Redundancy estimation and adaptive density control in wireless sensor networks

Renita Machado†, Guiling Wang‡, and Sirin Tekinay†
Department of Electrical and Computer Engineering†, Department of Computer Science‡
New Jersey Institute of Technology
Email: {rmm23,gwang,tekinay}@njit.edu

***Abstract*— While dense random deployments satisfy coverage and sensing requirements, constructing dense networks of sensor nodes poses economical constraints as well as the problem of redundancy. We provide an analytical framework for estimating the redundancy in a random deployment of nodes without the need of location information of nodes. We use an information theoretic approach to estimate the redundancy in a randomly deployed wireless sensor network and provide the Cramer-Rao bound on the error in estimating the redundancy in a wireless sensor network. We illustrate this redundancy estimation approach and calculate the bounds on the error in estimating the redundancy for a wireless sensor network with 1-redundancy. We also analytically show the inter-dependence between redundancy and network lifetime for random deployment. We further study the energy model of a WSN as interdependence between the environmental variation and its impact on the energy consumption at individual nodes. Defining network energy as the sum of residual battery energy at nodes, we provide an analytical framework for the dependence of node energy and sensitivity of network energy as a function of environmental variation and reliability parameters. Using a neural network based approach, we perform adaptive density control and show how reliability requirements and environment variation influences the rate of change of network energy.**

I. Introduction

A key challenge in WSNs is the deployment of nodes to satisfy the requirements of higher network lifetime with continuous coverage of the deployment region for reliable sensing. While continuous coverage and connectivity lead to higher reliability of sensing operation, it takes a toll on the battery life of individual nodes and consequently reducing network lifetime. One way of balancing this tradeoff is deploying more sensors than are required to cover the deployment region. While this redundancy approach can be used to increase network lifetime by the use of sleep-scheduling and power-aware routing protocols, the economic constraints of deploying large number of nodes poses a limitation. The problem of deployment has been widely studied in the context of providing uniform coverage, connectivity and redundancy while optimizing the number of sensor nodes required for the sensing operation [[1], [2]]. The random placement of nodes is preferred over the deterministic placement approach for remote and hostile environments, where it is not possible to place sensors in a particular pattern to cover the entire deployment region, for example, forest ecology sensing environments. Deployment for such applications is carried out by randomly scattering the nodes over the deployment region. While this approach has the advantage of eliminating the overhead of planning and deterministic placement, it also gives rise to the problem of not knowing the density and location of sensors in the deployment region. In addition, some network operations require higher density of sensors in regions of high-interest phenomenon. Since equipping nodes with location detecting GPS receivers is expensive and we do not know the precise location of sensors due to the random deployment, we need to be able to estimate the probability density function (PDF) of node distribution over the placement region and calculate the redundancy of nodes. The nodes can then be instructed to follow some degree of sleep-awake duty cycle to satisfy the requirements of sensing for that area of the deployment region while also contributing to higher network lifetime. The problem we study can hence be modeled as follows: In a wireless sensor network of randomly deployed, stationary, power-limited, homogenous sensor nodes designed to be operational for at least time T, where we have no location information, how do we estimate the redundancy of nodes in the network? The time constraint T ensures that all nodes in the network are designed to perform sensing and data processing without loss of battery energy for at least entire duration T, thus increasing reliability of system operation. One way of doing this would be intelligent processing at the nodes to discover the number of neighbors in their one-hop sensing range and calculating the redundancy. However, this approach, calls for higher number of transmissions at every node to discover the number of neighbors, which leads to increase in the depletion of battery energy for the energy-constrained nodes. Another way of estimating the redundancy would be from an analysis of the node deployment strategy. While this approach eliminates the need for sensors to discover the level of redundancy in their neighborhood, it results in greater variance of the error in estimating the redundancy due to the uncertainty of the deployment strategy information. In this paper, we propose an information theoretic approach to estimate the redundancy in a randomly deployed network. Every node transmits its node ID to the base station in the initial phase of the network operation. From these transmissions, the base station gathers the node IDs and the signal strength information. The signal strength data is used to estimate the PDF $p_X(x)$ of the node distribution over the deployment region using the MinMax measure [3]. We then decompose this PDF $p_X(x)$ into a primary distribution and an unknown number of secondary distributions. This unknown parameter θ is the redundancy that we estimate. Further, we provide bounds on the support of the random variable x that describes the PDF of the node distribution over the deployment region. Finally, we provide the Cramer-Rao lower bound and an upper bound using Barron's proof [4] on the error in estimating the redundancy in the deployment region.

The next part of this paper focuses on developing an energy model for WSNs that takes into account the interdependence between the environmental variation and its impact on the energy consumption at individual nodes. We use the information about redundancy of nodes in the WSN to perform adaptive density control with the help of a neural network algorithm that adapts the data dissemination according to the change in the environmental phenomenon being monitored.

The rest of the paper is organized as follows: Section II presents the preliminaries for the WSN model and the PDF estimation technique based on the MinMax measure [3]. In section III, we develop the analytical model to obtain the bounds on the Fisher information and the Cramer-Rao bounds for the error in estimating the redundancy parameter. We illustrate our approach with a case of 1-redundancy in the deployment region. Section IV discusses the variation of error bounds of redundancy estimation and validates the relationship between redundancy and network lifetime. Section V introduces the motivation for developing an energy model based on environmental variations. Section VI describes the adaptation of the Boltzmann learning algorithm for WSNs by taking into account reliability requirements of the sensing operation. In section VII, we develop an analytical model for the environment variation and its influence on node energy. Section VIII describes the use of the Boltzmann rule in this framework to determine the density of nodes for the given reliability requirements. Section IX presents the results of numerical simulation of the proposed energy model. Finally, section X concludes the paper and presents directions for future work.

