receives the data only from its neighbor nodes. However, it is unstable for large size networks. Numbers of hierarchical clustering algorithms [23–26] have been also proposed to improve the lifetime of network. In [27], the authors proposed least distance clustering (LDC) for improving the lifetime of WSN. The advantage of LDC is that it executes faster, because of assigning of non-CH nodes to nearest CH. The main drawback of LDC is the improper formation of clusters. However, as the size of the network increases the computational complexity varies exponentially for finding the optimal clusters. In [4], the authors proposed a clustering algorithm called GLBCA by considering BFS. The advantage of GLBCA is that it finds optimal solution when the sensors have equal load. The demerit of this algorithm is that the execution time is high for large scale network. Some of the algorithms [28–30] have been developed for topology control, delay tolerance and time sensitive based data collection and [31, 32] have been developed for WSNs for addressing on security issues and also application based algorithms including mobile cloud scheme for WSN were found in [33–35].

2.2 Nature inspired approaches

A plenty of clustering algorithms are also available in the literature based on nature inspired approaches [36, 37]. Some of the fault-tolerance cluster based routing algorithms are found in [38, 39]. Centralized LEACH (LEACH-C) [40] is implemented using simulated annealing. LEACH-C performs better than LEACH. Tillett et al. [41] have proposed a PSO approach to select the optimal location of CHs. However, it completely ignores the distance to the sink. Abbas et al. [42] have proposed an algorithm based on fuzzy logic and chaotic based genetic algorithm (FLCGA). However, it completely ignores the cluster formation phase. Enan et al. [43] have presented an energy-aware evolutionary routing protocol for dynamic clustering in WSNs (EAERP). However, it may select the sensor nodes as CHs which may not have sufficient energy and also may assign the non-CH sensors to nearer CHs with low energy. Guru et al. [3] have proposed a PSO based cluster formation. However, it does not consider the residual energy of the sensor nodes.

Latiff et al. [44] have proposed an energy aware cluster head selection using PSO called PSO-C by considering various parameters such as average intra-cluster distance and ratio of total initial energy of all nodes to the total current energy of the all CHs. The advantage of PSO-C is that it select the optimal CHs. The main drawback of this algorithm is that it assigns the non-CH sensor nodes to the nearest CH in the cluster formation phase, which may cause the energy inefficiency in the network and can decrease the network lifetime. But, it does not consider the sink distance, for the direct communication of CHs to BS where sink distance plays a vital role to reduce the energy consumption of the network. Buddha and Lobiyal [45] have proposed a novel energy aware CH selection algorithm using PSO. However, it completely ignores the cluster formation phase. In [46], the authors proposed a novel evolutionary approach for load balanced clustering algorithm for wireless sensor networks. The main advantage of this algorithm is that it improves the load of clusters comparably LDC and some existing algorithms. The demerit of this algorithm is that it selects the CHs randomly and also energy and distance parameters are not considered while balancing the load of clusters, which may cause energy inefficiency of the network. In [47], the authors proposed a novel differential evolution algorithm for clustering in WSNs. The advantage of this algorithm is that it efficiently extends the lifetime of first node death and then improves the lifetime of WSN. However, the authors form the clusters by considering the energy consumption and network lifetime of CHs. But, it ignores the sink distance in cluster formation and also CHs are selected randomly which may cause energy inefficiency of WSN. Some of the algorithms have been developed for targeting coverage problem [48, 49] and fuzzy based schemes for ambient intelligence applications also found in [50].

Our proposed algorithm has the following advantages over the existing algorithms:

- Compared with existing popular meta-heuristics which shows our proposed nCRO-ECA converges quickly with better quality of solution.
- It uses an efficient scheme for molecular structure encoding and potential energy function by considering various parameters, in contrast to the existing algorithms [4, 27, 46, 47] in the CH selection phase.
- It also assigns the non-CH sensor nodes to the CHs using an efficient molecular structure encoding and a new potential energy function by considering energy and various distance metrics, whereas, in the existing algorithms [4, 27, 44, 46] non-CH sensor nodes simply join the CHs by considering distance only, which may cause imbalance load of the CHs and may lead to serious energy inefficiency.

3 Preliminaries

3.1 Chemical reaction optimization

The CRO is one of the recent variable population based swarm intelligence meta-heuristics which inspired from the chemical reaction process. In CRO, a molecule, i.e., molecular structure represents a complete solution. A molecule possesses two kinds of energies, i.e., potential energy (PE) and kinetic energy (KE). The former one is characterized by the virtue of its structure, i.e., stability. Means a molecule with less PE is having more stable structure. The second one is the energy possessed by the molecule of its virtue of motion. The quality of a molecule is measured by the potential energy function. The objective function value corresponds to the PE of a molecule. In terms of mathematically the objective function can be expressed as

$$PE_\psi = f(\psi) \tag{1}$$

where f is an objective function also called as potential energy function and ψ is the structure of a molecule. For example, a molecule intends to change its structure from ψ to ψ' the change is always possible if $PE_\psi \geq PE_{\psi'}$ otherwise we allow change only when $PE_\psi + KE_\psi \geq PE_{\psi'}$. The KE of a molecule symbolizes its ability to escape from the local optimum.

The collisions among the molecules in a container are classified into two categories. The first one is uni-molecular collision and other is inter-molecular collision. Uni-molecular collisions are again classified into two types of reactions: on-wall ineffective collision and decomposition. Next, inter-molecular collisions are also classified into two types of reactions: inter-molecular ineffective collision and synthesis. We can describe each type of collision reaction in a container of molecules visually and also mathematically as follows.

3.1.1 On-wall ineffective collision

If a molecule hits the wall of a container and it simply bounces back. Some of the molecular attributes changes. The collision can be explained graphically is shown in Fig. 2.
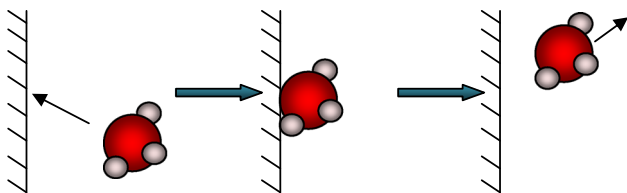
Let us assume that the current molecular structure is ψ, the obtained molecular structure is ψ' which is the neighbourhood structure of ψ. Therefore, the change is allowed only if

$$PE_\psi + KE_\psi \geq PE_{\psi'} \tag{2}$$

However, the collision is not so vigorous. The structure of the resultant molecule is not so different from the original molecule. We get, $KE_{\psi'} = (PE_\psi + KE_\psi - PE_{\psi'}) \times p, p \in [KELossRate, 1]$, where KELossRate is the system parameter bounded by [0, 1] and $(PE_\psi + KE_\psi - PE_{\psi'}) \times (1 - p)$ is the amount of energy lost to the environment when a molecule hits the wall of a container. The lost energy is stored in the central energy buffer. The stored energy can be used to encourage decomposition reaction.
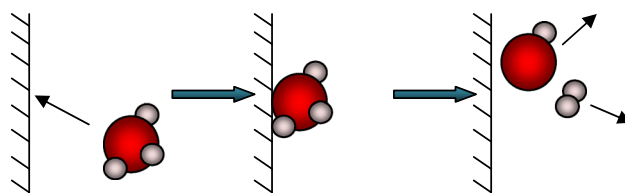
3.1.2 Decomposition

If a molecule hits the wall of a container and then splits into two or more pieces is called as decomposition. The collision is so vigorous, the structure of obtaining molecule is very different from the original one. The collision can be explained graphically is shown in Fig. 3.

The condition for the decomposition as: $(NHits[\psi] - MHits[\psi]) > \alpha$. Suppose the structure of an original molecule is ψ and the obtained structures are $\psi_{1'}$ and $\psi_{2'}$. If the original molecule ψ has sufficient energy $(PE + KE)$ to endow the PE of the resulting molecules then the change is allowed. That is as follows:

$$PE_\psi + KE_\psi \geq PE_{\psi 1'} + PE_{\psi 2'} \tag{3}$$
$$\text{Let } temp_1 = PE_\psi + KE_\psi - PE_{\psi 1'} - PE_{\psi 2'}$$

Therefore, $KE_{\psi 1'} = temp_1 \times q$ and $KE_{\psi 2'} = temp_1 \times (1 - q)$, where q is a random uniform number which is generated from the interval [0, 1]. However, Eq. (3) is unusual to happen. In general PE_ψ, $PE_{\psi 1'}$ and $PE_{\psi 2'}$ are having similar energy. Equation (3) holds only when KE_ψ is large enough. But due to sequence of on-wall ineffective collisions the KE of molecules tend to decrease. Thus, Eq. (3) is unusual to happen in normal cases. In this case, some part of energy is drawn from the central energy buffer to favour for the happening of decomposition reaction.

Fig. 2 Visualization of on-wall ineffective collision

Fig. 3 Visualization of decomposition

$$PE_\psi + KE_\psi + buffer \geq PE_{\psi 1'} + PE_{\psi 2'} \tag{4}$$

If Eq. (4) holds, the change is allowed and we can calculate

$$KE_{\psi 1'} = (temp_1 + buffer) \times q_1 \times q_2 \tag{5}$$

and

$$KE_{\psi 2'} = (temp_1 + buffer - KE_{\psi 1'}) \times q_3 \times q_4 \tag{6}$$

where q_1, q_2, q_3 and q_4 are random uniform numbers which are independently generated from the interval [0, 1]. Multiplication with the two random numbers in both Eqs. (5) and (6) is to ensure that the $KE_{\psi 1'}$ and $KE_{\psi 2'}$ are not too large. Because the energy stored in the buffer is usually large. Also the *buffer* is updated as $buffer = buffer + temp_1 - KE_{\psi 1'} - KE_{\psi 2'}$. If Eqs. (3) and (4) do not hold, the decomposition reaction fails, then the molecule retains its original molecular structure ψ, PE and KE.

3.1.3 Inter-molecular ineffective collision

If two or more molecules collide with each other and then bounce back. The effect of change is similar to that in an on-wall ineffective collision, but it involves more than one molecule, in this case two molecules are considered. No KE is drawn to the central energy buffer, in this collision there is an exchange of energy take place between the molecules. The collision can be shown graphically in Fig. 4.

