End-to-End Channel Capacity of a Wireless Sensor Network Under Reachback

Muhammad U. Ilyas, *Student Member, IEEE*, and Hayder Radha., *Member, IEEE*

Abstract—**Many *Wireless Sensor Network* applications will impose a many-to-one traffic flow pattern. This phenomena in communication networks is referred to as *reachback*. The higher volume of traffic in traffic hotspots, which usually form around the base station, limits the number of sensors that can participate in the formation of a Wireless Sensor Network with a fixed per sensor data rate. To alleviate the effects of reachback, spatial correlation of sensor data has enabled the use of Slepian-Wolf coding as a compression technique to reduce the volume of traffic in sensor networks. Within a cluster, Slepian-Wolf coding is used to compress the sensor readings. Clusterheads then transmit the compressed sensor readings over the *Overlay Network* which is formed by all clusterheads. We develop an expression for the end-to-end (sensor-to-base station) channel capacity of such a Wireless Sensor Network and observe the effect of variations in the number of clusterheads on end-to-end capacity.**

Index Terms—**End-to-End Channel Capacity, Reachback, Wireless Sensor Network.**

I. Introduction

OVER the past several years different assumptions have been made regarding the structure and capabilities of Wireless Sensor Networks (WSN) and the devices they are constituted of. In [1] Gupta and Kumar studied the scalability of a wireless networks with randomly chosen source destination pairs. Their conclusions offer two solutions to the scalability problem; 1) Design smaller networks or, 2) localize communication by clustering nodes. The idea of a WSN consisting of homogeneous devices gradually gave way to that of a network consisting of a homogeneous mix of nodes with non-uniform device capabilities. In this newly emerged view of WSNs, sensors are grouped into clusters. Each cluster of sensor nodes elects from among its members a clusterhead node (CLH) that acts as a gateway for all incoming and outgoing communications. Consequently, WSNs can be thought of as hierarchical networks with two levels. At this point we are not making any rigid assumptions about the networking technology or type of MAC protocol used.

Level 1, the lower level, refers to the network formed by a CLH and its associated sensors. Typically, all devices within a cluster are capable of communicating with each other directly. Therefore, communication between sensors and their CLH are assumed to take place over a single hop. We will use the terms cluster, intra-cluster or cluster-level communication interchangeably to refer to the exchange of messages between nodes belonging to the same cluster. Communication within a particular cluster proceeds at a common frequency, i.e. there is a potential for interference between transmissions of different sensors of a cluster. Different clusters may or may not use differing frequencies.

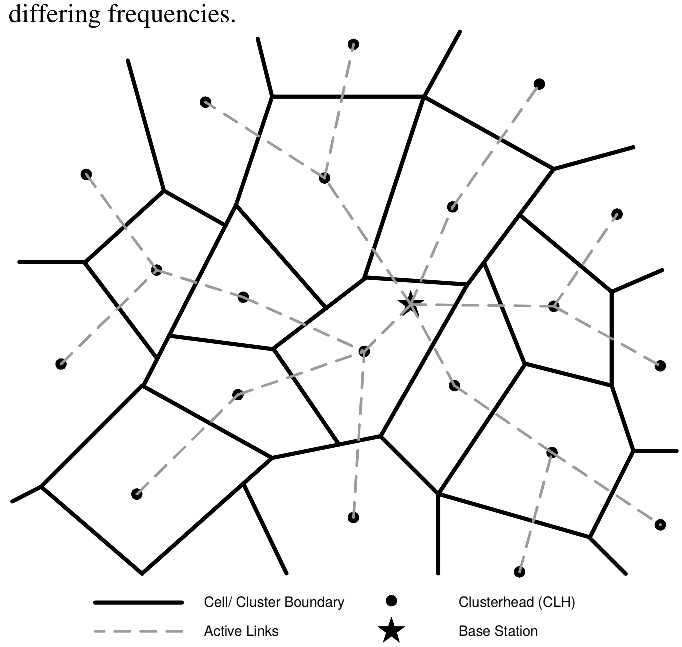

Figure 1. A tesselated or clustered wireless sensor network.

Level 2, the upper layer, refers to the network formed by the CLHs of all clusters in the WSN and the base station. We will also refer to this network as the *Overlay Network* (ON). We assume that the CLHs participating in the ON are capable of routing and relaying their own and other clusters' packets towards the base station. Moreover, since the most widely used WSN routing algorithms (DSR, AODV, Directed Diffusion) are different forms of shortest path routing algorithms, we assume that at any given time the routes from CLHs to base station in the ON form a tree rooted at the base station. We are discounting the possibility of using bifurcated routing, i.e. multiple paths from source to destination. We will use the terms overlay, overlay network or CLH-level communication interchangeably to refer to the exchange of messages between nodes belonging to the ON. Moreover, it is assumed that message exchanges between CLHs in the ON take place at one frequency, i.e. like sensors in a cluster, CLHs in the ON have the potential to produce interference for each other. However,

Manuscript received January 2, 2006.

