Distance Based Thresholds for Cluster Head Selection in Wireless Sensor Networks

Sang H. Kang, *Senior Member, IEEE,* and Thinh Nguyen, *Member, IEEE*

***Abstract*—Central to the cluster-based routing protocols is the cluster head (CH) selection procedure that allows even distribution of energy consumption among the sensors, and therefore prolonging the lifespan of a sensor network. We propose a distributed CH selection algorithm that takes into account the distances from sensors to a base station that optimally balances the energy consumption among the sensors. NS-2 simulations show that our proposed scheme outperforms existing algorithms in terms of the average node lifespan and the time to first node death.**

***Index Terms*—Wireless sensor network, cluster, energy consumption, distance to base station.**

I. INTRODUCTION

WIRELESS Sensor Networks (WSN) are often characterized by unattended self-organization of non-rechargeable nodes. To prolong the network lifespan of WSNs, cluster-based routing protocols are often used to obtain load balance and to reduce communication volume in a distributed manner. Low Energy Adaptive Clustering Hierarchy (LEACH) [1] is one of the most promising cluster-based routing protocols.

LEACH divides the network into several clusters of nodes. Through a cluster head (CH) selection procedure, each node in the network has an equal chance of becoming a CH overall. Each CH gathers and processes data from its members, then forwards the aggregate data to the base station (BS). In this way, LEACH attempts to balance the energy consumption among all the nodes. However, if the BS is far away from the sensor field, the energy expense for the CH to send data to the BS increases according to the 4-th power of its distance to the BS [1]. As such, even though all the nodes have an equal chance of becoming a CH, the ones far away from the BS will run out of energy before the closer ones. Therefore, in order to further balance the energy consumption, the probability of a node becoming a CH should be computed based on its distance to the BS.

Recently, several distance-based CH selection algorithms have been proposed. In [2], a CH election algorithm is proposed using the minimum and maximum of the distance to the BS. In [3], the authors investigate inner-CH multi-hop routing using Dijkstra's algorithm. In [4], the relative distance to the BS is considered with a weighting factor. In [5], far zone heads are adopted for multi-hop routing; but the interference between the cluster and its far zone needs to be addressed. Note that all the aforementioned approaches are heuristic based on the intuition that the far away nodes should communicate relatively less than the closer ones.

Rather than using a heuristic approach, we propose a distributed LEACH-based CH selection algorithm in which nodes are self-selected to become CHs with *different* probabilities based on their distances to the BS, in such a way that the energy consumption among the nodes are balanced. This is in contrast with the original LEACH protocol where all the nodes are self-selected to become CHs with a constant pre-designed probability, and thus is suboptimal. Next, we briefly describe the LEACH protocol.

II. THE LEACH ROUTING PROTOCOL

Assume there are N sensor nodes deployed randomly in a square of $M \times M$ [m^2]. All nodes start with an equal amount of energy. Each node is assumed to perform power control depending on the distance between the transmitter and receiver. LEACH adopts the following energy model: A node transmitting an l-bit message over a distance d [m] will dissipate an energy amount $E_T(l, d)$

$$E_T(l,d) = \begin{cases} l(E_e + \epsilon_f d^2), & \text{if } d < \delta \\ l(E_e + \epsilon_m d^4), & \text{if } d \geq \delta, \end{cases} \tag{1}$$

where E_e [J/bit] represents the energy being dissipated to operate the circuitry per bit, ϵ_f [J/bit/m^2] and ϵ_m [J/bit/m^4] denote the factors in Friss' free space model and the typical multi-path model, respectively, and $\delta = \sqrt{\epsilon_f/\epsilon_m}$. The energy dissipation for receiving an l-bit message is determined by

$$E_R(l) = lE_e. \tag{2}$$

When a node listens to the medium for t [sec], it's energy consumption is modeled as tE_L, where energy dissipation per unit time, E_L [Joules/sec], is assumed to be constant for the sake of simplicity.

In LEACH, time is divided into rounds, $r = 0, 1, 2, \cdots$. The number of CHs in each round is a random variable with expectation k, which is a pre-calculated value as a system parameter. Let $p = k/N$ be the desired percentage of CHs $(0 \leq p \leq 1)$. Let us use $[\cdot]$ to denote rounding off throughout this letter. LEACH groups $[1/p]$ successive rounds together to form a group of rounds (GOR). Let $G(r)$ be the set of nodes that have not been CHs within the GOR including the current

Manuscript received March 2, 2012. The associate editor coordinating the review of this letter and approving it for publication was I.-R. Chen.