Suppose the structures of original molecules are ψ_1 and ψ_2 and the obtained structures of two new molecules are $\psi_{1'}$ and $\psi_{2'}$, where $\psi_{1'}$ and $\psi_{2'}$ are neighbourhood structures of ψ_1 and ψ_2. The change is allowed only if the following conservation of energy condition holds.

$$PE_{\psi 1} + KE_{\psi 1} + PE_{\psi 2} + KE_{\psi 2} \geq PE_{\psi 1'} + PE_{\psi 2'} \tag{7}$$

Let $temp_2 = (PE_{\psi 1} + KE_{\psi 1} + PE_{\psi 2} + KE_{\psi 2} - PE_{\psi 1'} - PE_{\psi 2'})$, Therefore, $KE_{\psi 1'} = temp_2 \times r$ and $KE_{\psi 2'} = temp_2 \times (1 - r)$, where r is a random uniform number generated in the interval [0, 1]. If Eq. (7) fails, the molecules retain original structures such as ψ_1 and ψ_2.

3.1.4 Synthesis

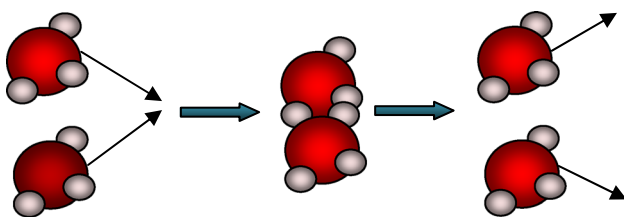If more than one molecule (here two molecules) collide with each other an intermediate molecule or new molecule

is formed. The change is so vigorous. The structure of the obtained molecule is very different from the original molecule. The collision can be explained graphically is shown in the following Fig. 5. The condition for synthesis as $KE_{\psi 1} \leq \beta$ and $KE_{\psi 2} \leq \beta$, suppose the original molecules are ψ_1 and ψ_2, the obtained molecule is ψ'. The change is allowed only if the following condition holds as:

$$PE_{\psi 1} + KE_{\psi 1} + PE_{\psi 2} + KE_{\psi 2} \geq PE_{\psi'}$$

we get, $$KE_{\psi'} = PE_{\psi 1} + KE_{\psi 1} + PE_{\psi 2} + KE_{\psi 1} - PE_{\psi'} \tag{8}$$

If Eq. (8) fails, the molecules retain the original structures. Interestingly $KE_{\psi'}$ is larger than $KE_{\psi 1}$ or $KE_{\psi 2}$ and $PE_{\psi'}$ is similar to $PE_{\psi 1}$ or $PE_{\psi 2}$. The obtained molecule is having greater ability to escape from the local minimum. In each iteration any one of the above reaction can occur. During the sequence of iterations the population of molecules may increase or decrease. The population changes only either decomposition or synthesis. If only one molecule is present in the population a uni-molecular reaction can occur.

3.2 Energy model

The energy model used in this paper is based on the same radio model as used in [15]. In this model, transmitter dissipates energy to run the radio electronics and the power amplifier. The receiver dissipates energy to run the radio electronics. The energy consumption of the node depends on the amount of the data and distance to be sent. In this model, energy consumption of a node is proportional to d^2 when the propagation distance (d) less than the threshold distance d_0 otherwise it is proportional to d^4. The total energy consumption of each node in the network for transmitting the l-bit data packet is given by the following equations.

$$E_{TX}(l, d) = \begin{cases} l \times E_{elec} + l \times \varepsilon_{fs} \times d^2, & if\ d < d_0 \\ l \times E_{elec} + l \times \varepsilon_{mp} \times d^4, & if\ d \geq d_0 \end{cases} \tag{9}$$

where, E_{elec} is the energy dissipated per bit to run the transmitter or receiver circuit, amplification energy for free space model ε_{fs} and for multi-path model ε_{mp} depends on

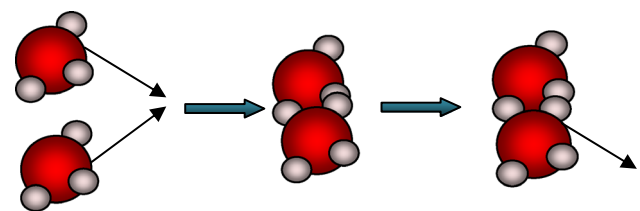**Fig. 4** Visualization of inter-molecular ineffective collision

Fig. 5 Visualization of synthesis

the transmitter amplifier model and d_0 is the threshold transmission distance. In the same way to receive l bit of data the energy consumed by the receiver is

$$E_{RX}(l) = l \times E_{elec} \qquad (10)$$

where, E_{elec} depends on several factors such as digital coding, modulation, filtering and the signal spreading.

3.3 Terminologies

For better understanding of the proposed algorithm, we tabulated the some terminologies as shown in Table 1.

The network lifetime can be defined in many ways, e.g., the time for first node dead (FND) or time elapsing from the initial deployment of nodes to the predefined percentage of nodes runs out its energy or half of nodes death (HNA) and last node death (LND) so on. However, in the proposed algorithm we consider the first node death [51].

3.4 Network model

The WSN scenario used in this paper is similar to the existing algorithms with the following properties. The sensors are randomly deployed throughout the sensing field and a node can compute the distance to the other node based on the received signal strength. Same methodology

has been adopted in the literature [52]. Therefore, it does not require any location finding system such as GPS. All the sensor nodes are assumed to be stationary after deployment and nodes are capable of operating in cluster head and normal sensor mode. Each node performs sensing periodically and has always data to send to its CH or BS. The sensing and communication distances are circular field, sensors use different levels of transmission power depending upon the distance to which data to be sent to the CH or BS; all the sensor nodes are homogeneous and have equal capability for processing and communication. The communication links are wireless, symmetric and established between the nodes when they are within the communication range of each other.

4 Proposed approach

The flow process of the proposed algorithms is shown in Fig. 6. From Fig. 6, it can be observed that after initialization of molecular structures, we introduce a novel deterministic decision making condition which is highlighted in orange colour. This condition can helpful for achieving the better quality of solution and faster convergence. We have extensively studied the theories and experimental chemical kinetics like collision theory,

Table 1 Notations used in the problem formulations and proposed algorithms

| Symbol	Description
S	The set of sensor nodes, i.e., $S = \{s_1, s_2, s_3, \ldots, s_n\}$
C	The set of CHs, i.e., $C = \{CH_1, CH_2, CH_3, \ldots, CH_m\}$, where $m < n$
p_j	Maximum no. of senors which are in the communication range of CH_j
k_{ij}	Maximum no. of CH_j which are in the communication range of senor node s_i
d_{max}	The maximum communication range of a sensor node
T_H	The threshold energy (average energy of the sensors) for being a CH
d_0	The threshold distance
E_{Si}	The initial energy of the sensor node s_i, $1 \le i \le n$
E_{CHj}	The current energy of cluster head CH_j, $1 \le j \le m$
$E_C(CH_j)$	The energy consumed by the cluster head CH_j
$E_R(CH_j)$	The residual energy of a cluster head CH_j
$dis(s_i, CH_j)$	The distance between a sensor node s_i and CH_j
$Comm(s_i)$	The set of nodes which are within the communication range of s_i, i.e.,
	$Comm(s_i) = \{S_j | \forall\ S_j \varepsilon\ S\}\ \{s_j | \forall s_j \in S \wedge dis(s_i, s_j) \le d_{\max}\}$
Pop	Population of molecules
$PopSize$	Initial population of molecules
U_{MR}	Uni-molecular reactions
$NHits$	The number of times a molecule undergoes collisions over the iterations
$MHits$	The number of collisions such that a molecule experience a minimum PE
α	The threshold value for decomposition
β	The threshold value for synthesis
T_R	The maximum number of iterations for an acceptable minimum solution (minimum PE)
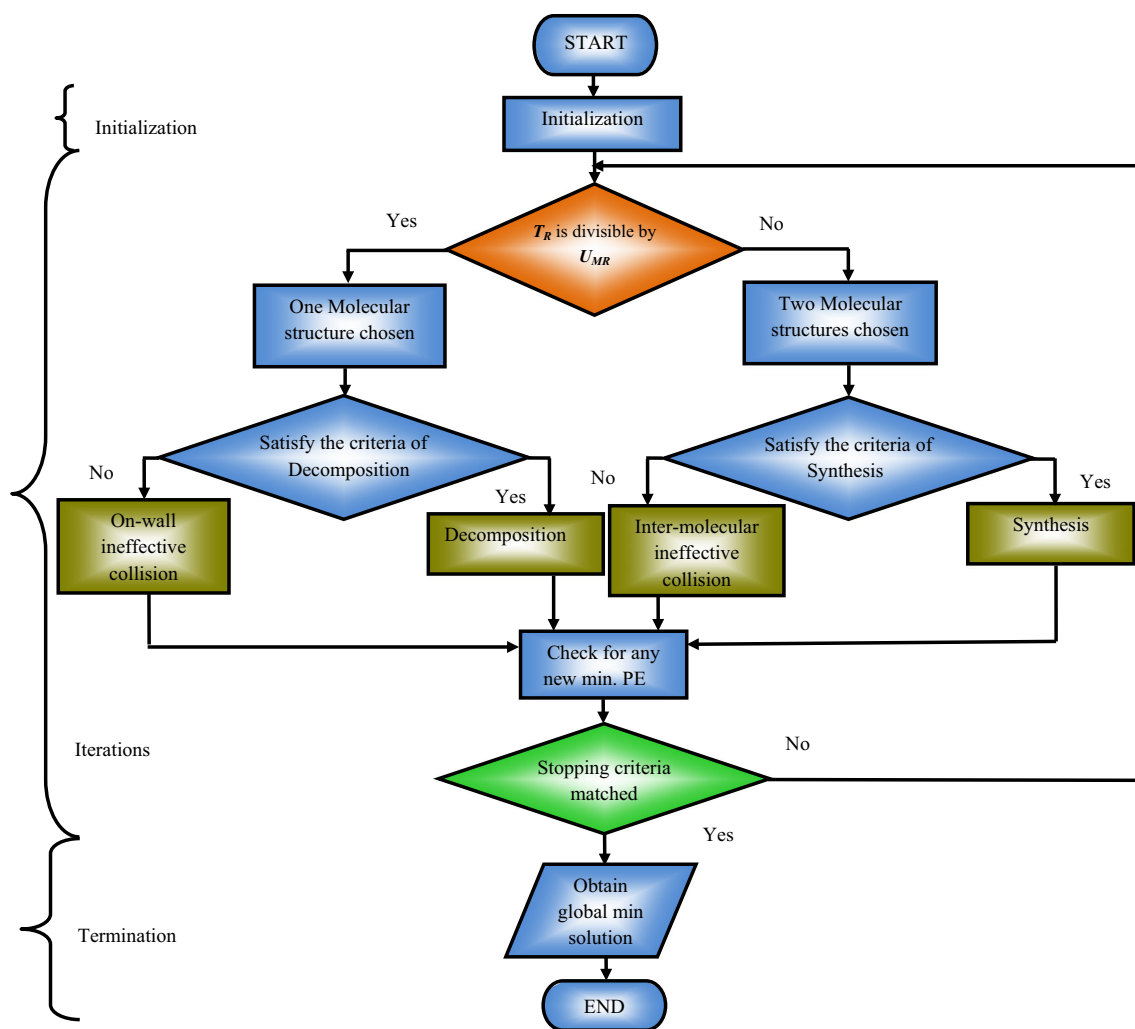
Fig. 6 Flow process of the proposed novel CRO based clustering algorithms

kinetic molecular theory and Avogadro's principle of diffusion of gases [6, 7]. The theories and experimental chemical kinetics proved that the number of uni-molecular collisions are greater than the inter-molecular collisions in a container, we follow the same natural phenomenon for proper mixture of uni and inter-molecular collisions which is an essential condition for the quick converge and better quality of solution. The value of U_{MR} is different for the two proposed algorithms. In each iteration any one of the elementary reaction can occur as shown in dark yellow coloured process blocks of the flowchart. Finally, stopping criteria in our algorithms is the number of iterations until no further change in the potential energy value is shown in the green colour process block. The decomposition and synthesis collisions make the molecules to diversify the search space, whereas the on-wall and inter-molecular ineffective collisions intensify the search space. After each collision, we check the energy conservation condition, if it satisfies we add the molecules to the *Pop* otherwise we

keep the initial molecules. For diversifying the search space we use the half-random change operator in decomposition and half-recombination for synthesis, for intensifying the search space we use the two-exchange operator.