M. U. Ilyas is with the Department of Electrical & Computer Engineering, Michigan State University, East Lansing, MI 48824 USA (phone: 517-355-3769; fax: 517-353-1980; e-mail: ilyasmuh@egr.msu.edu).

H. Radha is with the Department of Electrical & Computer Engineering, Michigan State University, East Lansing, MI 48824 USA (e-mail: radha@egr.msu.edu).

this frequency channel is assumed to be free of interference from cluster-level communication.

As Figure 1 shows, as traffic generated by the CLHs situated farther away approaches the base station, the expected volume of traffic carried by a link increases. This leads to a capacity bottleneck around the base station that subsequently limits the rate at which CLHs, and ultimately sensors, can inject data into the network. This is called the *reachback problem* [2]. To alleviate the effects of the reachback channel compression is used. A very popular compression method in WSNs is Slepian-Wolf coding [10] which will be discussed in more detail in section III.

Table 1. Tabular lising of relvant features of the three wireless networking technologies under consideration for use in Wireless Sensor Networks.

	IEEE 802.15.4	**IEEE 802.11b**	**IEEE 802.15.1/ Bluetooth**
Frequencies	~~868–868.6 MHz~~ ~~902-928 MHz~~ 2.4-2.4835 GHz	2.4-2.4835 GHz	2.4-2.4835 GHz
MAC type	1.TDMA in GTS 2. CSMA/CA in CTS	1.CSMA/CA in DCF 2. Polling in PCF	Polling

A. *Wireless Networking Standards for WSNs*

In our attempt to formulate a generalized expression for the end-to-end channel capacity of a channel between an arbitrary sensor and the base station we will have to remain open to the possibility of a number of different wireless networking technologies. Besides proprietary radio interfaces,the most commonly encountered standardized wireless networking technologies found in implementations of WSNs are the IEEE 802.11x Wireless Local Area Network (WLAN) standard [3], the IEEE 802.15.1/ Bluetooth standard [4] and the IEEE 802.15.4 Low Rate-Wireless Personal Area Network (LR-WPAN) standard [5]. The distinguishing features of these three networking standards are many. Table 1 lists only the features that we were concerned with in our work, i.e. the operating frequencies and the types of MAC protocols.

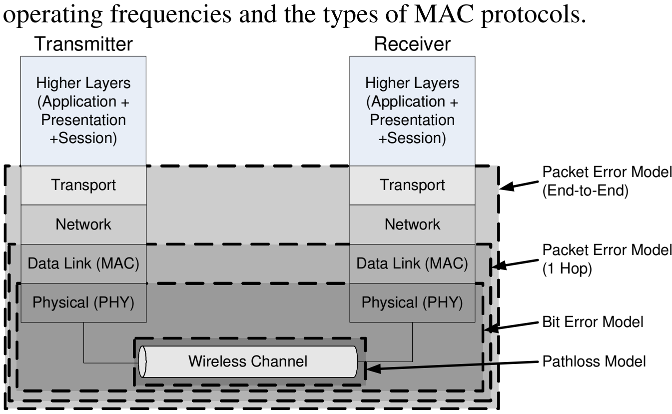

Figure 2. End-to-End channel between a transmitter and receiver on a multihop wireless network.

As we will show in a later section, our interpretation of how the end-to-end capacity of a wireless channel can be computed requires us to assume a pathloss model to model the physical channel. Since pathlosses are, besides numerous environmental factors, dependent on the frequency of the transmitted signal in addition to the distance between transmitter and receiver, we assume all transmissions to be happening in the 2.4–2.4835 GHz Industrial Scientific and Medical (ISM) band. This will allow us to use a single pathloss model that will be applicable to all three networking standards under consideration. Therefore, our end-to-end channel capacity expression will not be applicable to WSNs using IEEE 802.15.4 networks operating in the 868 – 868.6 MHz or 902 – 928 MHz bands. In addition to that, we are allowing for both collision and collision-free, CSMA/CA and TDMA, type MAC protocols.

II. End-to-End Channel

In this section we describe the general methodology of obtaining a channel model for the end-to-end channel between a transmitter and receiver communicating over a multihop wireless network.

The first step consists of identifying a suitable pathloss model for the wireless channel. For our purpose that means a pathloss model that is well suited for the 2.4–2.4835 GHz ISM band. We have identified the channel model proposed by the physical channel modeling subgroup of the IEEE 802.15 taskgroup 4 in [7] for this purpose. Figure 2 depicts a transmitter and receiver communicating over a multihop wireless channel. The transmitter and receiver are depicted by the two protocol stacks of the *Open Systems Interconnect* (OSI) model. At the Physical layer of the OSI model we assume a *Discrete Memoryless Channel* (DMC). The pathloss model is an abstraction of the physical channel between a single transmitter and receiver. In terms of the OSI model, the pathloss model represents an abstraction of everything that falls under the Physical layer. From the pathloss model we can determine a bit error model that will be representative of everything below the Data-link layer. The ultimate goal here is to determine a model that is capable of abstracting everything down from layer 5 of the OSI model. In the next step, we add another layer of abstraction to the bit error model by using it to obtain a packet error model. The packet error model will then abstract the network stack from the network layer down. Note that the packet error model can serve as an end-to-end model in single-hop networks. However, for multi-hop networks such as the ones we are considering we will have to modify the packet error model to obtain the desirable end-to-end packet error model that will abstract everything from the session layer down. This successive abstraction of the channel between transmitter and receiver based on the next lower model is depicted in Figure 2. The layers of the OSI model inside a dashed box represent the layers encapsulated in the corresponding model.