S. H. Kang was supported by the Basic Science Program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (No. 2011-0026363). T. Nguyen was supported by NSF (0834775) and NSF-CAREER (845476).

S. H. Kang is with the Department of Electrical and Computer Engineering, University of Seoul, Seoul, Korea (e-mail: shkang@uos.ac.kr).

T. Nguyen is with the School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, Oregon, USA (e-mail: thinhq@eecs.oregonstate.edu).

Digital Object Identifier 10.1109/LCOMM.2012.073112.120450

round r. In round r, node i ($i = 1, 2, \cdots, N$) elects itself to become a CH with probability:

$$T_{i,r} = \begin{cases} \frac{p}{1-p\cdot(r \mod [\frac{1}{p}])}, & i \in G(r) \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

Each round begins with the set-up phase which consists of the following sub-phases I and II, which operate sequentially during the pre-defined periods τ_1 and τ_2, respectively. In sub-phase I, CHs are selected using a distributed randomized procedure according to (3). Each node that has elected itself to be a CH broadcasts an advertisement (ADV) message using CSMA as a MAC protocol. The ADV message is a 2-tuple message, (ID, Code), containing the node identification (ID) of the CH and a spreading code to avoid inter-cluster interference. Then all nodes simultaneously enter sub-phase II, in which each non-CH node chooses a nearest CH, goes to the sleep mode, and then wakes up at a random time to send a join-request (JOIN) message to the nearest CH with CSMA using the relevant spreading code and goes to the sleep mode again.

After the set-up phase, CHs simultaneously broadcast TDMA scheduling messages using spreading codes to start the steady-state phase, which involves data signal sending from non-CH nodes to the CH nodes by TDMA scheduling, aggregation/beamforming, and routing to the BS. The maximum duration of the steady-state phase is denoted by τ_3. Note that, in LEACH, ADV, TDMA and JOIN messages are transmitted with a fixed power corresponding to a distance of $\sqrt{2}M$ [m]. Refer to [1] for further details for LEACH.

III. Energy Consumption Analysis

In this section, we analyze the average amount of energy consumption, $\overline{E}_{\text{CH}}(d)$ and $\overline{E}_{\text{non-CH}}$, in a round by a CH node with distance d from the BS and a non-CH node, respectively. Suppose that a sensor node senses and transmits data once in a round. If a node fails to find a CH, it sends its data to the BS itself. Note that except the ADV, all other messages including data signal are transmitted through spreading code systems with the spreading factor, h.

Let l_{adv} denote the length of the ADV message, l_{join}, the length of join message, l_{tdma}, the TDMA scheduling message, and l_{data}, the length of the data signal, all in [bits]. Let E_{DA} in [J/bit/signal] denote the energy for data aggregation using the beamforming by a CH. Also, we denote the energy expense for CSMA by E_{mac} [J], i.e., the energy consumption while a CH broadcasts an ADV which involves automatic repeat request (ARQ) retransmission by the binary exponential back-off scheme.

Let ζ denote the number of nodes in a cluster, which is a random variable with a variation around its stochastic expectation, $\text{E}[\zeta] = N/k$, with an arbitrary random distribution. For the sake of simplicity, we take the first moment, i.e., the expectation of ζ into account to develop our algorithm. Consider a CH node with the distance d to the BS. The expectation of its energy dissipation in a round is given by

$$\overline{E}_{\text{CH}}(d) = E_0 + E_T(hl_{\text{data}}, d), \tag{4}$$

where

$$\begin{aligned} E_0 = {} & E_{\text{mac}} + E_T(l_{\text{adv}}, \sqrt{2}M) + (\tau_1 + \tau_2 + \tau_3)E_L \\ & + (\text{E}[\zeta] - 1)\, E_R(hl_{\text{join}}) + E_T(hl_{\text{tdma}}, \sqrt{2}M) \\ & + (\text{E}[\zeta] - 1)\, E_R(hl_{\text{data}}) + \text{E}[\zeta] l_{\text{data}} E_{\text{DA}}. \end{aligned} \tag{5}$$

We next calculate the average amount of energy dissipation in a round, $\overline{E}_{\text{non-CH}}$, by a non-CH node. In general, each cluster is an arbitrary shaped region occupying area of M^2/k on average. In general, each cluster is an arbitrary shaped region occupying area of M^2/k on average. For the sake of simplicity, we assume the cluster area to be a circle with radius $R = M/\sqrt{\pi k}$. Often, this circular-cluster assumption might not be sufficiently accurate to model many real-world scenarios. However, it is only intended to capture the essence of energy dissipation of a CH node and a non-CH node in a round. One can develop a more realistic assumption at the cost of complicating the analysis.