4.1 A novel CRO based energy efficient cluster head selection algorithm

The algorithm is developed with an efficient molecular structure encoding and a new potential energy function which can efficiently select the CHs from the sensor nodes. The potential energy function is developed using various
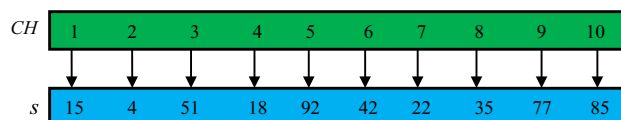

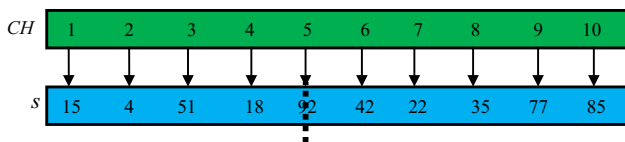**Fig. 7** Representation of a sample molecular structure

Fig. 8 Input structure of a molecule

parameters such as intra-cluster distance, sink distance and energy ratio.

4.1.1 LP formulation for CH selection

The main objective of the proposed algorithm is to select CHs among the normal sensor nodes by considering energy efficiency so that the network lifetime can be prolonged

For the efficient CH selection with energy efficiency, we consider various distance and energy parameters which include the minimum intra-cluster distance, minimum sink distance and maximum residual energy of the nodes. Let f_1 be a function of the minimum distance from their neighbor sensor nodes. We need to minimize f_1 for optimal CH selection. Let f_2 be a function of minimum sink distance and f_3 be a function of energy ratio, i.e., the ratio of the energy consumption of the sensor nodes to the remaining energy of the sensor nodes. Note that this ratio should be minimized for the optimal CH selection. We have normalized the all three objective functions between the range of 0 and 1 for efficient optimization of the function. Therefore, the Linear Programming (LP) of the optimal CH selection problem is as follows:Minimize $F = \alpha_1 \times f_1 + \alpha_2 \times f_2 + \alpha_3 \times f_3$ Subject to

$$dis(s_i, CH_j) \leq d_{max}, \forall s_i \in S, \ and \ CH_j \in C \quad (11)$$

$$E_{CH_j} > T_H, 1 \leq j \leq m \quad (12)$$

$$\alpha_1 + \alpha_2 + \alpha_3 = 1, \quad \alpha_1, \alpha_2 \ and \ \alpha_3 \in (0, 1) \quad (13)$$

The constraint (11) states that the sensor node S_i is within the maximum communication range of CH_j.The constraint (12) states that the energy of CH_j nodes must be greater than the threshold energy T_H. In the constraint (13), α_1, α_2 and α_3 are the control parameters (weights) of the function f_1, f_2 and f_3 respectively, and it also ensures that those values must not be 0 or 100 % weight.

4.1.2 Molecular structure representation and initialization

A molecular structure represents a complete solution. In this problem, the structure of a molecule represents the set of CHs have to choose from the sensor nodes. The characteristics (dimensions) of a molecular structure are the number of CHs. Let $M_i = [X_{i,1}(t), X_{i,2}(t), X_{i,3}(t), X_{i,4}(t),...,X_{i,D}(t)]$ be the ith molecule of the population of molecules, where each component $X_{i,d}(t)$ represents the

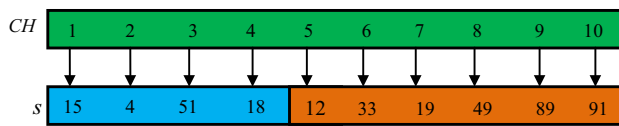node ID of CH, $1 \leq i \leq N_M$, $1 \leq d \leq D$. Then the structure of a molecule can be represented as follows.

$$M_i = [X_{i,1}(t), X_{i,2}(t), X_{i,3}(t), X_{i,4}(t),...,X_{i,D}(t)] \quad (14)$$

We illustrate it by means of a figurative example as shown in Fig. 7 in which CH denotes the index of cluster heads and s indicates the index of sensor nodes. We initialize the each component by a randomly generated sensor node ID. For example, let us assume that 100 sensor nodes are deployed in the network with node IDs ranges from 1 to 100. Therefore, the value of each component $X_{i,d}$ such as, $1 \leq X_{i,d} \leq 100$. In the CH selection algorithm, we consider the uni-molecular reaction occurs. Let us assume that the decomposition reaction happens. The process of decomposition reaction can be explained in illustration Sect. 4.1.4.

4.1.3 Derivation of potential energy function

The derivation of the potential energy function depends on the following parameters:

a. *Intra-cluster distance* We need to select the CHs such that minimum distance from its neighbor nodes. In the intra-cluster communication process, sensor nodes consume some energy to communicate data to the CHs. If we minimize intra-cluster distance, energy for the intra-cluster communication also reduces.

$$\text{Objective 1 : Minimize } f_1 = \sum_{j=1}^{m} \left(\sum_{i=1}^{p_j} dis(CH_j, s_i) \right) \quad (15)$$

b. *Sink distance* It is defined as the distance between a cluster head CH_j and the base station BS, i.e., $dis(CH_j, BS)$. Sink distance plays an important role for the direct communication of CHs to the sink. If we minimize the sink distance, energy for the communication cost of CHs to the sink also reduces.
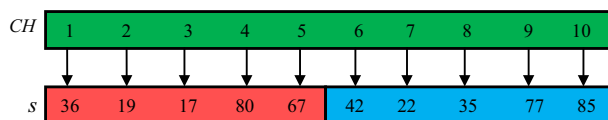

Fig. 9 Generated structure of a molecule 1

Fig. 10 Generated structure of a molecule 2

Fig. 11 Pseudo code of the proposed CH selection algorithm

Algorithm 1: A novel CRO based energy efficient CH selection

Input:	(1) Set of sensor nodes: $S = \{s_1, s_2, s_3, \ldots., s_n\}$.
	(2) Predefined *PopSize* of Molecules: N_M.
	(3) Number of characteristics of a Molecule: $D = m$.
Output:	Optimal cluster heads $CH = \{CH_1, CH_2, CH_3, \ldots., CH_m\}$.

Step 1: Assign parameter values to *PopSize*, *KELossRate* and *InitialKE*.
Step 2: *for* each of the molecules *do*
    2.1 Assign random molecular structures such as $M_i = [X_{i,1}(t), X_{i,2}(t), \ldots\ldots, X_{i,D}(t)]$, $1 \le i \le N_M$,
        $1 \le X_{i,d}(t) \le n$, $1 \le d \le D = m$, number of CHs are supposed to be selected.
    2.2 Calculate *PE* using Eq. (18)
    2.3 Assign *InitialKE= KE*
  endfor
Step 3: Let the central energy buffer be *buffer* and assign *buffer* =0
Step 4: *While* t =1 to TR *do* /*TR = Max. number of iterations*/
  if (T_R% U_{MR} ==0)
    Select a molecule, say ψ from *Pop* randomly
    if (NHits[ψ]-MHits[ψ])>α) *then*
      Decompose (ψ, *buffer*) into ψ'_1 and ψ'_2
      if satisfies the energy conservation condition either Eq. (3) or (4) *then*
        Remove ψ from *Pop*
        Add ψ'_1 and ψ'_2 to *Pop*
      endif
    else
      On-wall ineffective collision (ψ, *buffer*) into ψ'
      Add *KELost* to *buffer*
      if satisfies the energy conservation condition from Eq. (2) *then*
        Remove ψ from *Pop*
        Add ψ' to *Pop*
      endif
  else
    Select two random molecules say ψ_1 and ψ_2 from *Pop*
    if (KE $\psi_1 \le \beta$ and KE $\psi_2 \le \beta$) *then*
      Synthesis (ψ_1, ψ_2) into ψ'
      if satisfies the energy conservation condition from Eq. (8) *then*
        Remove ψ_1, ψ_2 from *Pop*
        Add ψ' to *Pop*
      endif
    else
      Inter-molecular ineffective collision (ψ_1, ψ_2) into ψ_1' and ψ_2'
      if satisfies the energy conservation condition from Eq. (7) *then*
        Remove ψ_1, ψ_2 from *Pop*
        Add ψ_1' and ψ_2' to *Pop*
    endif
  4.1 Check for any new optimal value (Minimum)
  endwhile
Step 5: Stop

$$\text{Objective 2 : Minimize } f_2 = \sum_{j=1}^{m} dis(CH_j, BS) \qquad (16)$$

c. *Energy ratio* It is defined as the ratio of energy consumed by the CHs to the residual energy of the CHs. If a CH consume less energy consumption during sensing, computation and communication activities and having more residual energy has lower energy ratio. Lower the energy ratio greater is the chance to the selection of CHs.
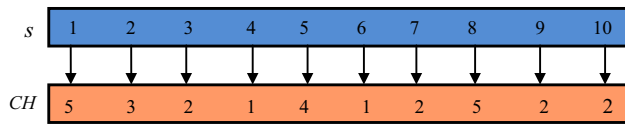
Fig. 12 Representation of a sample molecular structure

$$\text{Objective 3 : Minimize } f_3 = \sum_{j=1}^{m} \frac{E_C(CH_j)}{E_R(CH_j)} \quad (17)$$

In our nCRO approach, we use the weight summation method to minimize all the above three objective functions instead separately minimize them because all these objectives are weakly conflicting each other, so there exists a unique Pareto optimal solution. Therefore, we use the following potential energy function.