III. Brief Overview of Slepian-Wolf Coding

Slepian-Wolf coding was first proposed by Slepian and Wolf in [10]. However, we make some simplifying assumptions about Slepian-Wolf coding as it applies to WSNs. These assumptions have been taken from Marco and Neuhoff in [8] and [9]. Slepian-Wolf coding exploits the spatial correlation of sensor readings in neighboring sensors and is employed as a means to compress the data before transmitting it to the base station. In our case, Slepian-Wolf coding is employed by the CLH in intra-cluster communication in the collection of data from its associated sensors. Slepian-Wolf coding is based on the equation (1). Consider a cluster i as in Figure 3 consisting of N_i sensors and a CLH in which the first sensor $n_i(1)$ produces reading $X_{n_i(1)}$, the second sensor $n_i(2)$ produces reading $X_{n_i(2)}$ and so on.

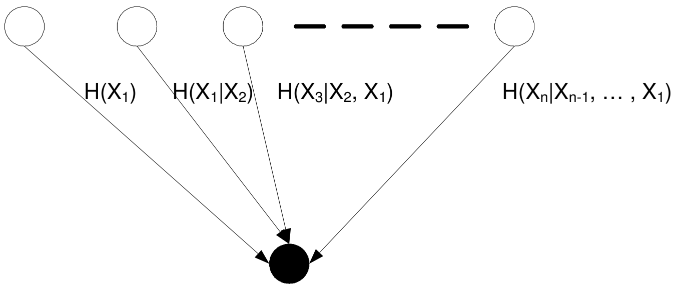

Figure 3. Slepian-Wolf coding in cluster-level communication.

$$H\left(X_{n_i(1)}\right)+H\left(X_{n_i(2)} \mid X_{n_i(1)}\right)+\ldots+H\left(X_{n_i(N_i)} \mid X_{n_i(N_i-1)},\ldots,X_{n_i(2)},X_{n_i(1)}\right) = H\left(X_{n_i(1)},X_{n_i(2)},\ldots,X_{n_i(N_i)}\right) \quad (1)$$

The size of all N_i sensor readings after lossless compression is lower-bounded by their joint entropy. Slepian-Wolf coding within a cluster proceeds as follows.

1. The CLH collects readings from sensors in *rounds*, compresses and transmits them to the base station.
2. In every round, sensors transmit their readings to their CLH according to a pre-defined sequence or schedule. For simplicity's sake, let us assume that the transmissions to the CLH are scheduled in the ascending order of sensor nodes' ID numbers.
3. The first transmission $X_{n_i(1)}$ by $n_i(1)$ is not compressed (see first term on left-hand side of (1)).
4. The second sensor $n_i(2)$ compresses its sensor reading to $H\left(X_{n_i(2)} \mid X_{n_i(1)}\right)$ based on the side information of $n_i(1)$'s transmission of size $H\left(X_{n_i(1)}\right)$. Therefore, the j-th node in cluster i transmits its data as $H\left(X_{n_i(j)} \mid X_{n_i(j-1)},\ldots,X_{n_i(2)},X_{n_i(1)}\right)$ bits. This is depicted in Figure 3.
5. This way the total volume of all transmissions approaches the joint entropy as shown in (1).

However, this coding scheme has one major disadvantage. Failure of the CLH to receive the k-th transmission results in its inability to reconstruct all subsequent transmissions $k+1$ through N_i for that round.

IV. Notation and Convention

In this section we define the notation and convention we have used for different parameters of the generalized WSN under consideration.

Let N be the total number of sensors in a WSN of M clusters. Let $C_i \quad \forall i \in \{1,2,3,\ldots,M\}$ denote an individual cluster consisting of one CLH and N_i sensors (therefore; $N=\sum_{i=1}^{M} N_i$). Sensors are addressed using a 2-dimensional coordinate system, the i-th cluster's j-th sensor node is referred to as $n_i(j) \quad \forall i=\{1,2,3,\ldots,M\} \quad \forall j=\{1,2,3,\ldots,N_i\}$, with $n_i(0)$ denoting its CLH, and $n_0(0)$ denoting the base station. We also define a function $f\left(n_i(j)\right)$ that returns the frequency at which the node or CLH provided in the argument communicates. This way $f\left(n_i(0)\right)$ denotes the intra-cluster communication frequency of cluster C_i. We use $f_i \quad \forall i \in \{1,2,3,\ldots,M\}$ as a shortened form to denote the cluster frequencies of all M clusters. Similarly, we abbreviate $f\left(n_0(0)\right)$ by f_0 to denote the frequency used by CLHs for communicating in the ON. The probability of $n_k(l)$ making an interfering transmission at the same time as $n_i(j)$ is making its transmission is denoted by $p_{n_k(l)}\left(n_i(j)\right)$. We also define a *Frequency Indicator Function*, $I_f\left(n_i(j),n_k(l)\right) = \begin{cases} 0 & if \quad f\left(n_i(j)\right) \neq f\left(n_k(l)\right) \\ 1 & if \quad f\left(n_i(j)\right) = f\left(n_k(l)\right) \end{cases}$. New notation will be introduced in the following sections where needed as we go on.