A. *Related Work*

In [5], the authors study redundancy in terms of redundant broadcasts as a consequence of broadcasting by flooding in a mobile ad hoc network. By simulation, they show that for k greater than or equal to four neighbors, the expected additional coverage is below 0.05 %, i.e. benefit of rebroadcast is small. In [6], the authors analyze sensor redundancy by finding bounds on the neighbor set of a sensor node. They provide an analytic framework to determine the percentage of redundant area with n number of neighbors and the probability that a node is completely redundant. Specifically, they show that if a sensor is completely redundant, at least three and at most five neighbors are needed to cover its sensing area. The analytical model developed in [6] shows that for a 90 % partial redundancy, i.e. requiring 90% of its sensing area to be covered, needs five neighbors, which is similar to the simulation results obtained in [5]. Our work differs in that we do not require the sensors to be aware of the number of neighbors in their sensing range and thereby eliminate the need for processing and storage of the information related to neighbor discovery.

II. Preliminaries

We assume a dense wireless sensor network of homogenous, stationary, power limited sensor nodes densely and randomly deployed over the deployment region designed to operate for at least T time units. This time constraint on operation ensures that the resulting deployment and operation of the network is reliable for the desired interval. The problem is to estimate the redundancy in such a deployment without the knowledge of location information of sensor nodes. We assume that the base station initiates a phase of node-discovery, where it broadcasts a query transmission asking every node to respond to the base station with its node ID. From the signal strength of the transmissions of the nodes transmitting their node IDs, the base station obtains the relative location of the nodes in the deployment region. This problem is known as the direct problem [3], which refers to the problem of finding the initial probability assignment consistent with available information about a probabilistic system. In our problem formulation, the signal strength of individual transmissions comprises the available information about the node locations in the deployment region. The signal strength information constitutes the sample data and is used to estimate the PDF of node distribution over the deployment region. In this paper, we use the approach developed in [3] to obtain the PDF. We refer the reader to [3] for a detailed explanation of the approach. We proceed to estimate the redundancy of nodes in the deployment region from the PDF of the node location. We introduce the following definitions to aid in the redundancy calculations:

Primary distribution: The primary distribution refers to the PDF of sensor nodes' distribution that is necessary and sufficient to provide coverage and connectivity in the deployment region. The primary distribution ensures continuous sensing throughout the deployment region.

Secondary distributions: In order to provide continuous sensing in spite of battery exhaustion or device failure in nodes belonging to the primary distribution, as well as to improve the reliability of the sensing operation, we deploy additional nodes over the deployment region. The distribution of these redundant nodes has a PDF called the secondary distribution.Depending on the sensing requirements, minimum network lifetime constraints and the economic constraints and resources available for the sensing operation, we can deploy *k* distributions of sensor nodes in addition to the primary distribution over the deployment region resulting in *k*-redundancy.

The final step in the problem of redundancy estimation would be to estimate the number of secondary distributions (redundancy) over the deployment region from the initial sample set of signal strengths that provides the relative location of nodes in the deployment region. We also provide Cramer-Rao bounds on the error in estimating the redundancy in the deployment. To do this, we determine the Fisher information of the redundancy parameter θ, which is obtained from the estimated PDF (solution to the direct problem). This is a more general version of the redundancy estimation problem [5] in the absence of location information. Our results bring insight into the general problem of redundancy in dense, large WSNs where the estimate of redundant distributions can be useful to selectively power down or sleep schedule nodes in certain distributions to satisfy lifetime constraints or increase the density of sensing operation in high interest areas of the deployment region. The advantage of this approach lies in the fact that nodes do not need to possess computational complexity to process information about their neighbors to calculate redundancy in their coverage areas. The base station performs the processing to determine the redundancy in the entire deployment region. This approach can also be used to control the density of 'awake' nodes in sections of the deployment region by applying the redundancy processing approach to specific areas that exhibit high interest phenomenon.

III. Cramer-Rao bounds in error in redundancy estimation using the Fisher information of redundancy

To illustrate this method of estimating redundancy, we assume a primary distribution and a single secondary distribution of nodes over the deployment region, thus creating a 1-redundant network of nodes, i.e. every point is sensed by two nodes at all times.

A. Bounds on the PDF of node distribution over the deployment region

Let the primary distribution have a PDF $p_S(x)$ and let the secondary distribution have a PDF $p_Z(x)$, where S is a random variable describing the primary distribution, and $X = S + Z_s^{(\tau)}$ describes the resultant distribution of all nodes in the deployment region. The secondary distribution $Z_s^{(\tau)}$ is a normal distributed random variable $\sim N(0, \tau)$ independent of S. Thus, X which denotes the PDF of all nodes in the entire deployment region is a perturbed random variable with continuously differentiable density $p_X^x(\tau)$. Defining the score function as $\rho_X(x)$,

$$\rho_X(x) = p'_X(x)/p_X(x) \tag{1}$$

and $p^{(2\tau)}$ for the density of $S + Z_S^{(2\tau)}$ there exists a constant [4]

$$c_{\tau,k} = \sqrt{2}\,(2k/\tau e)^{k/2} \tag{2}$$

such that for all x [4],

$$p_X^{(\tau)}(x)\,|\rho_X(x)|^k \leq c_{\tau,k}\,p^{(2\tau)}(x) \tag{3}$$

B. Illustration of the redundancy estimation approach for 1-redundancy

A widely used assumption to model random deployment of nodes over a deployment region [1], [2] is the Poisson point process. Let the primary distribution $p_S(x)$ denote a Poisson point process of intensity λ.

$$p_S(x) = \frac{\lambda^x e^{-\lambda}}{x!} \tag{4}$$

Since $X = S + Z_s^{(\tau)}$,

$$p_X^{(\tau)}(x) = p_S(x) \otimes p_Z(x)$$

.