Potential energy function $= \alpha_1 \times f_1 + \alpha_2 \times f_2 + \alpha_3 \times f_3$
$$\quad (18)$$

where $\alpha_1 + \alpha_2 + \alpha_3 = 1$, Also $0 < f_1,\ f_2,\ f_3 < 1$ Our objective is to minimize the potential energy function. The lower the value of *PE*, the better is the stability of the molecule, i.e., the better is the CH selection.

4.1.4 Illustration

We can illustrate in terms of implementation point of view as follows. Let us consider the structure of a molecule for decomposition be ψ and the obtaining molecular structures after decomposition are ψ_1' and ψ_2'.

The condition for the decomposition as : $(NHits[\psi]$
$$-MHits[\psi]) > \alpha \quad (19)$$

Assume that $NHits[\psi] = 7$, $MHits[\psi] = 1$ and $\alpha = 5$. From Eq. (19), we get, $(7–1) > 5$, so the reaction favours for the decomposition. Assume that $KE_\psi = 0.1$, *Energy of buffer* $= 5$, $PE_\psi = 0.5$, $PE_{\psi1'} = 0.3$ and $PE_{\psi2'} = 0.4$

$$PE_\psi + KE_\psi \geq PE_{\psi1'} + PE_{\psi2'} \quad (20)$$

From Eq. (20), we get, $0.5 + 0.1 \geq 0.3 + 0.4$. Equation (20) unusual to happen and decomposition reaction does not occurs. It can draw some part of energy stored in the *buffer* for the happening of decomposition reaction.

$$PE_\psi + KE_\psi + buffer \geq PE_{\psi1'} + PE_{\psi2'} \quad (21)$$

From Eq. (21), we get, $0.5 + 0.1 + 5 \geq 0.3 + 0.4$. Equation (21) holds, the decomposition can occur. The molecule splits into two others molecules. Calculate the $KE_{\psi1'}$ and $KE_{\psi2'}$ as follows, assume the values of $q_1 = 0.25$, $q_2 = 0.67$, $q_3 = 0.92$ and $q_4 = 0.12$. Also $temp_1 = PE_\psi + KE_\psi - (PE_{\psi1'} + PE_{\psi2'}) = -0.1$

$$KE_{\psi1'} = (temp_1 + buffer) \times q_1 \times q_2 \quad (22)$$

From Eq. (22), we get, $KE_{\psi1'} = (-0.1 + 5) \times 0.25 \times 0.67 = (4.9) \times 0.167 = 0.82$

$$KE_{\psi2'} = (temp_1 + buffer - KE_{\psi1'}) \times q_3 \times q_4 \quad (23)$$

From Eq. (23), we also get, $KE_{\psi2'} = (-0.1 + 5 - 0.82) \times 0.92 \times 0.1 = (4.08) \times 0.092 = 0.375$.

Also, we can update the $buffer = buffer + temp_1 - (KE_{\psi1'} + KE_{\psi2'}) = 5 - 0.1 - (1.195) = 3.705$. Hence, the molecular structure ψ must be removed from the *Pop* and the nascent molecules ψ_1' and ψ_2' can be added to the *Pop*. We can also explains in terms of operators as follows.

Let us consider 10 randomly selected sensor nodes as CHs and 100 sensor nodes are randomly deployed in a region of interest. Consider a molecule from Fig. 7, i.e., an initial random representation of a molecule as an input for the decomposition. In the decomposition, we consider half-random change operator to escape from the local search and explore the search away from neighbouring region. In the half-random change operator, we select the position of slicing a molecule is exactly half of the characteristics in case of an odd number of characteristics otherwise it slices by considering the floor value. In this example the position of slicing is the 5th position, i.e., sensor node 92. Input molecule for decomposition is shown in Fig. 8 which is decomposed into two molecules are shown in Figs. 9 and 10. The structures of obtaining molecules are very different from the original molecule because the reaction is so vigorous. We check the conservation of energy condition for the obtained molecules, which favour for the decomposition, we can remove a molecule from Fig. 8 and add the molecules from the Figs. 9 and 10 to the *Pop*. Half-recombination operator is used in synthesis, means a new molecule is formed from the first half of characteristics from the first molecule and other half from the second molecule. The pseudo code of proposed cluster head selection algorithm is shown in Fig. 11.

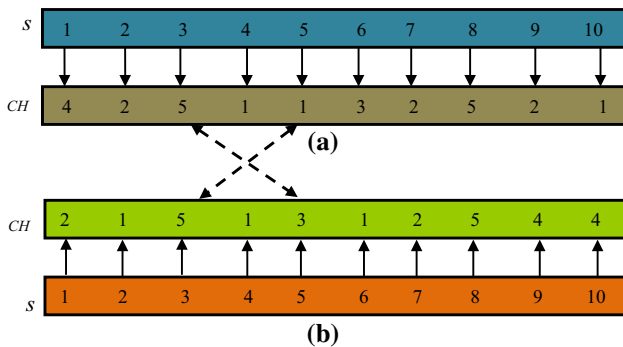## 4.2 A novel CRO based energy balanced cluster formation algorithm

After the selection of optimal CHs, the cluster formation algorithm is run at the BS for the optimal assignment of non-cluster head sensor nodes to the CHs based on the derived potential energy function.

4.2.1 LP formulation for cluster formation

Now we present the cluster formation problem, our main objective is to prolong the lifetime of network and decrease the energy consumption of the network. The network

Table 2 Sensors with the possible list of cluster heads

Sensor node (s_i)	$Comm(s_i)$
s_1	$\chi_1 = \{CH_1, CH_2, CH_3, CH_4, CH_5\}$
s_2	$\chi_2 = \{CH_1, CH_3, CH_4, CH_5\}$
s_3	$\chi_3 = \{CH_1, CH_2\}$
s_4	$\chi_4 = \{CH_1\}$
s_5	$\chi_5 = \{CH_1, CH_2, CH_4, CH_5\}$
s_6	$\chi_6 = \{CH_1, CH_5\}$
s_7	$\chi_7 = \{CH_1, CH_2, CH_3\}$
s_8	$\chi_8 = \{CH_1, CH_2, CH_3, CH_4, CH_5\}$
s_9	$\chi_9 = \{CH_2, CH_4\}$
s_{10}	$\chi_{10} = \{CH_2, CH_4, CH_5\}$

Fig. 13 a, **b** Input structures of molecules

Fig. 14 a, **b** Generated structures of molecules

$$dis(s_i, CH_j) \times a_{ij} \le d_{max}, \forall s_i \in S, \;\; and \;\; CH_j \in C \tag{25}$$

$$\sum_{j=1}^{m} a_{ij} = 1 \tag{26}$$

The constraint (25) states that the sensor node s_i can be assigned to the CH_j when it is in the maximum communication range of a sensor node s_i and (26) states that a sensor node s_i can be assigned to only one CH_j.

4.2.2 Molecular structure representation and initialization

The representation of molecular structure is different from the CH selection algorithm. For the cluster formation algorithm a structure of a molecule represents the assignment of sensor nodes to the CHs. Let $M_{1i} = [Y_{i,1}(t), Y_{i,2}(t), Y_{i,3}(t), Y_{i,4}(t),...,Y_{i,D}(t)]$ be the ith molecule from Pop, where each component $Y_{i,d}(t)$ maps the sensor nodes to the CH node IDs which are in the communication range of each sensor node, $1 \le i \le N_{M1}, 1 \le d \le D$. Then the molecular structure can be represented as follows and shown in Fig. 12. From Fig. 12, it can be observed that the index of s denotes the set of sensor nodes and CH indicates that set of CHs.

$$M_{1i} = \left[Y_{i,1}(t), Y_{i,2}(t), Y_{i,3}(t), Y_{i,4}(t), ..., Y_{i,D}(t) \right] \tag{27}$$

The generation of each characteristic of a molecule is such that to assign a sensor node to the unique cluster head within the communication range. Our proposed algorithm does not depend on any particular algorithm for initial generation of molecules. The idea for generating the initial random population can be explained as follows.

Example 1 We consider a WSN with 10 sensor nodes and 5 cluster heads, i.e., $S = \{s_1, s_2,...,s_{10}\}$ and $\chi = \{CH_1, CH_2, CH_3, CH_4, CH_5\}$.

Table 2 shows the set of sensor nodes with their possible cluster head nodes to which sensors to be assigned. Here, the number of characteristics of the molecule is equal to the
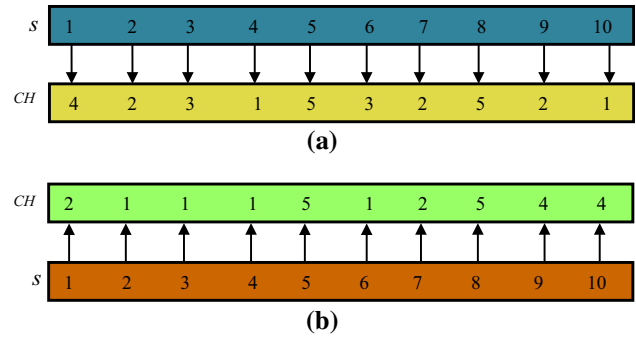
lifetime can be maximized by minimizing the energy consumption of the CH nodes, to assign the sensor nodes to the CHs in such a way that to balance the energy consumption of the CHs. And the lifetime of the network can be maximized by minimizing the distance between the sensor nodes and the CHs.

Let g_1 be a function such that to assign a sensor node to the CH with maximum residual energy. We need to maximize this for the optimal formation of clusters. Let g_2 be a function to minimize the distance from a sensor node to CH and CH to BS. We have normalized these two objective function values between the range of 0 and 1 in such a way that we minimize the linear combinations of these two functions in an efficient manner.

Let a_{ij} be a Boolean variable defined by

$$a_{ij} = \begin{cases} 1 & if \; s_i \; is \; assigned \; to \; CH_j, \forall i,j : 1 \le i \le n, 1 \le j \le m \\ 0 & otherwise \end{cases} \tag{24}$$

Minimize $W = W_1 \times g_2 + W_2 \times 1/g_1$ subject to

10. Suppose for the 3rd characteristic of a molecule, a number is generated randomly either 1 or 2, for the 5th characteristic, the randomly generated number among the 1, 2, 4 or 5. In the same way for the 10th characteristic of a molecule, a number is generated randomly amongst 2, 4 or 5. From the Fig. 12, s_3 can be assigned to the cluster head CH_2, s_5 can selects the CH_4 and so on. It should be noted

that the structure of a molecule represents the potential solution of the search space. This is due to that all sensor nodes are assigned to their corresponding cluster head nodes and also each sensor node is assigned to the unique cluster head. In the cluster formation algorithm, we consider the inter-molecular reaction occurs. The process can be explained in illustration Sect. 4.2.4.