V. Cluster Communication

From the pathloss model in [7] we can obtain the expression in equation (2) to obtain the interference power $P_{n_k(l)}\left(n_i(j)\right)$ of a signal transmitted by $n_k(l)$ at receiver $n_i(j)$.

$$P_{n_k(l)}\left(n_i(j)\right) = I_f\left(f\left(n_i(j)\right), f\left(n_k(l)\right)\right) \cdot K_0 \cdot P_{TX-amp} \cdot \eta_{TX-ant} \cdot \eta_{RX-ant} \times \frac{PL_0}{\left(d\left(n_i(j),n_k(l)\right)/d_0\right)^2 \cdot \left(f/f_c\right)^{2\mathrm{K}+2}} \quad (2)$$

Here, P_{TX-amp} (typically 1000 mW for IEEE 802.15.4 compliant devices) is the signal power at the transmitter after amplification before it is passed to the antenna, η_{TX-ant} is the transmitter antenna efficiency, η_{RX-ant} is the receiver antenna efficiency and K, PL_0, and K_0 are environmental parameters that depends on the operating environment (see [7]). Since we are assuming that a WSN consists of devices with identical radio interfaces transmitting at a fixed power we consider these terms to be constant across the entire network. We also define two reference parameters f_c and d_0 for this pathloss model. For the sets of parameters provided in [7], f_c is set to 5 GHz and d_0 is set to 1 m. This leaves the expression dependent on the transmission frequency f and the distance between sensor nodes $n_i(j)$ and $n_k(l)$ that is returned by the distance function $d\left(n_i(j),n_k(l)\right)$.

In order to obtain a bit error model for a DMC we need an expression for the *Signal-to-Interference & Noise Ratio* (SINR). The general expression for the SINR is given in (3). Using the expression for the pathloss model described above we can arrive at an expression for the SINR in the terms that we have previously defined.

$$SINR = \frac{P_{TX}}{P_A + \sum P_{\mathrm{int}}} \quad (3)$$

where P_{TX} is the power of the received signal for which the SINR is being computed and P_{int} is the power of interfering signals caused by undesired concurrent transmissions elsewhere in the network. P_A is the ambient noise power of interference produced by sources that are not part of the WSN and are not modeled by any terms in P_{int}. Sources of ambient noise power may include but are not limited to other networks that are co-located with the WSN under consideration or devices or appliances (microwave ovens) that operate in the same frequency band. If we substitute the pathloss model from (2) in the SINR expression in (3) we obtain (4). Since at the cluster-level all transmissions are directed from sensors to their respective CLH, the term $SINR(n_i(j))$ represents the SINR of the transmission from $n_i(j)$ to its CLH $n_i(0)$. Note that we have omitted the indicator function term and the probability of simultaneous transmission terms from the numerator. Since both sensor $n_i(j)$ and CLH $n_i(0)$ are operating at the same frequency, $I(n_i(j), n_i(0)) = 1$. In the general case, all sensors and CLHs can be considered as potential sources of interference. Obviously, we assign $p_{n_i(j)}(n_i(j)) = 0$.

$$SINR(n_i(j)) = \frac{K_0 \cdot P_{TX-amp} \cdot \eta_{TX-ant} \cdot \eta_{RX-ant} \cdot \dfrac{PL_0}{\left(d(n_i(j), n_i(0)) / d_0\right)^2 \cdot \left(f / f_c\right)^{2K+2}}}{P_A + \sum_{k=1}^{M} \sum_{l=0}^{N_k} I_f\left(f(n_i(j)), f(n_k(l))\right) \cdot p_{n_k(l)}(n_i(j)) \cdot K_0 \cdot P_{TX-amp} \cdot \eta_{TX-ant} \cdot \eta_{RX-ant} \cdot \dfrac{PL_0}{\left(d(n_k(l), n_i(0)) / d_0\right)^2 \cdot \left(f / f_c\right)^{2K+2}}} \tag{4}$$

We simplify (4) by removing all constants from thenumerator and defining a new term $P_A{}'$ in place of P_A. If,

$$P_A{}' = \frac{P_A \cdot \left(f / f_c\right)^{2K+2}}{K_0 \cdot P_{TX-amp} \cdot \eta_{TX-ant} \cdot \eta_{RX-ant} \cdot PL_0}, \text{ then}$$

$$SINR(n_i(j)) = \frac{1}{\left(d(n_i(j), n_i(0)) / d_0\right)^2} \times \frac{1}{P_A{}' + \sum_{k=1}^{M} \sum_{l=0}^{N_k} I_f\left(f(n_i(j)), f(n_k(l))\right) \cdot p_{n_k(l)}(n_i(j)) \cdot \dfrac{1}{\left(d(n_k(l), n_i(0)) / d_0\right)^2}} \tag{5}$$