Using Fourier transforms to obtain the convolution, the PDF $p_X^{(\tau)}(x)$, is given by

$$p_X^{(\tau)}(x) = \frac{1}{x!\,j^{(2x+2)}t^{x+2}}\sqrt{\frac{2\pi}{\tau}} \tag{5}$$

The score function $\rho_X(x)$ for (5) is given by

$$\rho_X(x) = -\left[\frac{x\ln(-t)+1}{x}\right] \tag{6}$$

The density $p^{(2\tau)}$ is obtained as follows

$$p^{(2\tau)} = \sqrt{\frac{\pi}{\tau}} \frac{1}{x! j^{2x+2} t^{x+2}} \tag{7}$$

Substituting (9) and (10) in (3) and simplifying we get the bounds on the density function of the nodes in the deployment region,

$$p_X(x) \leq \frac{-\sqrt{\frac{2\pi}{\tau}} \left(\frac{2k}{\tau e}\right)^{k/2} \frac{1}{x!(-t)^x t^2}}{\left[\frac{x \ln(-t)+1}{x}\right]^k} \tag{8}$$

Substituting (8) in (11) to obtain the support for *x*, and evaluating *t* in the finite integral limits from 0 to T , where T is defined as the minimum time for which the network is designed to be operational (i.e. no node has run out of battery energy), the support for *x* is given by,

$$x \geq \frac{2}{\sqrt{\tau e}} - \ln(-T) \tag{9}$$

Next, we evaluate the Fisher information of the redundancy parameter θ. The Fisher information $J(X)$ is given by the variance of the score function ρ and satisfies the following bound for the random variables *S* and *X* [4],

$$J(X) = E\left|\rho_X(x)\right|^2 \leq \frac{5.658}{\tau e} \tag{10}$$

Substituting for τe from (12) in (13),

$$J(X) = E\left|\rho_X(x)\right|^2 \leq \sqrt{2}\left(x + \ln(-T)\right)^2 \tag{11}$$

In our problem formulation, the unknown parameter, which we estimate, is the redundancy θ , the number of secondary distributions. To obtain the Fisher information of the redundancy parameter $J(\theta)$, we note that the Fisher information $J(X)$ can also be written as

$$J(X) = \int f(x-\theta)\left[\frac{\partial}{\partial x} \ln f(x-\theta)\right]^2 dx \tag{12}$$

The Fisher information for the redundancy parameter is given by

$$J(\theta) = \int f(x-\theta)\left[\frac{\partial}{\partial \theta} \ln f(x-\theta)\right]^2 dx \tag{13}$$

Further, since $\frac{\partial}{\partial x} \ln f(x-\theta) = -\frac{\partial}{\partial \theta} \ln f(x-\theta)$,

$$\int f(x-\theta)\left[\frac{\partial}{\partial x} \ln f(x-\theta)\right]^2 dx = \int f(x-\theta)\left[\frac{\partial}{\partial \theta} \ln f(x-\theta)\right]^2 dx. \tag{14}$$

which implies $J(\theta) = J(X)$.

The Cramer-Rao gives a lower bound on the estimation of variance of an unknown parameter. For a random variable *X* with mean μ and variance σ^2,

$$\sigma^2 \geq \frac{1}{J(X)} \tag{15}$$

with equality if and only if *X* is $N(\mu, \sigma^2)$.

The inverse of the Fisher information given by $1/J(X)$ is the Cramer-Rao bound. The Cramer-Rao inequality places a lower bound on the mean square error of an unbiased estimate of *X* given by σ^2. Since $J(\theta) = J(X)$,

$$\frac{1}{\sigma^2} \leq J(\theta) \leq \sqrt{2}\left[x + \ln(-T)\right]^2 \tag{16}$$

The Cramer Rao bound on the error in estimating the redundancy θ for the case of a primary Poisson distribution and a single secondary normal distribution is thus given by

$$\frac{1}{\sqrt{2}\left[x + \ln(-T)\right]^2} \leq \frac{1}{J(\theta)} \leq \sigma^2 \tag{17}$$

IV. Results

We now show the significance of (21) in establishing the bounds of error in redundancy estimation as a function of the expected network lifetime. We calculate the variance of $p_X(x)$ from (11) as

$$\sigma^2 = (-T)^{-\left(\frac{2}{\sqrt{\tau e}} - \ln(-T)\right)}\left[\frac{-2K_1}{\sqrt{\tau e}} - \frac{K_2\,(-T)^{-\left(\frac{2}{\sqrt{\tau e}} - \ln(-T)\right)}}{\left(\frac{2}{\sqrt{\tau e}} - \ln(-T)\right)^2}\right] \tag{18}$$

where

$$K_1 = \frac{-1}{T^2}\sqrt{\frac{2\pi}{\tau}} \ln(-T)$$

and

$$K_2 = \left[\frac{K_1}{\ln^2(-T)}\right]^2$$

Evaluating (22) at the support of *x* from (12), the bounds on the error in estimating redundancy are as follows,

$$\frac{\tau e}{4\sqrt{2}} \leq \frac{1}{J(\theta)} \leq \sigma^2 \tag{19}$$

Let τ be a fraction α of the desired network lifetime *T*. We show that the error bounds exist value as long as α constant c_1. For $c_1 \leq \alpha \leq c_2$, bounds may hold true depending on designed network lifetime *T*. For α¿ c_2, there is no upper bound on the error information, i.e. the error in estimating the redundancy is infinite. These bounds are illustrated as shown in Figure 1. The concentric circles represent the areas where the bounds exist for corresponding values of *T* and α. As seen from Fig. 1, the circles are plotted in increasing order of desired network lifetime. We see that with an increase in network lifetime, the upper bound of the error in estimating the redundancy keeps decreasing. Since the lower bound in (24) is a constant, this implies that the error bound gap decreases for higher desired network lifetime. This can be intuitively explained as follows: For a larger desired network lifetime with given number of nodes, we need to maintain a lower value of redundancy in the deployment region. A lower value of redundancy translates to lesser number of 'awake' nodes covering the deployment region at any given time instant, thus increasing the network lifetime. Conversely, for lower desired network lifetime, higher number of nodes can be made to cover the deployment region by staying 'awake' for longer time intervals leading to a higher redundancy. Thus, the error in estimating a lower value of redundancy (higher network lifetime) is smaller than that in estimating a higher value of redundancy (lower network lifetime). Numerical results from simulating (24) for different desired network lifetimes T, shows that for $\tau \leq 0.7$, the error is bounded. The upper bound of error for T =1000 hours and 100 hours are close and shown as overlapping circles ($\tau \leq 0.01$). The next larger circle shows the error bound for T =10 hours with $\tau \leq 0.03$. The shaded region in Fig.1 represents the value of lower network lifetimes for which the error bounds may not be satisfied due to decrease in T. For extremely low value of desired network lifetime, the error bounds for redundancy estimation are not satisfied, implying that the error in estimating redundancy is infinite. Fig.2 shows the upper bound of error estimation as a function of support of the random variable representing the node distribution for 1-redundancy with T = 5 hours. We see that for increasing lower limit of the support, the error in estimating the redundancy increases due to the inverse dependence of support on T.