Fig. 15 The pseudo code of proposed cluster formation algorithm

---

Algorithm 2: A novel CRO based energy balanced cluster formation

Input:
(1) Set of sensor nodes: $S = \{s_1, s_2, s_3, ..., s_n\}$.
(2) Possible sets of CHs:$\chi = \{\chi_1, \chi_2, \chi_3, ..., \chi_n\}$ such that
$\chi_i \in \{CH_d\}, \forall\ 1 \le\ i \le n, 1 \le d \le m$
(3) Predefined *PopSize*: N_{M1}.
(4) Number of characteristics of a Molecule: $D = n$.
Output: Optimal sensor assignment function $\omega: S \to \chi$

Step 1: Assign parameter values to *PopSize, KELossRate* and *InitialKE*.
Step 2: *for* each of the molecules *do*
    2.1 Assign random molecular structures as $M1i = [Yi,1(t), Yi,2(t),........,Yi,D(t)], 1 \le i \le NM1,$
       $1 \le\ d \le D = n$, number of sensors are supposed to be assigned.
    2.2 Calculate *PE* using Eq. (30)
    2.3 Assign *InitialKE= KE*
  endfor
Step 3: Let the central energy buffer be *buffer* and assign *buffer* =0
Step 4: *While* t =1 to TR *do* /*TR = Max. number of iterations*/
  if (T_R% U_{MR} ==0)
    Select a molecule, say ψ from *Pop* randomly
    if (NHits[ψ]-MHits[ψ])>α) *then*
      Decompose (ψ, *buffer*) into ψ'_1 and ψ'_2
        if satisfies the energy conservation condition either Eq. (3) or (4) *then*
          Remove ψ from *Pop*
          Add ψ'_1 and ψ'_2 to *Pop*
        endif
      else
        on-wall ineffective collision (ψ, *buffer*) into ψ'
        Add *KELost* to *buffer*
        if satisfies the energy conservation condition from Eq. (2) *then*
          Remove ψ from *Pop*
          Add ψ' to *Pop*
        endif
  else
    Select two random molecules say ψ_1 and ψ_2 from *Pop*
    if ($KE\ \psi_1 \le \beta$ and $KE\ \psi_2 \le \beta$) *then*
      Synthesis (ψ_1, ψ_2) into ψ'
      if satisfies the energy conservation condition from Eq. (8) *then*
        Remove ψ_1, ψ_2 from *Pop*
        Add ψ' to *Pop*
      endif
      else
      Inter-molecular ineffective collision (ψ_1, ψ_2) into ψ_1' and ψ_2'
      if satisfies the energy conservation condition from Eq. (7) *then*
        Remove ψ_1, ψ_2 from *Pop*
        Add ψ_1' and ψ_2' to *Pop*
    end
    4.1 Check for any new optimal value (Minimum)
  endwhile
Step 5: Stop

Table 3 Network parameters

Parameter	Value
Target area	500×500 m^2
Base station locations	(250, 250) and (500, 250)
Number of sensor nodes	200–1000
Number of cluster heads	40–90
Energy of a sensor node	2 J
E_{elec}	50 nJ/bit
ε_{fs}	10 pJ/bit/m^2
ε_{mp}	0.0013 pJ/bit/m^4
d_{max}	120 m
d_0	67 m
Packet length	4000 bits
Message size	500 bits

Table 4 CRO parameters

Parameters	Value
Number of molecules	50–60
InitialKE	0.5 J
KELossRate	0.08 J
Initial energy of buffer	0.0 J
α_1	0.22
α_2	0.20
α_3	0.58
W_1	0.67
W_2	0.33
U_{MR} (Algorithm 1)	47–78
U_{MR} (Algorithm 2)	42–70
Number of iterations	65–105

4.2.3 Derivation of potential energy function

The derivation of potential energy function depends on the following parameters.

a. *Residual energy of cluster head* The primary objective in the assignment of a sensor node to CH is the amount of energy left in the CH, i.e., residual energy of CH. The sensor node can choose a CH among possible CHs such that it has more residual energy than others.

$$\text{Objective 1 : Maximize} \quad g_1 = \sum_{i=1}^{n} \max_{j=1}^{k_{ij}} E_R(CH_j) \quad (28)$$

b. *Cluster head and sink distance* The secondary objective in the formation of cluster is the distance from the sensor node to the CH and CH to sink distance. A sensor node can be assigned to the CH based on the minimum distance from that node to CH and CH to sink node. It can be expressed as mathematically as:

Objective 2 : Minimize

$$g_2 = \sum_{i=1}^{n} \min_{j=1}^{k_{ij}} (dis(s_i, CH_j) + dis(CH_j, BS)) \quad (29)$$

We use the weight summation approach to minimize all the above two objective functions instead of separately minimize them because these two objectives are weakly conflicting each other and there exists a unique Pareto optimal solution. Therefore, we use the following potential energy function as follows.

$$\begin{aligned} Potential~energy~function &= W_1 \times g_2 + W_2 \times 1/g_1 \\ \text{Also } 0 < g_1, g_2 < 1, \quad &\text{where} \quad W_1 + W_2 = 1 \end{aligned} \quad (30)$$

Our objective is to minimize the potential energy function. The lower the value of *PE*, better is the stability of the molecule, i.e., optimal assignment of non-CH sensor nodes to CHs for the formation of balanced clusters.

4.2.4 Illustration

We can illustrate in terms of implementation point of view as follows. Assume that $KE_{\psi 1} = 1.8$, $KE_{\psi 2} = 2.1$ and $\beta = 1$. The condition for synthesis is $KE_{\psi 1} \leq \beta$ and $KE_{\psi 2} \leq \beta$. Also consider the $PE_{\psi 1} = 0.89$ and $PE_{\psi 2} = 0.56$. $PE_{\psi 1'} = 0.85$ and $PE_{\psi 2'} = 0.55$. The molecules fails the condition for synthesis, therefore an inter-molecular ineffective collision occurs. Check for the conservation of energy condition.

$$\begin{aligned} PE_{\psi 1} + KE_{\psi 1} + PE_{\psi 2} + KE_{\psi 2} &\geq PE_{\psi 1'} + PE_{\psi 2'} \\ 0.89 + 1.8 + 0.56 + 2.1 &\geq 0.85 + 0.55 \end{aligned} \quad (31)$$

Let $temp_2 = PE_{\psi 1} + KE_{\psi 1} + PE_{\psi 2} + KE_{\psi 2} - PE_{\psi 1'} - PE_{\psi 2'} = 3.95$, where $KE_{\psi 1'} = temp_2 \times r$ and $KE_{\psi 2'} = temp_2 \times (1 - r)$ and r ε [0, 1]. Therefore, $KE_{\psi 1'} = 3.95 \times 0.2 = 0.79$ and $KE_{\psi 2'} = 3.95 \times 0.8 = 3.16$.

In terms of operators point of view can be explained as follows.

Let us consider 10 sensor nodes and assumes that 5 CHs are in the communication range of each sensor. Let us pick up two molecules from the *Pop* which are shown in Fig. 13(a, b). We consider the two-exchange or pair-exchange operator to find the neighbourhood molecules for the local search. In this operation

We first pick two random positions of molecular characteristics in each input molecule and exchange the elements as follows. For example, pick the two random characteristics 5, 1 and 5, 3 from Fig. 13(a, b) respectively. Exchange these two characteristics mutually, i.e., $5 \rightarrow 3$ and $1 \rightarrow 5$. Therefore, the nascent molecules are generated as shown in Fig. 14(a, b). During on-wall ineffective collision, within a molecule two random positions are

Fig. 16 Comparison in terms of energy consumption for **a** 60 CHs, **b** 90 CHs in WSN#1
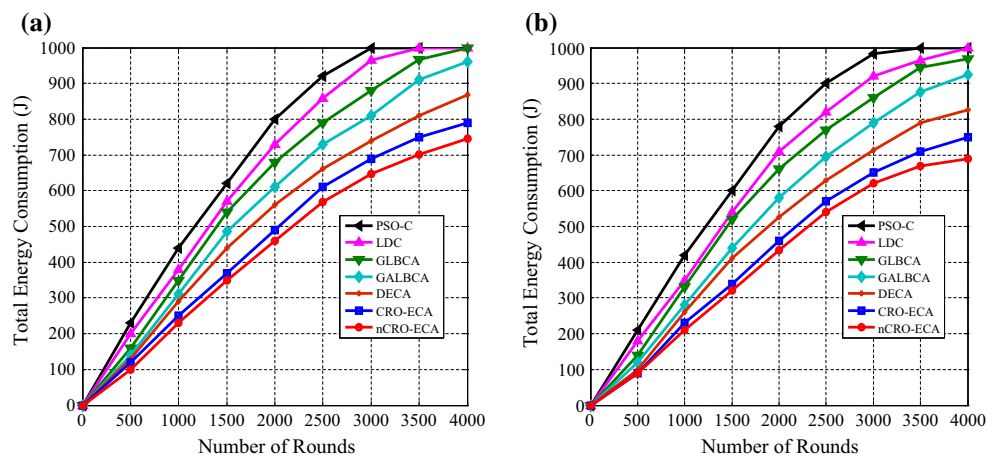
exchanged. The pseudo code of the proposed cluster formation algorithm is shown in Fig. 15.

5 Simulation results and performance evaluation

5.1 Simulation environment

The proposed algorithm was tested using C programming and the results are plotted using MATLAB (version 7.11) on an Intel core i7 processor with chipset 2600, 3.40 GHZ CPU 2 GB RAM running on the platform Microsoft Windows 7. The simulations were performed over varying number of sensor nodes from 200 to 1000 with 40 to 90 CHs. In the simulation run, we have used the parameter values as used by [15] is shown in Table 3. We have considered various network scenarios, two of which are presented here as follows. For the first scenario WSN#1 the position of base station was taken in the center of the field, i.e., at (250, 250), for the second scenario WSN#2 it was placed at (500, 250).

We tested our algorithm at variable size of the initial population of molecules ranging from 30 to 100, however, we found that the proposed CH selection algorithm was shown comparably better results at 60 molecules and cluster formation algorithm at 50 molecules. We have considered the decomposition constant (α) 4, 5 and Synthesis constant (β) 2, 3 for the CH selection algorithm and cluster formation algorithms respectively. The other parameters to run our novel CRO based algorithms are shown in Table 4.