Then (5) is the final expression for $SINR(n_i(j))$. In the next few steps we derive the bit error model. For that we use the *Lognormal Shadow Fading Model* described by Rappaport in [11]. If $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{u}{2}} du$, then the probability of receiving a bit error at $n_i(0)$ in a transmission originating at $n_i(j)$ is called the *Bit Error Rate* (BER) $P_{BER}(n_i(j))$ and is described by equation (6). This can be thought of as the probability of error in a Binary Symmetric Channel (BSC). The corresponding channel capacity in terms of the BER is obtained from equation (7), where $H_b(\cdot)$ is a function that returns the entropy of a Bernoulli random variable with the parameter provided in the argument.

$$P_{BER}(n_i(j)) = Q\left(\sqrt{2 \cdot SINR(n_i(j))}\right) = \frac{1}{\sqrt{2\pi}} \int_{\sqrt{2 \cdot SINR(n_i(j))}}^{\infty} e^{-\frac{u}{2}} du \tag{6}$$

$$C_{BSC}(n_i(j)) = 1 - H_b\left(P_{BER}(n_i(j))\right) \tag{7}$$

From the BER we now determine expression (8) for obtaining the probability of a packet loss or the *Packet Error Rate* (PER) $P_{PER}(n_i(j))$ for a transmission from $n_i(j)$ to $n_i(0)$. The PER corresponds to the probability of error of a Binary Erasure Channel (BEC). We assume that a received packet is discarded if a single bit is in error, a valid assumption considering our choice of wireless standards. The term in the exponent represents the packet length which is the number of bits in the header h and the number of bits of the payload $\left\lceil H\left(X_{n_i(j)} \mid X_{n_i(1)}, X_{n_i(2)}, \ldots, X_{n_i(j-1)}\right)\right\rceil$. From (8) we trivially obtain expression (9) for the channel capacity $C_{PER}(n_i(j))$ in terms of $P_{PER}(n_i(j))$.

$$P_{PER}(n_i(j)) = 1 - \left(1 - P_{BER}(n_i(j))\right)^{h + \left\lceil H\left(X_{n_i(j)} \mid X_{n_i(1)}, X_{n_i(2)}, \ldots, X_{n_i(j-1)}\right)\right\rceil} \tag{8}$$

$$C_{PER}(n_i(j)) = \left(1 - P_{PER}(n_i(j))\right) \cdot \prod_{k=1}^{j} \left(1 - P_{PER}(n_i(k))\right) \tag{9}$$

VI. Overlay Network Communication

In this section we now turn our attention to the end-to-end capacity of the channel between an arbitrary CLH and the base station communicating over a multi-hop wireless network. The case in which all CLHs are directly communicating with the base station in the ON becomes a special case of the more general case of a multi-hop ON. We will consider two types of communication in the ON; 1) Without Slepian-Wolf Coding in ON in which CLHs simply forward packets from downstream CLHs , and 2) With Slepian-Wolf coding in ON in which intermediate CLHs apply a second level of compression to their own packets by recoding them based on the downstream packets they have received and forwarded. In this case, a loss of a single downstream packet that is used for compressing data causes the inability to perform Slepian-Wolf coding and causes a packet loss for the current round. Alternatively, the loss of a packet results in a packet error for all upstream CLHs in the ON. As before in the case of the Cluster-level communication, we start from the expression for SINR, this time for a signal transmitted by a CLH $n_i(0)$ to its upstream neighbor. This is given in (10).

Before proceeding further we define a set of new functions that will subsequently be used in this section. $R^{1\uparrow}(n_i(0))$ returns the immediate upstream neighbor of CLH $n_i(0)$, where upstream denotes the direction towards the base station in the network topology. $R^{1\downarrow}(n_i(0))$ returns the set of CLHs that is 1 hop downstream from $n_i(0)$, where the term downstream refers to the direction away from the base station in the network topology. $R^{\uparrow}(n_i(0))$ returns the set of all CLHs that are upstream from $n_i(0)$ and $R^{\downarrow}(n_i(0))$ returns the set of all CLHs that are downstream from $n_i(0)$.

$$SINR_{ON}\left(n_i(0)\right)=\frac{K_0\cdot P_{ON-TX-amp}\cdot\eta_{TX-ant}\cdot\eta_{RX-ant}\cdot\dfrac{PL_0}{\left(d\left(n_i(0),R^{1\uparrow}\left(n_i(0)\right)\right)/d_0\right)^2\cdot\left(f/f_c\right)^{2K+2}}}{P_{ON-A}+\sum_{\substack{k=1\\k\neq i}}^{M}p_{n_k(0)}\left(n_i(0)\right)\cdot K_0\cdot P_{ON-TX-amp}\cdot\eta_{TX-ant}\cdot\eta_{RX-ant}\cdot\dfrac{PL_0}{\left(d\left(n_i(0),n_k(0)\right)/d_0\right)^2\cdot\left(f/f_c\right)^{2K+2}}} \quad (10)$$