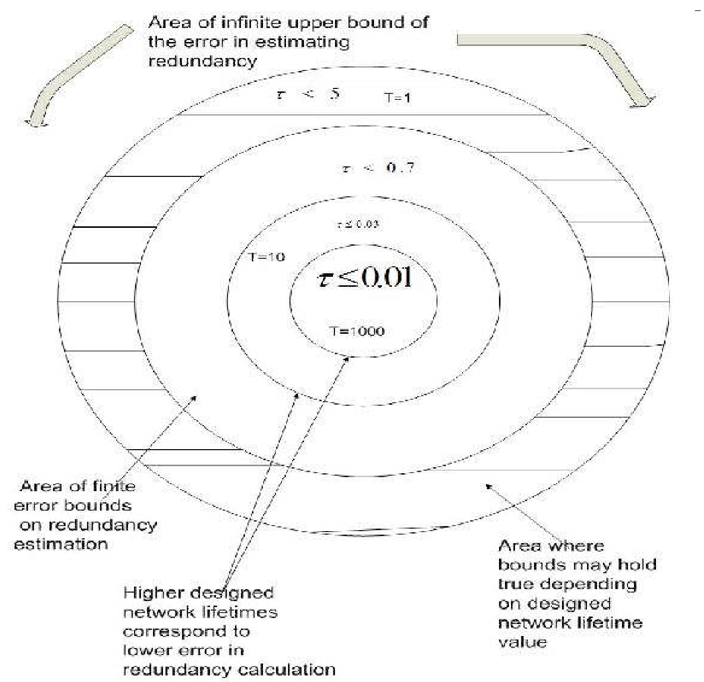

Fig. 1. Interdependence of network lifetime and error in redundancy estimation for 1-redundancy. Higher network lifetimes correspond to lower redundancy and hence lower error in estimating redundancy. For the converse case of lower network lifetime, the error in redundancy estimation is higher.

V. NEURAL-NETWORK BASED ADAPTIVE DENSITY CONTROL

In this section, we describe a neural-network approach for density control based on the interdependence between environment variation and node energy. Our first objective is to develop an approach that takes into account reliability of the sensing operation. The goal is to vary the density of 'awake' nodes according to a user-defined reliability requirement. For instance, a higher reliability requirement requires a higher density of nodes in the deployment region that are continuously sensing and transmitting to a central base station (BS) that acts as a sink. The second objective is to account for variation of environment in parts of the deployment region. Since some areas of the deployment region may exhibit different environment variations, the rate of reporting data from nodes to BS should allow for adaptive density control in different areas of the deployment region. Specifically, we use the Boltzmann learning rule to choose the set of active sensors for a given sensing cycle. Our scheme incorporates features of reliability by providing means to increase the density of active sensors and/or increasing the rate of reporting sensed data to the base station (BS).

This approach for density control draws upon two important areas of research for intelligent WSNs. The first of these is energy optimization, and a key challenge in energy optimization for densely deployed WSNs is selecting the set of sensors that remain 'awake' for a given cycle. Some of the criteria developed for choosing the set of active nodes are environment probing [6] to determine active neighbors, k-coverage [7] and connectivity-based participation in multi-hop network [8]. While these approaches target the WSN architecture and individual node lifetimes, they overlook the dependence of the battery variation on the variation in the sensed environment. Since nodes typically report data to the base station when the sensed data exhibits large variance, a rapidly changing environment influences the energy consumption at nodes due to higher number of transmissions from nodes to base station/other nodes. Hence, it is essential to model the WSN system with interdependence between the WSN and the environment. This kind of a system modeling approach reflects the sensitivity of network lifetime to the pattern of variation in the environment with the help of bifurcation parameters of the environment model.

The second is the use of a neural-network based approach to create intelligent, adaptive WSNs. Approaches from

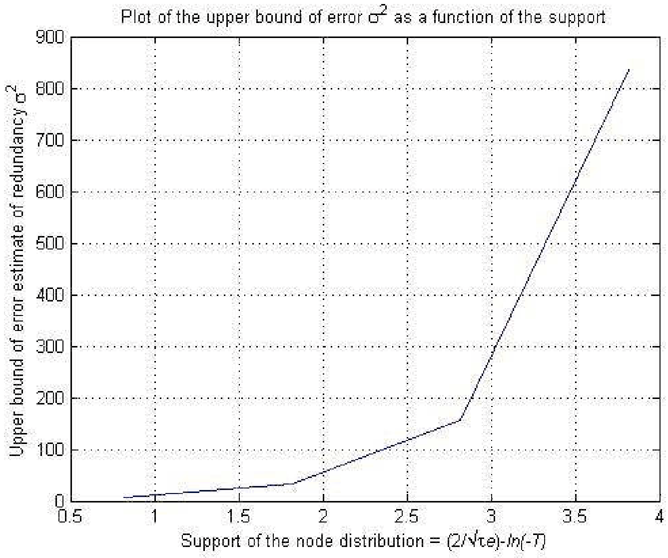

Fig. 2. Variation of error bound gap as a function of support of the random variable describing node distribution. The support is inversely dependent on the network lifetime. A lower value of the support implies a higher network lifetime and hence lower error in redundancy and its estimation.

neural networks (NN) have been widely used in WSNs for routing, fault recognition and modeling [9], [10], [11] and [12]. Recent research on specific WSN applications has exposed the impact of environment modeling on the WSN performance parameters [13]. In [13], the authors study the data gathering problem, in which they model the physical phenomena being sensed as a system of partial differential equations and have sensors transmit estimates of data rather than the raw data. Our work differs in that we develop the WSN energy model by taking into account the environment variation and the dependence of node energy change on this variation. We then use the Boltzmann rule to train the BS about the environment variation, so that it can predict the sensed data values and reduce the energy consumption involved in transmission of data from nodes to BS. It does this by taking into account the desired reliability constraints of the sensing operation and thus performs adaptive density control.