Suppose we have a non-CH node in a circular shaped cluster with radius R. Let a random variable r $(0 < r < R)$ denote the distance from the non-CH node to its CH at the center of the circular cluster. The probability density function $f(r)$ of r is given by $f(r) = \frac{2r}{R}$ $\quad (0 < r < R)$. When the non-CH node transmits an l_{data}-bit data signal to its CH, it consumes energy $E_T(hl_{\text{data}}, r)$, which has randomness due to r. The expectation of data transmission energy, $\text{E}[E_T(hl_{\text{data}}, r)]$, from the non-CH node to its CH is calculated as follows: if we have $R < \delta$,

$$\begin{aligned} \text{E}[E_T(hl_{\text{data}}, r)] &= hl_{\text{data}} \left(E_e + \int_0^R \epsilon_f r^2 f(r)\, dr \right) \\ &= hl_{\text{data}} \left(E_e + \frac{\epsilon_f R^2}{2} \right) \end{aligned} \tag{6}$$

else if we have $R \geq \delta$,

$$\begin{aligned} & \text{E}[E_T(hl_{\text{data}}, r)] \\ &= hl_{\text{data}} \left(E_e + \int_0^\delta \epsilon_f r^2 f(r)\, dr + \int_\delta^R \epsilon_m r^4 f(r)\, dr \right) \\ &= hl_{\text{data}} \left(E_e + \frac{\epsilon_f \delta^4}{2R^2} + \frac{\epsilon_m}{3} \left(R^4 - \frac{\delta^6}{R^2} \right) \right). \end{aligned} \tag{7}$$

We now calculate the average energy dissipation in a round, $\overline{E}_{\text{non-CH}}$, of a non-CH node as follows:

$$\begin{aligned} \overline{E}_{\text{non-CH}} &= kE_R(l_{\text{adv}}) + E_T(hl_{\text{join}}, \sqrt{2}M) + \tau_2 E_L + E_R(hl_{\text{tdma}}) \\ &\quad + \text{E}[E_T(hl_{\text{data}}, r)] \\ &= E_e(kl_{\text{adv}} + hl_{\text{tdma}} + hl_{\text{data}}) + E_T(hl_{\text{join}}, \sqrt{2}M) + \tau_2 E_L \\ &\quad + hl_{\text{data}} \left(E_e + \frac{\epsilon_f a^4}{2R^2} + \frac{\epsilon_m}{3} \left(b^4 - \frac{\delta^6}{b^2} \right) \right) \end{aligned} \tag{8}$$

where $a = \min\{R, \delta\}$ and $b = \max\{R, \delta\}$. We also note that different $f(r)$ can be used to model other distribution for node placement.

IV. Distance Based Thresholds for CH Selection

In contrast with the original LEACH, we propose a new protocol called LEACH with Distance-based Thresholds (LEACH-DT) based on the analysis in the previous section. In LEACH-DT, the probability for a node to become a CH depends on its distance to the BS. We assume that the BS is at least δ [m] away from the sensing area. CHs dissipate energy according to the multi-path model (d^4 power loss) in (1).

Let d_i denote the distance from node i to the BS. Given N, k, M, and d_i, we investigate the optimal degree of contribution of node i to CHs through the differentiated percentage p_i of CHs as a function of the distance to the BS d_i as opposed to a fixed $p = k/N$ in LEACH. We keep the expectation of the total number of CHs to be k as in the original LEACH by a constraint:

$$\sum_{i=1}^{N} p_i = k. \tag{9}$$

Consider nodes i and j taking their turns as CHs with probabilities p_i and p_j, respectively. We construct the following balance equation for the expected energy expense:

$$p_i\overline{E}_{\text{CH}}(d_i) + (1-p_i)\overline{E}_{\text{non-CH}} = p_j\overline{E}_{\text{CH}}(d_j) + (1-p_j)\overline{E}_{\text{non-CH}}, \tag{10}$$

which simplifies to:

$$p_i = \frac{p_j(\overline{E}_{\text{CH}}(d_j) - \overline{E}_{\text{non-CH}})}{\overline{E}_{\text{CH}}(d_i) - \overline{E}_{\text{non-CH}}}. \tag{11}$$

Let $c = p_1(\overline{E}_{\text{CH}}(d_1) - \overline{E}_{\text{non-CH}})$ and $\xi_i \equiv 1/(\overline{E}_{\text{CH}}(d_i) - \overline{E}_{\text{non-CH}})$. Then we have $p_i = c\xi_i$. Using (9), we have $c = k/\sum_{j=1}^{N}\xi_j$ and p_i can be computed as:

$$p_i = k\frac{\xi_i}{\sum_{j=1}^{N}\xi_j} \quad (0 \le p_i \le 1), \tag{12}$$

which provides us p_i's that balance the energy consumption.