If,

$$P_{ON-A}'=\frac{P_{ON-A}\cdot\left(f/f_c\right)^{2K+2}}{K_0\cdot P_{ON-TX-amp}\cdot\eta_{TX-ant}\cdot\eta_{RX-ant}\cdot PL_0}$$

$$SINR_{ON}\left(n_i(0)\right)=\frac{1}{\left(d\left(n_i(0),R^{1\uparrow}\left(n_i(0)\right)\right)/d_0\right)^2}\times\frac{1}{P_{ON-A}'+\sum_{\substack{k=1\\k\neq i}}^{M}\dfrac{p_{n_k(0)}\left(n_i(0)\right)}{\left(d\left(n_i(0),n_k(0)\right)/d_0\right)^2}} \quad (11)$$

Note that for a TDMA protocol in the ON $p_{n_k(0)}\left(n_i(0)\right)=0$, and hence (11) simplifies to (12).

$$SINR_{ON}\left(n_i(0)\right)=\frac{1}{P_{ON-A}'\left(d\left(n_i(0),R^{1\uparrow}\left(n_i(0)\right)\right)/d_0\right)^2} \quad (12)$$

Applying the Lognormal Shadow Fading Model leads us to a similar expression (13) for the BER as before.

$$P_{ON-BER}\left(n_i(0)\right)=Q\left(\sqrt{2\cdot SINR_{ON}\left(n_i(0)\right)}\right)=\frac{1}{\sqrt{2\pi}}\int_{\sqrt{2\cdot SINR_{ON}\left(n_i(0)\right)}}^{\infty}e^{-\frac{u}{2}}du \quad (13)$$

From (13) we can obtain a recursive definition for the BER of the multi-hop, end-to-end channel between a CLH and the base station.

$$P_{ON-BER-E2E}\left(n_i(0)\right)=P_{ON-BER}\left(n_i(0)\right)\cdot\left[1-P_{ON-BER-E2E}\left(R^{1\uparrow}\left(n_i(0)\right)\right)\right]+P_{ON-BER-E2E}\left(R^{1\uparrow}\left(n_i(0)\right)\right)\cdot\left[1-P_{ON-BER}\left(n_i(0)\right)\right] \quad (14)$$

The expression for the PER for a packet originating at $n_i(0)$ on a link between $n_k(0)$ and its upstream neighbor $R^{1\uparrow}\left(n_k(0)\right)$ is then provided by equation (15).

$$P_{ON-PER-n_k(0)}\left(n_i(0)\right)=1-\left(1-P_{ON-BER}\left(n_k(0)\right)\right)^{h+\left[H\left(X_{n_{i}(1)},X_{n_i(2)},\ldots,X_{n_i(N_i)}\right)\right]} \quad (15)$$

This leads us to the final expression (16) for the end-to-end PER from an arbitrary CLH $n_i(0)$ to the base station. This term is simply obtained by subtracting from 1 the product of the probabilities of a successful transmission on all links along the route from source $n_i(0)$ to the base station. Equation (17) gives the corresponding expression for the end-to-end capacity.

$$P_{ON-PER-E2E}\left(n_i(0)\right)=1-\prod_{n_k(0)\in R^{\uparrow}\left(n_i(0)\right)}\left(1-P_{ON-PER-n_k(0)}\left(n_i(0)\right)\right) \quad (16)$$

$$C_{ON-PER-E2E}\left(n_i(0)\right)=1-P_{ON-PER-E2E}\left(n_i(0)\right) \quad (17)$$

Now we come to the second case in which we assume a second level of Slepian–Wolf coding based on downstream packets. The derivation for the capacity expression for this case proceeds exactly along the same lines as in the first case between equations (10) and (13). The expression for $P_{ON-PER-n_k(0)}\left(n_i(0)\right)$, the packet error rate of the link from $n_k(0)$ to $R^{1\uparrow}\left(n_k(0)\right)$ for a packet originated at CLH $n_i(0)$, is given by equation (18). The exponential term is further complicated by the fact that the joint entropy is replaced by the conditional entropy conditioned on all sensor readings originated at $R^{\downarrow}\left(n_i(0)\right)$.

$$P_{ON-PER-n_k(0)}\left(n_i(0)\right)=1-\left[\left(1-P_{ON-BER}\left(n_i(0)\right)\right)^{h+\left[H\left(X_{n_i(1)},X_{n_i(2)},\ldots,X_{n_i(N_i)}\mid X_{n_l(1)},X_{n_l(2)},\ldots,X_{n_l(N_i)}\right)\right]}\right]\times\left[\prod_{n_k(0)\in R^{\downarrow}\left(n_i(0)\right)}\prod_{n_j(0)\in R^{\downarrow}\left(n_i(0)\right)}\left(1-P_{ON-PER-n_k(0)}\left(n_j(0)\right)\right)\right] \quad (18)$$

where $P_{ON-PER-n_k(0)}\left(n_j(0)\right)=1$ if $n_k(0)\notin R^{\uparrow}\left(n_j(0)\right)$. This leads to the expression for the end-to-end channel capacity between $n_i(0)$ and the base station in equation (19).