VI. Boltzmann Learning Rule for WSN power management

We assume that the WSN is deployed for *on-demand* sensing, where nodes report sensed data to the BS and the sensed data is transmitted to the BS with the help of direct communication links between nodes and the BS. In order to reduce the number of transmissions from the nodes to the BS and thereby conserve battery energy, the nodes report data only when the difference between the data sensed in

the current cycle differs from that in the previous cycle by a value that exceeds a certain threshold. This threshold is determined by the user demand for reliability, and is known as the reliability factor γ. The approach we take to integrate the Boltzmann machine [14] with a WSN can be detailed as follows.

We assume that the nodes in the WSN are in one of the following two states: visible or hidden. The visible/hidden state indicates the state of the transceiver in individual nodes, and thereby whether the node is 'visible' to the BS in terms of its ability to transmit/receive data to/from the BS. Studies of energy consumption in wireless sensor nodes indicate that the transceiver energy consumption is an order of magnitude higher than that of other components in a wireless sensor node. To facilitate further energy savings, within each of the visible/hidden states, a node can be either in the '0' or '1' modes which signifies the on/off states of the various components that make up the wireless sensor node. Thus at any given instant during the network operation, a node can be in one of the following energy-saving states: visible-0 (V0), visible-1(V1), hidden-0 (H0), hidden-1 (H1). The description of each of these states is as follows:

V1 (visible transmit): transceiver, sensors and data processor on.

V0 (visible receive): transceiver and data processor on

H1 (hidden sense): only transducer/sensor on

H0 (hidden sleep): all components off.

Boltzmann learning rule: The aim of the Boltzmann learning rule-based network density control is to obtain reliable operation while prolonging network lifetime. We do this by adaptively learning the variation of the environment and minimizing the amount of transmission-related energy expenditure at the nodes. The network goes through two distinct phases: training/clamped phase and the trained/free-running phase. The nodes enter *V1*, *V0*, *H1* or *H0* states in the training and trained phase as follows:

Training/clamped phase:

V1: Initially, all nodes are in the *V1* state. Nodes continuously gather and transmit data to the BS. Most real-world systems such as temperature variations, humidity and light variations follow an oscillatory pattern with diurnal, seasonal or daily variations. WSNs deployed to sense such parameters can exploit this pattern of variation with external and parametric noise and the bifurcation parameters. The BS learns the pattern of environment variation from this data. Since the data collected follows a high degree of temporal correlation, the BS decides to stop data collection from nodes in the training phase when it begins to observe the identical patterns of data collected, for e.g., photodetectors measuring the amount of light incident on an area at certain times of the day on consecutive days. The BS then instructs the nodes to enter the trained/free-running phase.

Trained/free-running phase:

V0: In order to reduce the number of transmissions from the nodes to the BS and thereby conserve battery energy, the nodes report data only when the difference between the data sensed in the current cycle differs from that in the previous cycle by a value that exceeds a certain threshold. This threshold is determined by the user demand for reliability, and is known as the reliability factor γ. The value of the reliability factor γ lies in the range 0 to 1, where γ=1 implies a WSN with all on nodes for the entire duration of network operation, i.e. γ=1 implies the θ- or k- redundancy of operation. Having obtained the knowledge of redundancy from the method outlined in Section 3, the BS instructs γ % of nodes to enter the visible-receive (*V0*) mode, while the other nodes enter the hidden mode such that the density of 'awake' nodes satisfies the user requirements of reliability. The highest reliability is obtained when all nodes are in the V1 state and continuously reporting data to the BS. The data obtained by the BS in the *V1* training phase is retrieved for analysis and prediction in the trained phase. To this end, the BS then uses prediction functions to evaluate the data values sensed by the nodes across the network and broadcasts these values to the nodes. Nodes compare the predicted values received in the BS broadcast message with the sensed value obtained in the previous cycle. If the difference fails the reliability requirement i.e. exceeds the threshold, the nodes enter the *V1* mode and transmit the sensed value to the BS. The BS can thus fine tune its prediction algorithm to better reflect the changes in the sensed environment in the whole or part of the deployment region depending upon the number of nodes who enter the *V1* training phase again from the *V0* trained phase. Thus we see that the reliability factor controls not only the density of awake nodes but also the rate of reporting and thereby the energy consumption of nodes. A higher reliability requirement for the sensing operation implies a smaller threshold, which can be rapidly exceeded due to faster variation of the sensing environment in the deployment region.

H1 and *H0*: These states constitute the hidden mode of the learning process. The amount of time spent in the hidden mode is inversely proportional to the time spent by the network in the visible/training mode. This is because, a larger duration of the training phase implies a complex pattern of environment variation which requires longer time for the BS to learn and develop prediction functions for them. Hence, we program the nodes for shorter hidden mode intervals so that the WSN can continue to reliably sense the environment.

The nodes instructed to enter the hidden mode first enter the *H0* (hidden-off mode) for γ of the hidden mode duration and then switch into the *H1* mode for the remaining duration of the hidden mode. This is done so that when the nodes enter the visible-receive (*V0*) mode again after leaving the hidden mode, the data values sensed by the sensors in the *H1* mode can be used as the latest reference with which to compare the BS prediction.

Assumptions: Since the accuracy of the Boltzmann learning rule-based density control algorithm relies on the accuracy of the prediction algorithms and approximating functions at the BS, we assume that the BS has superior processing capacity and is not power-limited like the nodes in the WSN. Nodes have buffer space to store data values obtained in the previous cycle and possess data processors with comparator algorithms. The clocks at individual nodes are synchronized with one another and the BS to achieve correct hidden mode/visible mode transitions. The communications links are assumed to be error-free. The only source of error is the error in the sensed data at nodes, which reflects as an increase in the energy consumption when the nodes perform comparisons with the BS's predicted values and find that the difference fails the reliability requirement.