We have run the proposed algorithms over 30 times and considered the average of these instances of data for evaluating the performance.

5.2 Performance metrics

To measure the performance of the proposed algorithms we use the following metrics.

A. Energy consumption: This is the total energy consumption for a certain number of rounds, in each round, CHs collect data, aggregate and route it to the BS. Note that energy consumption is increased as the number of rounds increases.

B. Network lifetime: As earlier stated, we consider the network lifetime as first node death (FND). More the network lifetime, performance of the network is better.

C. Packets receiving: This is the total data packets received by the BS in the whole life time of the network. More packets received better is the performance.

D. Convergence rate: It is defined as the number of iterations an algorithm takes to reach the global optimal solution.

5.2.1 Performance measurement in terms of energy consumption

We have extensively tested performance of the proposed algorithm in terms of the total energy consumption of the network over various scenarios with varying number of sensor nodes. The proposed algorithm outperforms the existing algorithms over various scenarios with varying no. of sensors and CHs. For example, we ran the proposed algorithm for comparing the total energy consumption of the network with 500 sensor nodes and 60, 90 CHs. The comparison results of the proposed algorithm with PSO-C [44], LDC [27], GLBCA [4], GALBCA [46], DECA [47] and CRO-ECA are shown in Figs. 16 and 17. It can be observed that the proposed algorithm outperforms some of

Fig. 17 Comparison in terms of energy consumption for **a** 60 CHs, **b** 90 CHs in WSN#2
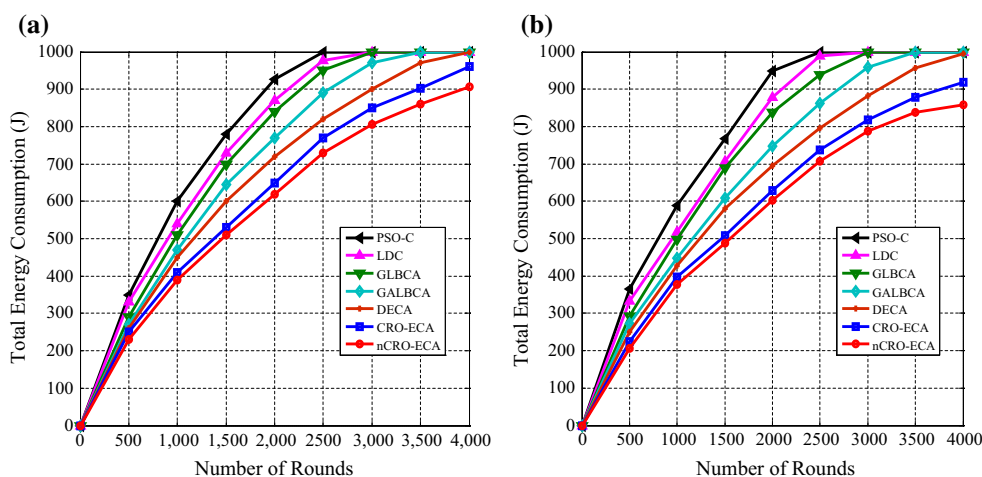
the existing algorithms in terms of total energy consumption. This is due to the fact that our derived potential energy functions taken care of energy consumption of the normal sensor nodes by reducing the distance between sensor nodes and their CHs in the cluster head selection and non-CH sensor nodes can be assigned to the CHs with less distance, it also considers the minimum distance from assigned CHs to BS in the cluster formation algorithm.

In the existing algorithms the CHs are selected randomly which may cause imbalance load of clusters. Also, there is no rotation of CHs which may cause severe energy inefficiency of the network and also in the cluster formation various essential parameters like distance and energy were not considered which may cause imbalance energy consumption.

5.2.2 Performance measurement in terms of network lifetime

Next, we ran the algorithms for comparing the network lifetime in terms of rounds by varying number of sensor nodes from 200 to 1000 and 60 to 90 CHs. From Figs. 18 and 19, it can be observed that the proposed algorithm outperforms PSO-C, LDC, GLBCA, GALBCA, DECA and CRO-ECA. The justification is that the selection of CHs in the proposed method considers the residual energy of the sensor nodes. As a CH, it actually consumes more energy than the normal sensor nodes. Therefore, the energy of CHs gets depleted more quickly than the normal sensor nodes. If a sensor node is selected with low energy, it can die quickly and hamper the network lifetime. In the proposed algorithm for the CH selection, we select the CHs from the non-CH sensor nodes with higher residual energy. Also in the cluster formation algorithm we assigned the non-CH sensor nodes to the CHs with higher residual energy.

The above Figs. 18 and 19 explain the performance of the proposed nCRO-ECA with the existing algorithms. From Figs. 18 and 19, we can observe that the proposed nCRO-ECA outperforms PSO-C, LDC, GLBCA, GALBCA, DECA and CRO-ECA in all WSN scenarios with varying number of sensor nodes. In the existing algorithms LDC, GLBCA, GALBCA and DECA the authors deployed the CHs randomly and also do not consider the energy of CHs in the assignment sensors to the CHs in the cluster formation phase. In PSO-C, the authors selected CHs in an energy efficient manner. However, normal sensor nodes are assigned to the CHs by considering distance only, does not taken care of the energy of CHs. It causes serious energy imbalance. In nCRO-ECA, the quality of the solution is obtained by the proper mixture of uni and inter-molecular collisions, which is an essential decision making condition for the quality of solution. However, in original CRO based algorithm CRO-ECA the decision making was done randomly during inter or uni-molecular collision condition, which may not yield proper mixture of collisions.

5.2.3 Performance measurement in terms of packets received

We ran the algorithms for comparing the number of data packets received by the base station by varying number of sensor nodes from 200 to 1000 and 60 to 90 CHs. From Figs. 20 and 21, it can be observed that the proposed algorithm outperforms PSO-C, LDC, GLBCA, GALBCA, DECA and CRO-ECA in terms of receipt of data packets.

The reason behind the number of packets received by the BS depend on both the energy consumption and network lifetime. The proposed algorithm receipt more number of data packets compared to existing algorithms because it consumes less energy and has more network
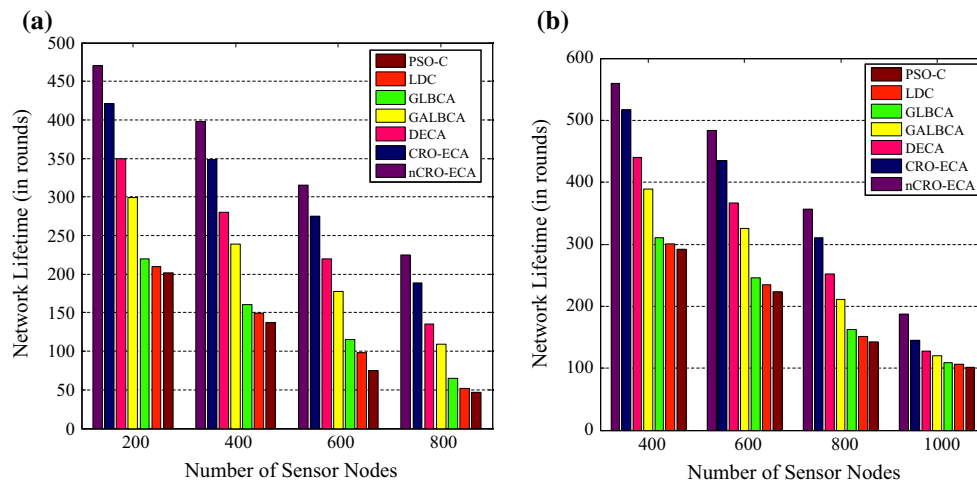
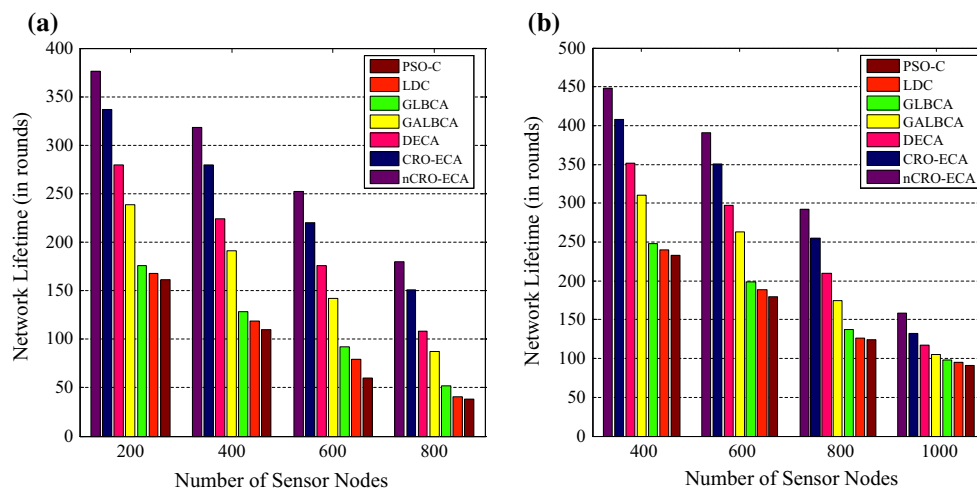**Fig. 18** Comparison in terms of network lifetime for **a** 60 CHs, **b** 90 CHs in WSN#1

Fig. 19 Comparison in terms of network lifetime for **a** 60 CHs, **b** 90 CHs in WSN#2
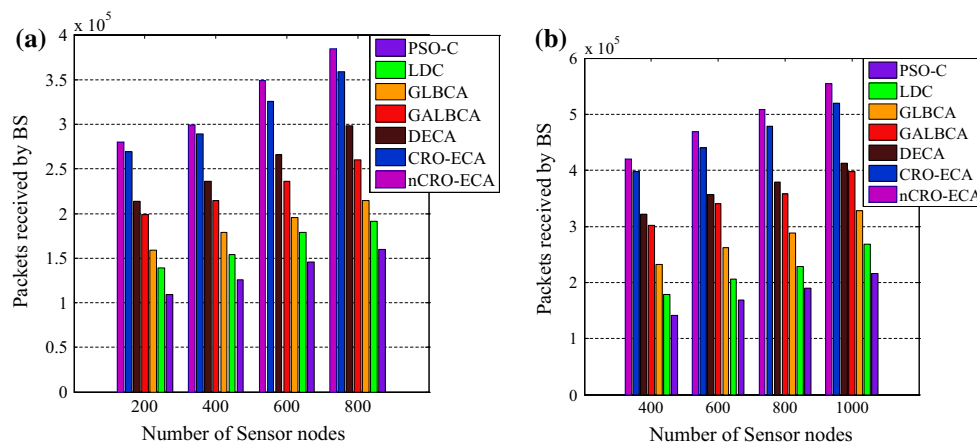