$$P_{ON-PER-E2E}\left(n_i(0)\right)=1-\prod_{n_k(0)\in R^{\uparrow}\left(n_i(0)\right)}\left(1-P_{ON-PER-n_k(0)}\left(n_i(0)\right)\right) \quad (19)$$

$$C_{ON-PER-E2E}\left(n_i(0)\right)=1-P_{ON-PER-E2E}\left(n_i(0)\right) \quad (20)$$

VII. Sensor-to-Base Station Capacity

Using the above results the sensor-to-base station BER and PER for any arbitrary sensor $n_i(j)$ can be obtained by multiplying the cluster-level BER $P_{BER}\left(n_i(j)\right)$ in (6) or PER $P_{PER}\left(n_i(j)\right)$ in (8) with the end-to-end BER $P_{ON-BER-E2E}\left(n_i(0)\right)$ in (14) or PER $P_{ON-PER-E2E}\left(n_i(0)\right)$ in equation (19), respectively. This yields an expression for the sensor-to-base station BER in equation (21) and a corresponding sensor-to-base station PER expression in equation (22). From (21) and (22) we can obtain the corresponding capacity expressions in equations (23) and (24)

$$P_{BER-S2BS}\left(n_i(j)\right)=P_{ON-BER-E2E}\left(n_i(0)\right)\cdot\left[1-P_{BER}\left(n_i(j)\right)\right]+P_{BER}\left(n_i(j)\right)\cdot\left[1-P_{ON-BER-E2E}\left(n_i(0)\right)\right] \quad (21)$$

$$P_{PER-S2BS}\left(n_i(j)\right)=1-\left[1-P_{PER}\left(n_i(j)\right)\right]\cdot\left[1-P_{ON-PER-E2E}\left(n_i(0)\right)\right] \quad (22)$$

$$C_{BER-S2BS}\left(n_i(j)\right)=1-H_b\left(P_{BER-S2BS}\left(n_i(j)\right)\right) \quad (23)$$

$$C_{PER-S2BS}\left(n_i(j)\right)=1-P_{PER-S2BS}\left(n_i(j)\right)=\left[1-P_{PER}\left(n_i(j)\right)\right]\cdot\left[1-P_{ON-PER-E2E}\left(n_i(0)\right)\right] \quad (24)$$

VIII. Results

For the following experiments we have used the channel model for residential environments in [7]. Initially we consider a WSN consisting of $N=100$ sensor nodes and $M=5$ CLHs randomly placed over a square region of dimensions $10m\times 10m$ according to a uniform random distribution. To create routes with greater number of hops in the overlay network we place the base station at coordinates $(0,0)$. We assume a set of 15 available frequencies for cluster-level communication in addition to one frequency reserved for communication between CLHs in the ON. Furthermore, we assume the IEEE 802.15.4 MAC protocol using TDMA/ GTS enabled mode for transmission of data frames. Figure 4 depicts a WSN with the above specified parameters. Circles denote the positions of CLHs while the x-marks denote the positions of sensors. Sensors in closest proximity of the means obtained by the k-means clustering algorithm [12] are assigned the role of the CLH for that cluster. Solid lines depict the topology of the ON while broken lines indicate each sensor's association with its respective CLH. Moreover, for reasons of brevity we only present the case in which Slepian-Wolf coding is used in cluster-level communication only. Applying the procedure described in the preceding sections to the WSN in Figure 4 we obtain $P_{BER-S2BS}\left(n_i\left(j\right)\right)$, $P_{PER-S2BS}\left(n_i\left(j\right)\right)$, $C_{BER-S2BS}\left(n_i\left(j\right)\right)$ and $C_{PER-S2BS}\left(n_i\left(j\right)\right)$ which are plotted in Figure 5. Since we are assuming that a single bit error in a packet at any hop causes a packet drop, the throughput/ goodput rate computed from the PER is expected to be lower than for the BER for all sensors. This implies that the PER is higher than the BER and is reflected in Figure 5.a. Similarly, the capacity in Figure 5.b in terms of PER is lower than the capacity in terms of BER. Finally, we investigate the effect of the number of clusters on the error probabilities. We reduce N to 50 while varying the number of clusters from 2 to 25 and generate 5 results for each set of parameters. Figure 6 shows the corresponding error probability and capacity plots. As we can see, as the number of clusters increases, contention decreases causing an expected decrease in the probability of error.

IX. Conclusions

We derive expressions for the probability of error and end-to-end channel capacity for a WSN modeled as 1) a cascade of BSCs and 2) a cascade of BECs. The application of these models to a number of randomly generated WSNs shows that the capacity of the BEC equivalent is lower than the capacity of the BSC based model, (as dictated by the data processing inequality).

References

[1] P. Gupta, and P. R. Kumar, "The Capacity of Wireless Networks," *IEEE Transactions on Information Theory*, Vol. 46, No. 2, March 2000.