VII. Analytical Framework of the Model

A. Training/Clamped phase

We model the environment by the Van der Pol system of equations. Van der Pol equations are widely used to model oscillatory equations in systems such as electric circuits and population dynamics. The choice of the equations used to model the environment variation depends on the sensed environment. As we shall show in this section, the choice of the equations used to model the environment bears no influence on the energy model, thereby reducing the model to a set of equations with bifurcation parameters and noise in the sensed data. However, the bifurcation parameters of the system model describing the environment affect the stability of the system and can be used to design the length of the energy saving states within a cycle. The model for the rate of change of environment is given by the noisy Van der Pol equation [15] in x as a function of time. The node energy level is given by y, whose rate of change is proportional to the rate of change of the environment by a factor p, which we call as the dependence function and also to the residual node energy at that time.

$$x'' + (\beta + \sigma_1 \xi_1)\, x + \alpha \left(x^2 - 1\right) x' + \sigma_2 \xi_2 = 0 \tag{20}$$

$$y' = y - px - \varepsilon \tag{21}$$

where,α and β are the bifurcation parameters,$\sigma_1 \xi_1$ and $\sigma_2 \xi_2$ denote the external and parametric noise, ε is the energy per bit per node [16].

Constraint: We evaluate this environment variation vs node-energy variation model subject to a reliability constraint γ. The intuition behind the constraint function is this: for prolonged network operation, we require that the reliability requirement should not require the nodes to remain in the *V1* (all-awake) mode at all times, since this causes rapid energy depletion and decrease in network lifetime. Also, the rate of change of the sensing environment should be less than the rate of change of battery level at the nodes to reduce the number of transmissions at nodes. Taking the product of these constraints, we introduce the reliability constraint for the WSN as follows: the product of the reliability requirement and the rate of change of the sensing environment should be proportional and less than or equal to the square of the rate of change of node energy. This constraint ensures that the system remains stable by providing a means by which the WSN trains the BS reliably according to the rate of change in the environment model.

$$\gamma x' \leq k y'^2 \tag{22}$$

Using the Lagrangian to find the extrema of (20) subject to the constraint posed by (22), the optimization function can be written as

$$\Delta(x, y, \lambda) = f(x, y) + \lambda g(x, y) \tag{23}$$

where the objective function $f(x, y)$ and the constraint function $g(x, y)$ are as follows,

$$f(x,y) = x'' + \alpha \left[\left(\left(y' - y + \varepsilon \right) / p \right)^2 - 1 \right] x' + \beta \left(y' - y + \varepsilon \right) / p \tag{24}$$

$$g(x, y) = \gamma x' - k y'^2 \tag{25}$$

To find the extrema of the function, we obtain the partial derivatives of (23) and equating them to zero we get,

$$x' = k y'^2 / \gamma \tag{26}$$

Assuming that the higher partial derivatives tend to zero, we obtain

$$\frac{\alpha}{\gamma} y'^2 \left(\varepsilon + y - y' + 1 \right) = 0 \tag{27}$$

To find the solutions of (27), since $\alpha/\gamma \neq 0$,

1) $y'^2 = 0$, i.e.

$$\frac{dy}{dt} = 0 \Rightarrow y = -px - \varepsilon \tag{28}$$

Battery level is a function of environment variation and node energy consumption.

2) $(\varepsilon + y - y' + 1) = 0$, i.e.

$$y' = \varepsilon + y + 1 \tag{29}$$

But $y' = y - px - \varepsilon$ from case 1. Hence,

$$x = -\left(2\varepsilon + 1 \right) / \tag{30}$$

Solving the differential equation in (28) to obtain y,

$$y = c_0 e^t - (\varepsilon + 1) \tag{31}$$

where, c_0 is a constant of integration.

From (30),

$px = -\left(2\varepsilon + 1 \right)$

Let $p(x) = px$, thus the extremum for the objective function is

$$(x, y) = \left(-\left(2\varepsilon + 1 \right) / p, c_0 e^t - (\varepsilon + 1) \right) \tag{32}$$

Thus, the choice of the system of equations used to describe the environment does not influence the energy model. The dependence function p which describes the dependence of node energy on environment variation is used in the calculation of weights to decide the node states. The objective function $f(x, y)$ evaluated at this extremum is

$$f\left(x, y\right) = \beta \left(2\varepsilon + 1 \right) / p \tag{33}$$

which is a function of the dependence p of node energy on the environmental variation.

B. Trained phase/ Free-running phase

In the free-running phase, the BS uses the data acquired during the training phase to reduce the number of transmissions from the nodes to the BS. The BS broadcasts the predicted values to the visible nodes, which in turn match it with their sensed values. If the difference between the received prediction and sensed data exceeds a threshold, i.e. fails the reliability requirement, the visible nodes enter the visible transmit mode where they broadcast the sensed data to the BS. At the end of the free-running phase, the BS increases the density of visible-receive nodes for the next free-running cycle. This process goes on until the BS error converges to less than the threshold set at the nodes. If the procedure fails to converge until the point where the density of visible node equals the total density of nodes in the network, the BS initiates the training phase over again. The system can thus be modeled as follows [17]:

$$\begin{aligned} x' &= z \\ z' &= -\left(\beta + \sigma_1 \xi_1 \right) x - \alpha \left(x^2 - 1 \right) z - \sigma_2 \xi_2 \\ y' &= c \left(y - px - \varepsilon \right) \left(x - \gamma \right) \end{aligned} \tag{34}$$

Proceeding similar to the training phase, the value of the objective function at the extremum is

$$f\left(x, y\right) = p \gamma^2 c \tag{35}$$

This shows that the battery level is a function of

- reliability,
- dependence of node energy variation on environment variation and
- energy required per bit per node.

VIII. Calculation of node weights for density control

In this section, we illustrate the computation performed at the BS to determine the set of visible nodes for the next trained phase. The goal of the weight computation process is to balance the rate of energy consumption across the network. At the end of a trained phase, it assigns the nodes with lowest weights (lowest rate of change of battery energy) to enter the visible mode. The BS calculates the weight of nodes based on the number of transmissions from the node and the history of modes it has been in for the previous cycles.