Fig. 20 Comparison in terms of packets receipt by BS for **a** 60 CHs, **b** 90 CHs in WSN#1
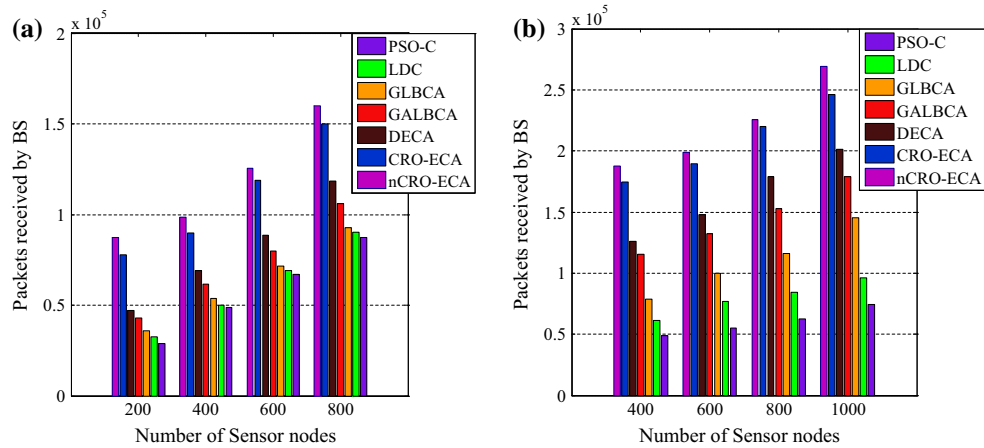
Fig. 21 Comparison in terms of packets receipt by BS for **a** 60 CHs, **b** 90 CHs in WSN#2

Fig. 22 Comparison in terms of convergence rate for **a** 40 CHs, **b** 60 CHs in WSN#1
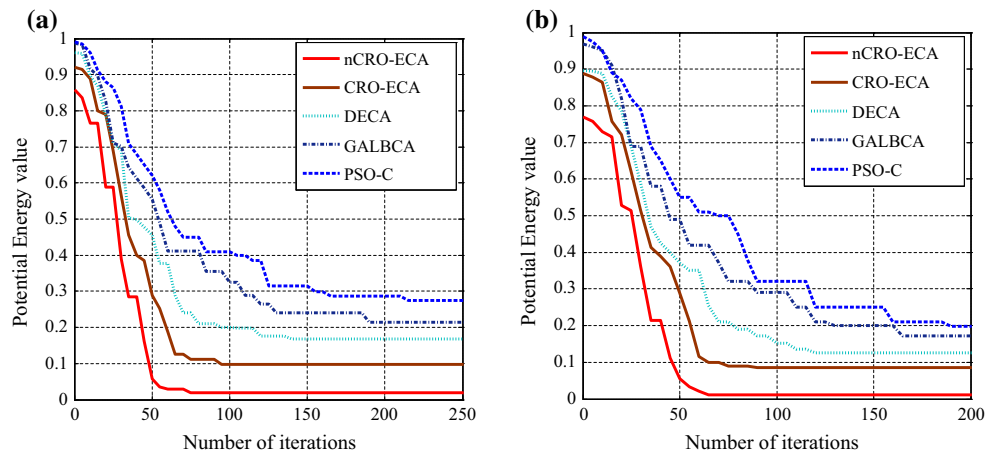
Fig. 23 Comparison in terms of convergence rate for **a** 40 CHs, **b** 60 CHs in WSN#2
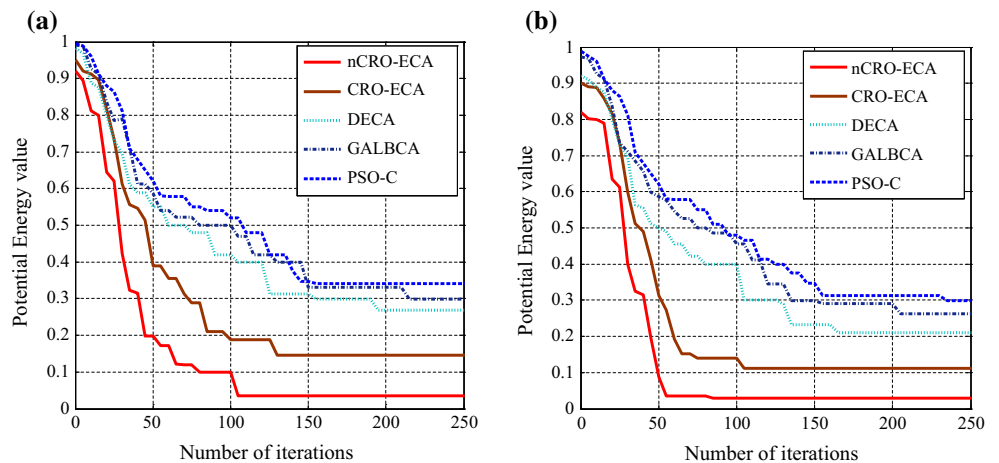
lifetime. It is also important to note that the number of packets received is predominantly more, when the base station is placed at the center of the target area compared to the edge of a target region. If the position of the base station is changed from the center to the edge of the target region, the number of packets is fallen down.

This decline is more for the existing algorithms due to the fact that the proposed algorithm takes care of proper

selection of the CHs and also formation for cluster also taken care with the help of efficient potential energy functions.

5.2.4 Performance measurement in terms of convergence rate

We have extensively tested the convergence rate of the proposed algorithm with CRO-ECA and some existing algorithms over varying number of sensors from 200 to 1000 and also varying number of CHs.

The proposed algorithm shows quick convergence and better quality of solution over the various scenarios with varying number of sensor nodes. For example, here we present the convergence rate with 500 sensors and 40–60 CHs. From Figs. 22 and 23, we can observe that the proposed algorithm nCRO-ECA converges quickly and reaches better quality of the solution compared to some existing algorithms PSO-C, GALBCA, DECA and CRO-ECA. Note that, the number of iterations and potential energy value for the convergence of nCRO-ECA is the average number of iterations and potential energy value of the two proposed algorithms. Because, the number of iterations for the convergence of the two proposed algorithms varies differently. The justification for the faster convergence with better quality of solution of nCRO-ECA is the incorporation of local improvement decision making condition by following the theoretical and experimental chemical kinetics. According to the NFL (No Free Lunch) theorem, no meta-heuristic is neither superior nor inferior to others, if a problem matches to a particular meta-heuristic it outperforms, i.e., nCRO paradigm matches to this problem. Therefore, it outperforms over the existing algorithms.

6 Conclusion

In this paper, first the Linear Programmings have been formulated for two optimization problems such as CH selection and cluster formation. Then, two algorithms have been presented for the same based on novel CRO technique using efficient molecular structure representations and potential energy functions. For the energy efficiency of the proposed nCRO-ECA, we considered various parameters such as intra-cluster distance, sink distance and residual energy of sensor nodes in the CHs selection phase. In the cluster formation phase, we considered various distance and energy parameters. We have shown simulation results along with their comparisons with various existing algorithms, namely, PSO-C, LDC, GLBCA, GALBCA, DECA and original CRO based algorithm CRO-ECA. The algorithm has been extensively tested with several scenarios

of WSN and varying number of sensors and cluster heads. The experimental results have shown that the proposed algorithm performs better than the existing algorithms in terms of total energy consumption, network lifetime, number of data packets received by the base station and convergence rate. However, we have not considered any routing algorithm in the proposed algorithm. Our future work will aim to develop a routing algorithm using the same meta-heuristic approach. For such algorithm, we shall consider various issues such as fault tolerance and hot spot problem of WSNs.

References

1. Akyildiz, I. F., Weilian, S., Sankarasubramaniam, Y., & Cayirci, E. (2002). A survey on sensor networks. *IEEE Communication Magazine, 40*(8), 102–114.
2. Abbasi, A. H., & Mohamad, Y. A. (2007). A survey on clustering algorithms for wireless sensor networks. *Computer Communications, 30*(14), 2826–2841.
3. Guru, S. M., Halgamuge, S. K., & Fernando, S. (2005). Particle swarm optimisers for cluster formation in wireless sensor networks. *In Proceedings of international conference on intelligent sensors sensor networks and information processing* (pp. 319–324).
4. Low, C. P., Fang, C., Ng, J. M., & Ang, Y. H. (2008). Efficient load-balanced clustering algorithms for wireless sensor networks. *Computer Communications, 31*(4), 750–759.
5. Lam, A. Y. S., & Li, V. (2010). Chemical-reaction-inspired metaheuristic for optimization. *IEEE Transactions on Evolutionary Computation, 14*(3), 381–399.
6. Xu, J., Lam, A., & Li, V. O. (2011). Chemical reaction optimization for task scheduling in grid computing. *IEEE Transactions on Parallel and Distributed Systems, 22*(10), 1624–1631.
7. Lam, A., Li, V. O., & Yu, J. J. (2013). Power-controlled cognitive radio spectrum allocation with chemical reaction optimization. *IEEE Transactions on Wireless Communications, 12*(7), 3180–3190.
8. Lam, A. Y., & Li, V. O. (2012). Chemical reaction optimization: A tutorial. *Memetic Computing, 4*(1), 3–17.
9. Atkins, P., & de Paula, J. (2010). *Physical chemistry* (9th ed.). Oxford, UK: Oxford University Press (Part1 and Part 3).
10. Oxtoby, D. W., Gill, H. P., & Campion, A. (2012). *Principles of modern chemistry* (7th ed.). United States of America: Cengage Learning (Unit 3 and Unit 5).
11. Afsar, M. M., & Tayarani-N, M. H. (2014). Clustering in sensor networks: A literature survey. *Journal of Network and Computer Applications, 46*, 198–226.
12. Liu, Y., Xiong, N., Zhao, Y., Vasilakos, A. V., Gao, J., & Jia, Y. (2010). Multi-layer clustering routing algorithm for wireless vehicular sensor networks. *IET Communications, 4*(7), 810–816.
13. Xiang, L., Luo, J., & Vasilakos, A. (2011). Compressed data aggregation for energy efficient wireless sensor networks. In *Sensor, mesh and ad hoc communications and networks (SECON), 2011 8th annual IEEE communications society conference on* (pp. 46–54).
14. Liu, X. A. (2012). Survey on clustering routing protocols in wireless sensor networks. *Sensors, 12*(8), 11113–11153. doi:10.3390/s120811113.