[2] J. Barros, S. D. Servetto, "On the Capacity of the Reachback Channel in Wireless Sensor Networks," *IEEE Workshop on Multimedia Signal Processing*, December 2002.

[3] "ANSI/IEEE Std 802.11, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," 1999.

[4] "IEEE Std 802.15.1, Part 15.1: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Wireless Personal Area Networks (WPANs)," June 14, 2002.

[5] "IEEE P802.15.4/D18, Draft Standard: Low Rate Wireless Personal Area Networks," February 2003.

[6] Soo Young Shin, Hong Seong Park, Sunhyun Choi, Wook Hyun Kwon, "Packet Error Rate Analysis of IEEE 802.15.4 under IEEE 802.11b Interference," *International Conference on Wired/ Wireless Internet Communications 2005 (WWIC'05)*, Xanthi, Greece, May 11-13, 2005.

[7] Andreas F. Molisch, Kannan Balakrishnan, Chia-Chin Chong, Shahriar Emami, Andrew Fort, Johan Karedal, Juergen Kunisch, Hans Schantz, Ulrich Schuster, Kai Siwiak, "IEEE 802.15.4a channel model - final report," 2004.

[8] D. Marco, and D. L. Neuhoff, "Reliability vs. Efficiency in Distributed Source Coding for Field-Gathering Sensor Networks," *IEEE International Conference on Information Processing in Sensor Networks (IPSN'04)*, Berkeley, CA, 2004.

[9] D. Marco, E. Duarte-Melo,M. Liu, D. L. Neuhoff, "On the Many-to-One Transport Capacity of a Dense Wireless Sensor Network and the Compressibility of Its Data," *International Workshop on Information Processing in Sensor Networks (IPSN'03)*, Palo Alto, CA. Lecture Notes in Computer Science vol. 2634, Springer: New York, April 2003.

[10] David Slepian, and Jack K. Wolf, "Noiseless Coding of Correlated Information Sources," *IEEE Transactions on Information Theory*, Vol. IT-19, No. 4, July 1973.

[11] T.S. Rappaport, "Wireless Communications – Principles and Practice, 2nd ed," Pearson Education, Singapore, 2002.

[12] J. Hartigan and M. Wong, "A k-means clustering algorithm," Appl. Statist. 28 (1979), pp. 100–108.

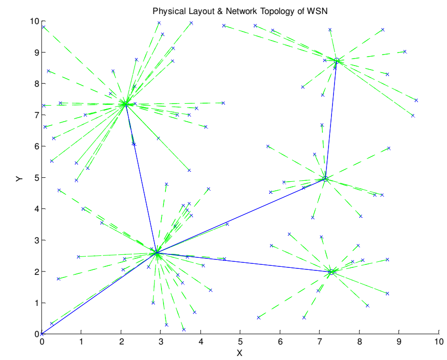

Figure 4. Physical layout & network topology of WSN

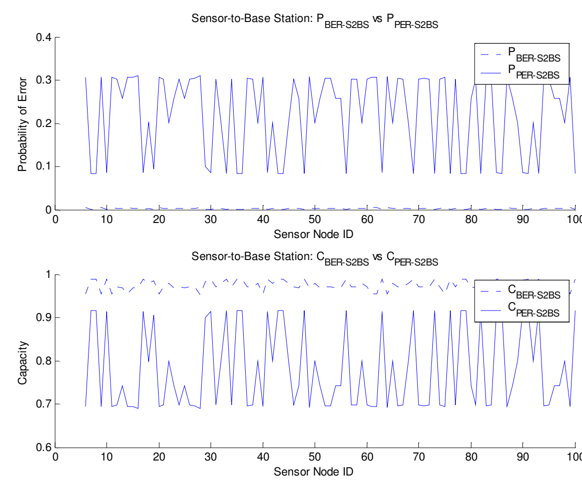

Figure 5. a) Sensor-to-Base Station Probability of Error for BER and PER, b) Sensor-to-Base Station channel capacity for BER and PER.

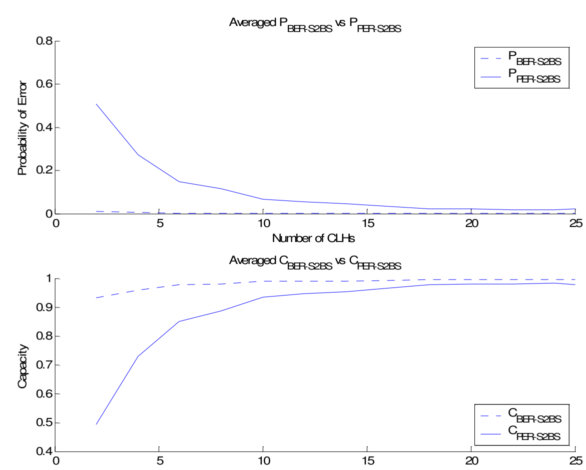

Figure 6. a) Sensor-to-base station channel capacity against number of clusters, averaged over 20 simulations.