Consequently, we have a differentiated GOR for each node according to p_i. Node i has its first GOR $\{r; r = 0, 1, \cdots, [1/p_i] - 1\}$, second GOR $\{r; r = [1/p_i], [1/p_i] + 1, \cdots, 2[1/p_i] - 1\}$, and so on. Let $G_i(r)$ be the indicator function determining whether or not node i has been a CH in its current GOR including the current round r, i.e., $G_i(r) = 1$ if node i has been a CH, and $G_i(r) = 0$ otherwise. We now propose new thresholds for the CH selection by having node i choose to become a CH at round r with probability

$$T^*_{i,r} = \begin{cases} \frac{p_i}{1-p_i\cdot(r \mod [\frac{1}{p_i}])}, & G_i(r) = 0 \\ 0, & G_i(r) = 1. \end{cases} \tag{13}$$

Note that at the beginning of each GOR, $G_i(r)$ is set to 0 to ensure that node i becomes a CH once in every GOR of $[1/p_i]$ rounds. The BS initially estimates d_i based on the signal strength from nodes to calculate p_i, and broadcasts them to all nodes. Sensor nodes then make autonomous decisions without any centralized control.

Multi-hop Extension. In the scenarios where some sensors are far away from the BS, a single-hop sensor network is not appropriate due to the limited power of a sensor to transmit signals over a long distance. In this letter, we will outline an extension of multi-hop protocol based on the proposed single-hop LEACH-DT. The main idea of the multi-hop extension is to have the sensors formed different sensor groups (SG) based on their distances of the BS. Each SG runs the proposed single-hop LEACH-DT to select the CH as before. Data is then relayed from the far-away SGs to the closer ones on a hop-by-hop basis with each CH closer to the BS acting as the BS for the sensors in the upstream SG. Specifically, our multi-hop extension can be explained with an example shown in Fig. 1.

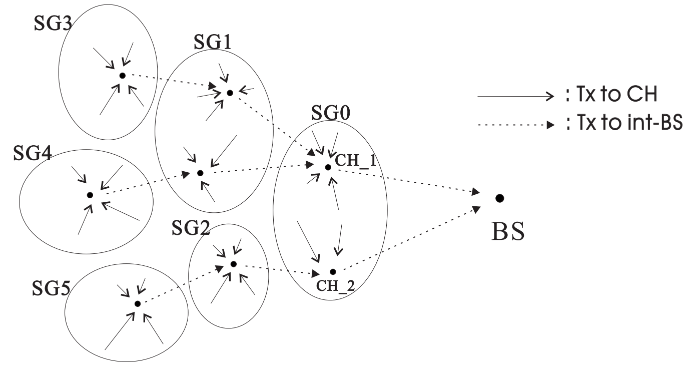

Fig. 1. An example of multi-hop extension.

Each round is initiated by the BS when it broadcasts an ADV message composed of 3-tuple (ID, Level, Code) where the node ID and Level of the BS are defined as ID = 0 and Level = 0, and Code is the spreading code system for CHs to send data signals to the BS. All nodes that have heard the ADV from the BS (Level = 0) are called to have Level 1 and form a sensor group (SG) for the current round, as SG0 in Fig. 1. SG0 runs our proposed LEACH-DT in the current round.

Suppose that 2 CHs, CH_1 and CH_2, are selected in SG0. In the current round, each of the 2 CHs in SG0 broadcasts an ADV message composed of (ID, Level = 1, Code) with appropriate ID and Code for each of the CHs. Note that Codes used by CHs should be designed to avoid inter-SG interference. Now consider the nodes hearing the ADV from CH_1 but not from the BS, we call these nodes belonging to SG1 with Level = 2. In the next round, nodes in SG1 runs LEACH-DT treating CH_1 as their intermediate BS (int-BS). For our multi-hop extension, we add sub-phase IV to the set-up phase for an int-BS to broadcast p_i's to its SG. In general, CHs in SGs with Level = n acts as intermediate BSs for nodes with Level = $(n+1)$, and so on. Consequently, the nodes in the entire network can form SGs with appropriate levels recursively. Data is then flowed in the opposite direction.