15. Heinzleman, W. B., Chandrakasan, A. P., & Balakrishnan, H. (2000). Energy efficient communication protocol for wireless microsensor networks. In *Proceedings of the 33rd Hawaii international conference on system sciences.*

16. Liu, X. Y., Zhu, Y., Kong, L., Liu, C., Gu, Y., Vasilakos, A. V., & Wu, M. Y. (2014). CDC: Compressive data collection for wireless sensor networks.

17. Hu, S., Han, J., Wei, X., & Chen, Z. (2015). A multi-hop heterogeneous cluster-based optimization algorithm for wireless sensor networks. *Wireless Networks, 21*(1), 57–65.

18. Lindsey, S., & Raghavendra, C. S. (2002). PEGASIS: Power efficient gathering in sensor information systems. *In Proceedings of IEEE aerospace conference* (Vol. 3, pp. 1125–1130).

19. Younis, O., & Fahmy, S. (2004). HEED: Hybrid energy efficient distributed clustering approach for Ad Hoc sensor networks. *IEEE Transactions on Mobile Computing, 3*(4), 366–379.

20. Yanjun, Y., Qing, C., & Vasilakos, A. V. (2015). EDAL: An energy-efficient, delay-aware, and lifetime-balancing data collection protocol for heterogeneous wireless sensor networks. *IEEE/ACM Transactions on Networking's, 23*(3), 810–823.

21. Han, K., Luo, J., Liu, Y., & Vasilakos, A. V. (2013). Algorithm design for data communications in duty-cycled wireless sensor networks: A survey. *Communications Magazine, IEEE, 51*(7), 107–113.

22. Wei, G., Ling, Y., Guo, B., Xiao, B., & Vasilakos, A. V. (2011). Prediction-based data aggregation in wireless sensor networks: Combining grey model and Kalman filter. *Computer Communications, 34*(6), 793–802.

23. Bandyopadhyay, S., & Coyle, E. J. (2003). An energy efficient hierarchical clustering algorithm for wireless sensor networks. In *IEEE INFOCOMM* (Vol. 3, pp. 1713–1723).

24. Banerjee, S., & Khuller, S. (2001). A clustering scheme for hierarchical control in wireless networks. In *Proceedings of IEEE INFOCOMM* (Vol. 2, pp. 1028–1037).

25. Xu, X., Ansari, R., Khokhar, A., & Vasilakos, A. V. (2015). Hierarchical data aggregation using compressive sensing (HDACS) in WSNs. *ACM Transactions on Sensor Networks (TOSN), 11*(3), 45.

26. Yao, Y., Cao, Q., & Vasilakos, A. V. (2013). EDAL: An energy-efficient, delay-aware, and lifetime-balancing data collection protocol for wireless sensor networks. In *Mobile ad-hoc and sensor systems (MASS), 2013 IEEE 10th international conference on* (pp. 182–190).

27. Bari, A., Jaekel, A., & Bandyopadhyay, S. (2008). Clustering strategies for improving the lifetime of two-tiered sensor networks. *Computer Communications, 31*(14), 3451–3459.

28. Rao, P. C. S., Banka, H., & Jana, P. K. (2015). PSO-based multiple-sink placement algorithm for protracting the lifetime of wireless sensor networks. In *Proceedings of the second international conference on computer and communication technologies* (pp. 605–616). Springer India.

29. Li, M., Li, Z., & Vasilakos, A. V. (2013). A survey on topology control in wireless sensor networks: Taxonomy, comparative study, and open issues. *Proceedings of the IEEE, 101*(12), 2538–2557.

30. Dvir, A., & Vasilakos, A. V. (2011). Backpressure-based routing protocol for DTNs. *ACM SIGCOMM Computer Communication Review, 41*(4), 405–406.

31. Jing, Q., Vasilakos, A. V., Wan, J., Lu, J., & Qiu, D. (2014). Security of the internet of things: Perspectives and challenges. *Wireless Networks, 20*(8), 2481–2501.

32. Yan, Z., Zhang, P., & Vasilakos, A. V. (2014). A survey on trust management for internet of things. *Journal of Network and Computer Applications, 42*, 120–134.

33. Phan, D. H., Suzuki, J., Omura, S., Oba, K., & Vasilakos, A. (2014). Multiobjective communication optimization for cloud-integrated body sensor networks. In *Cluster, cloud and grid computing (CCGrid), 2014 14th IEEE/ACM international symposium on* (pp. 685–693).

34. Xiao, Y., Peng, M., Gibson, J., Xie, G. G., Du, D. Z., & Vasilakos, A. V. (2012). Tight performance bounds of multihop fair access for MAC protocols in wireless sensor networks and underwater sensor networks. *IEEE Transactions on Mobile Computing, 11*(10), 1538–1554.

35. Rahimi, M. R., Ren, J., Liu, C. H., Vasilakos, A. V., & Venkatasubramanian, N. (2014). Mobile cloud computing: A survey, state of art and future directions. *Mobile Networks and Applications, 19*(2), 133–143.

36. Lin, J., Xiong, N., Vasilakos, A. V., Chen, G., & Guo, W. (2011). Evolutionary game-based data aggregation model for wireless sensor networks. *IET Communications, 5*(12), 1691–1697.

37. Logambigai, R., & Kannan, A. Fuzzy logic based unequal clustering for wireless sensor networks. *Wireless Networks*, 1–13.

38. Guo, W., Xiong, N., Vasilakos, A. V., Chen, G., & Yu, C. (2012). Distributed k–connected fault–tolerant topology control algorithms with PSO in future autonomic sensor systems. *International Journal of Sensor Networks, 12*(1), 53–62.

39. Guo, W., Park, J. H., Yang, L. T., Vasilakos, A. V., Xiong, N., & Chen, G. (2011). Design and analysis of a MST-based topology control scheme with PSO for wireless sensor networks. In *Services computing conference (APSCC), 2011 IEEE Asia-Pacific* (pp. 360–367).

40. Heinzelman, W. B., Chandrakasan, A. P., & Balakrishnan, H. (2002). An application specific protocol architecture for wireless microsensor networks. *IEEE Transactions on Wireless Communications, 1*(4), 660–670.

41. Tillet, J., Rao, R., & Sachin, F. (2002). Cluster head identification in adhoc sensor networks using particle swarm optimization. In *IEEE international conference on personal wireless communications* (pp. 201–205).

42. Abbas, K., Abedini, S. M., Faraneh, Z., & Al-Haddad, S. A. R. (2013). Cluster head selection using fuzzy logic and chaotic based genetic algorithm in wireless sensor network. *Journal of Basic and Applied Scientific Research, 3*(3), 694–703.

43. Enan, A., Bara, A., & Attea, A. (2011). Energy-aware evolutionary routing protocol for dynamic clustering of wireless sensor networks. *Swarm and Evolutionary Computation, 1*(4), 195–203.

44. Latiff, N. M. A., Tsemenidis, C. C., & Sheriff, B. S. (2007). Energy-aware clustering for wireless sensor networks using particle swarm optimization. In *Proceedings of 18th annual IEEE international symposium on personal, indoor and mobile radio communications* (pp. 1–5).

45. Buddha, S., & Lobiyal, D. K. (2012). A novel energy-aware cluster head selection based on particle swarm optimization for wireless sensor networks. *Human-Centric Computing and Information Sciences, 2*(1), 2–13.

46. Kuila, P., Gupta, S. K., & Jana, P. K. (2013). A novel evolutionary approach for load balanced clustering problem for wireless sensor networks. *Swarm and Evolutionary Computation, 12*, 48–56.

47. Kuila, P., & Jana, P. K. (2014). A novel differential evolution based clustering algorithm for wireless sensor networks. *Applied Soft Computing, 25*, 414–425.

48. Sengupta, S., Das, S., Nasir, M., Vasilakos, A. V., & Pedrycz, W. (2012). An evolutionary multiobjective sleep-scheduling scheme for differentiated coverage in wireless sensor networks. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 42*(6), 1093–1102.

49. Song, Y., Liu, L., Ma, H., & Vasilakos, A. V. (2014). A biology-based algorithm to minimal exposure problem of wireless sensor networks. *Network and Service Management, IEEE Transactions on, 11*(3), 417–430.

50. Acampora, G., Gaeta, M., Loia, V., & Vasilakos, A. V. (2010). Interoperable and adaptive fuzzy services for ambient intelligence applications. *ACM Transactions on Autonomous and Adaptive Systems (TAAS), 5*(2), 8.

51. Dietrich, I., & Dressler, F. (2009). On the lifetime of wireless sensor networks. *ACM Transactions on Sensor Networks, 5*(1), 1–38.

52. Xu, J., Liu, W., Lang, F., Zhang, Y., & Wang, C. (2010). Distance measurement model based on RSSI in WSN. *Wireless Sensor Networks, 2*(8), 606–611.

Haider Banka received his Ph.D. from Jadavpur University in 2008. He joined in the Department of Computer Science and Engineering, Indian School of Mines, Dhanbad, India in 2007 as an Assistant Professor. Currently he is an Associate Professor of the Department of Computer Science and Engineering, Indian School of Mines, Dhanbad, India. He has contributed more than 50 research papers and authored 2 books. He acted as referee in many reputed journals including IEEE Transaction on Evolutionary Computation, IEEE Transactions on Systems, Man, Cybernetics, Pattern Recognition, etc. Dr. Banka has served as the editorial board members, Co-chairs and PC members of several conferences of national and international reputes. He awarded as EU-INDIA fellowship, from April, 2006 to July, 2006, Department of Computer and Information Science (DISI), University of Genova, Italy. He was Visiting Research Fellow at Centre of Soft Computing Research (ISI Kolkata) funded by Department of Science & Technology (DST), from July, 2006 to August, 2007 and awarded INAE Teacher Mentoring Fellowship funded by Indian National Academy of Engineering (INAE) from June, 2008 to July, 2008. His current research interest includes Soft Computing, Algorithms, Data Mining, Pattern Recognition, Combinatorial Optimization and other interdisciplinary research areas.

P. C. Srinivasa Rao received his MCA from JNTU, Hyderabad in 2011. Currently he is a Senior Research Scholar in the Department of Computer Science and Engineering, Indian School of Mines, Dhanbad, India and pursuing his Ph.D. work in the field of Wireless Sensor Networks using Soft Computing. Before that Pasupuleti worked as a Software Consultant in ECIL-RAPIS-CAN LTD, Hyderabad during December, 2011 to August, 2012. His main area of research includes Wireless Sensor Networks and Nature Inspired Optimization Techniques.