A. Training phase

Let the total number of nodes in the network be N_T, where N_T , =N_h+N_b, N_h being the number of hidden nodes and N_b is the number of visible nodes. In the training phase, all nodes are visible. Thus, the number of hidden nodes, L = 0. Modifying the equation for the weight of a node according to the Boltzmann rule from [14] to adapt it to the WSN, we get

$$\Delta w_i = \eta \left(\rho_i^+ - \rho_i^-\right) \tag{36}$$

where,

$$\rho_i^+ = \sum_\alpha \sum_\beta P_{\alpha\beta}^+ s_{j|\alpha\beta} s_{i|\alpha\beta}$$

and

$$\rho_i^- = \sum_\alpha \sum_\beta P_{\alpha\beta}^- s_{j|\alpha\beta} s_{i|\alpha\beta}$$

The states of the hidden nodes are denoted by β, where β =$1, 2, ...2^L$, The states of the visible nodes are denoted by α, where α =$1, 2, ...2^K$

$P_{\alpha\beta}^-$: joint probability that the visible nodes are in state α and the hidden nodes are in state β, given that the network is in its free-running condition,

$P_{\alpha\beta}^+$: joint probability as above on the states of nodes , but for the network in its clamped condition.

$s_{i|\alpha\beta}$: state of node *i* given that the visible nodes are in state α and the hidden nodes are in state β.

η is the learning-rate parameter given by $\eta = \varepsilon/T$. The weight of a node is proportional to the rate of change of battery level at that node. From (31),

$$w_i = k_c \left[c_0 e^t - (\varepsilon + 1)\right] \tag{37}$$

Since $s_{j|\alpha 1} = s_{i|\alpha 1} = 1$ in the training phase i.e., *V1*,

$$\rho_i^+ = \left(2^{N_T}\right)/4$$
$$\rho_i^- = 0 \tag{38}$$

$$w_i = k_c \left[c_0 e^t - (\varepsilon + 1)\right] = \eta \left(2^{N_T}\right)/4 \tag{39}$$

which is only a function of time. This shows that the weights are only a function of time in the training phase. The intuition behind this is that in the longer the training phase, it implies a complex pattern of environment variation in the deployment region. This requires more number of transmissions from the nodes to the BS for the BS to learn about the environment and hence the weights of a node are only a function of time.

B. Trained phase

Adopting a similar procedure as above for calculation of weights in the trained phase, we show the weight of a node in the trained phase is

$$w_i = 0 - \rho_i^- = \sum_{\alpha=1}^{2^{N_b}} \sum_{\beta=1}^{2^{(N_T - N_b)}} \frac{s_{j|\alpha\beta} s_{i|\alpha\beta}}{2} = k_c \left(\varepsilon - p\gamma\right) \tag{40}$$

This shows that the weights are only a function of the reliability parameter γ. This is because, in the trained phase, the energy expenditure at a node due to communication with the BS is influenced by whether the difference between the BS predicted value and the sensed value stored in the node exceeds γ.

IX. Results

In this section, we obtain the simulation results of the Boltzmann learning-rule based WSN. We simulate a network of 100 nodes randomly scattered across a square deployment region of side 20 meters. Defining network energy as the sum of the battery energy of all nodes in the deployment region, we plot the rate of change of variation of the network energy in Ampere-hours/kilo-hours) as a function of time in kilo-hours for different a. dependence models of battery energy on environmental variation in the training phase (Fig.3) and b. reliability parameters in the trained phase (Fig.4).

Training phase: Fig. 3 shows the variation of the network energy as a function of the dependence model between the node energy and the environment variation model in the training phase. For this study, we use a reliability requirement of 0.5 for a network of 100 nodes. We see that when this dependence assumes the form of a quadratic polynomial, the rate of change of network energy is the highest compared to when the dependence is a constant. This sensitivity analysis to p illustrates the energy- conserving nature of the free-running trained phase where the BS optimizes the density of visible nodes to suit the reliability requirements. The rate of change of network energy approaches zero when the BS's prediction error matches the reliability requirement and the network enters

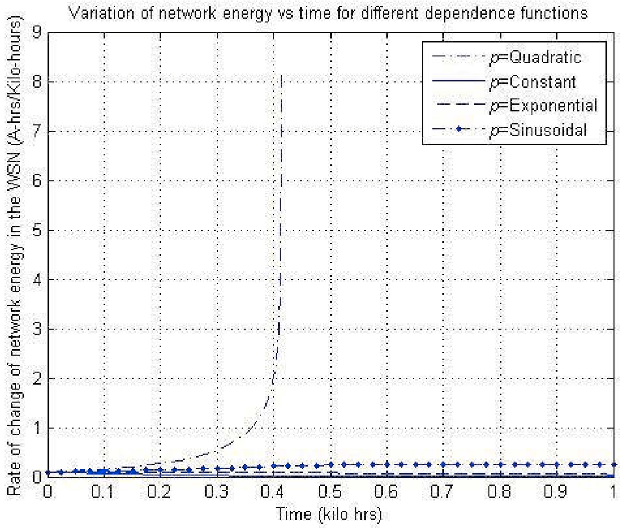

Fig. 3. **Training phase**:Rate of change of network energy as a function of dependence of node energy on environmental model in the training phase. The constant dependence results in lowest rate of change of network energy.

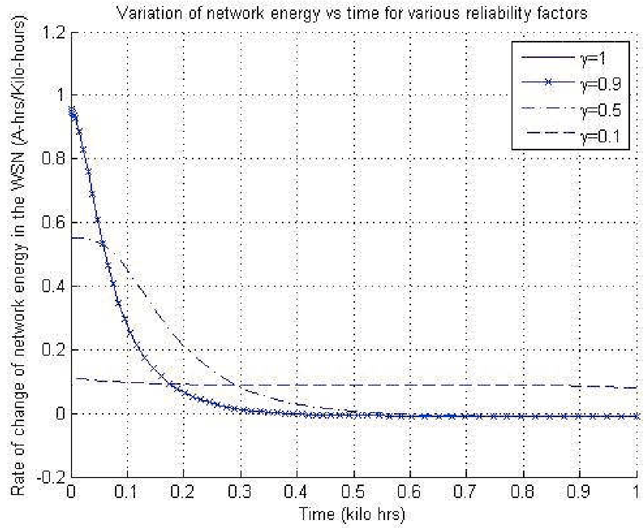

Fig. 4. **Trained phase** : Rate of change of network energy as a function of time for different reliability requirements. The reliability requirements dictate the density of visible nodes, which in turn affects the rate of change of network energy in the WSN. estimation.

the trained phase. A comparison of the rate of change of network energy in training (Fig. 3) and trained phase (Fig. 4) shows that this transition to the trained phase causes the gradient of the network energy variation to be less than that in the training phase.

Trained phase: In Fig. 4, we model the dependence function p, as an exponential function of the environment change parameter. The sensitivity analysis to reliability requirements shows that the rate of network energy varies with the desired reliability parameter γ. For high values of reliability of the sensing operation, the rate of change of network energy is higher. This is because, for higher ?, the BS increases the density of nodes in the visible mode, the density of nodes in the hidden-sense mode and also the rate of reporting, thus causing faster network energy depletion. We also model the case, where all the nodes are sensing, processing and transmitting for the entire duration of the deployment, i.e. all nodes are in the visible-transmit mode for γ=1. As seen from Fig.1, this reliability requirement causes higher variation of network lifetime than for lesser values of γ. For γ =0.1, which represents the case where the node energy has minimal dependence on the environment variation due to majority of the nodes always being in the hidden mode, the rate of change of network lifetime is much lower. For instance, for γ=1, the rate of change of network energy for the interval between the first 50 - 80 hours in the trained phase is higher by an order of magnitude than for γ =0.1.

X. Conclusions

The paper presented an analytical model to estimate the redundancy in a randomly deployed WSN without the use of location information. We illustrated the redundancy calculation approach and obtained the bounds on the error in estimating the redundancy for 1-redundancy in a randomly deployed WSN. We also showed the mutual dependence of network lifetime and redundancy in node deployment. These results can be used to design efficient sleep scheduling mechanisms to improve network lifetime by selectively controlling the density of 'awake' nodes in the deployment region. Further, the use of redundancy information can be employed to increase the reliability of sensing operation for regions of high-interest phenomenon that require larger number of sensors to accurately sense the environment. We also presented an energy model for wireless sensor networks by taking into account reliability requirements of the sensing operation and the impact of sensing environment variation on the rate of change of network energy. We presented an analytical framework for a NN- based (Boltzmann-learning rule) model to calculate the density of 'awake' nodes in the deployment region to satisfy the reliability requirements and accurately model the impact of environment variation on node energy. We observed that a higher reliability requirement and rapidly fluctuating sensing environments increased the rate of change of network energy. These results show the significance of sensitivity analysis of environment modeling on the lifetime of the WSN by creating sensing-environment and reliabilitycentered WSN topology. Our future work would involve the development of accurate prediction algorithms at the BS by modeling the system with a game-theoretic approach. The aim of the BS would be minimize the prediction error so that it would have to perform fewer computations to determine the set of awake nodes for the next cycle. Future work would also include improvement of this energy model by including lossy communication links in the deployment region.

REFERENCES

[1] P. Wan and C. Yi, "Coverage by randomly deployed wireless sensor networks," in *Proceedings of the IEEE International Symposium on Network Computing and Applications*, 2005, pp. 275–278.

[2] B. Wang, K. Chua, V. Srinivasan, and W. Wang, "Sensor density for complete information information coverage in wireless sensor networks," in *Proceedings of the European Workshop on Wireless Sensor Networks*, Zurich, 2006, pp. 69–82.

[3] S. Munirathnam, H. Kesavan, and P.H.Roe, "Probability density function estimation using the minmax measure," in *IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews*, ser. 1, vol. 30, Feb. 2000, pp. 77–83.

[4] O.Johnson, *Information Theory and the Central Limit Theorem*. London: Imperial College Press, 2004.

[5] Y. Gao, K. Wu, and F. Li, "Analysis of the redundancy of wireless sensor networks," in *Proceedings of the 2nd ACM International Workshop on Wireless Sensor Networks and Applications*, San Diego, USA, Sept. 2003.

[6] F.Ye, G. Zhang, S. Lu, and L. Zhang, "Peas: a robust energy conserving protocol for long-lived sensor networks," in *Proceedings of the 23rd International Conference on distributed Computing Systems*, Rhode Island, USA, May 2003, pp. 28–37.

[7] X. Wang, G. Xing, Y. Zhang, C. Lu, R. Pless, and C. Gill, "Integrated coverage and connectivity configuration in wireless sensor networks," in *Proceedings of the ACM Conference on Embedded Networked Sensor Systems*, Nov. 2003, pp. 28–39.

[8] D. Estrin and A. Cerpa, "Ascent: Adaptive self-configuring sensor networks topologies," in *IEEE Transactions on Mobile Computing*, ser. 3, vol. 3, July 2004, pp. 272–285.

[9] D. Gracanin, K. Adams, and M. Eltoweissy, "Data replication in collaborative sensor network systems," in *Proceedings of the 25th International Conference on Performance, Computing and Communications Conference*, Phoenix, USA, Apr. 2006.

[10] T. Bokareva, N. Bulusu, and S. Jha, "Sasha: Toward a self-healing hybrid sensor network architecture," in *Proceedings of the 2nd International Workshop on Embedded Networked Sensors*, Phoenix, USA, May 2005, pp. 71–78.

[11] F. Oldewurtel and P. Mahonen, "Neural wireless sensor networks," in *Proceedings of the International Conference on Systems ad Networks Communications*, Oct. 2006.

[12] H. Chen and A. F. Murray, "Continuous restricted boltzmann machine with an implementable training algorithm," in *IEEE Proceedings of Vision, Image and Signal Processing*, ser. 3, vol. 150, 2003, pp. 153–158.

[13] L.A.Rossi, B. Krishnamachari, and C.-C. J. Kuo, "Hybrid data and decision fusion techniques for model-based data gathering in wireless sensor neworks," in *Proceedings of the 60th Vehicular Technoogy Conference*, vol. 7, Sept. 2004, pp. 4616–4620.

[14] S. Haykin, *Neural Networks: A Comprehensive Foundation*. NJ: McMillan Publishing Company, 1994.

[15] H. Crauel and M. Gundlach, *Stochastic Dynamics*. Springer, 1999.

[16] L.C.Zhong, J. Rabaey, and A.Wolisz, "An integrated data link energy model for wireless sensor networks," in *Proceedings of the International Conference on Communications*, Paris, Jun. 2004.

[17] J.M.Cushing, *Differential Equations: An Applied Approach*. Upper Saddle River: Pearson Prentice Hall, 2004.