V. Performance Evaluation

We simulate our newly proposed LEACH-DT using NS-2. We set $N = 100$ nodes with $k = 5$. The sensor field is located on rectangular coordinates from $(x, y) = (0, 0)$ to $(x, y) = (M, M)$ with $M = 200$ [m]. We consider two cases of BS locations: (i) BS=(100,300) and (ii) BS=(100,500). All nodes start with an initial energy of 2 [J]. The duration of each round is 20 s. As in [1], we use ϵ_f = 10 [pJ/bit/m^2], ϵ_m = 0.0013 [pJ/bit/m^4], E_{DA} = 5[nJ/bit/signal], and E_e = 50 [nJ/bit], which gives $\delta = 87.7$ [m] for (1). Including the length of a data signal l_{data} = 500 [bits], we use the values of $h, l_{\text{adv}}, l_{\text{join}}, l_{\text{tdma}}, l_{\text{data}}, \tau_1, \tau_2$ and τ_3 as in [6]. We observe E_{mac} = 0.05 [J] from running the NS-2 for LEACH. We use $E_L = 0.045$ [Joules/sec] referring [7].

Fig. 2 shows the desired percentage p_i as a function of the distance d_i to the BS for LEACH, LEACH-DT, and the

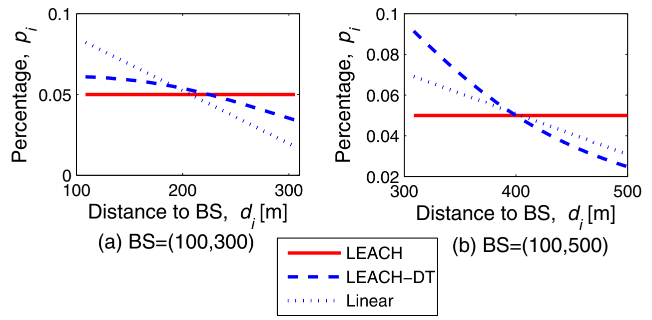

Fig. 2. Percentages p_i of CHs as a function of the distances d_i to the BS.

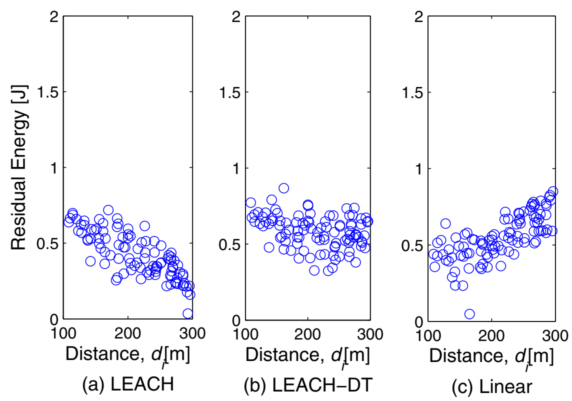

Fig. 3. Residual energy at round 220 as a function of the distance d_i; BS=(100,300), initial energy = 2 [J].

linear scheme. We use MATLAB to implement the analytical derivations in the previous sections. The linear scheme is based on [2] which uses a linear function based on the minimum $d_{\min}$ and maximum $d_{\max}$ of d_i through the factor $\alpha = (d_{\max} - d_{\min})/d_{\max}$. Unlike LEACH, p_i in LEACH-DT and the linear scheme varies inversely with the distances as expected.

Figs. 3 (a), (b), and (c) show the amount of residual energy of each node as a function of the distance d_i at round $r = 220$ for LEACH, LEACH-DT, and the linear scheme, respectively, for the case of BS=(100,300). This figure is drawn from a sample path out of 100 simulations. For LEACH, nodes with large d_i go through fast energy depletion and the graph slants to the right whereas LEACH-DT shows balanced energy consumption. Notably for the linear scheme, the nodes near the BS run out of energy before the others due to overly high probability of being a CH. During the period of rounds 0 through 220 in the figure, the average amount of energy expense per node per round is 0.0072 [J] by LEACH, 0.0063 [J] by LEACH-DT, and 0.0070 [J] by the linear scheme. Fig. 4 shows the mean time to the n-th node death with 95% confidence intervals, where $(n = 1, 2, \cdots, N)$ for the three LEACH, LEACH-DT, and linear schemes. The time in the vertical axes is in rounds. As seen, LEACH-DT outperforms other schemes with larger mean times. Table I also shows the improvement of LEACH-DT over the other two schemes in both mean time to first node death and mean time of node lifespan. We note that the increase of lifespan by the proposed LEACH-DT is greater than 10% over LEACH.

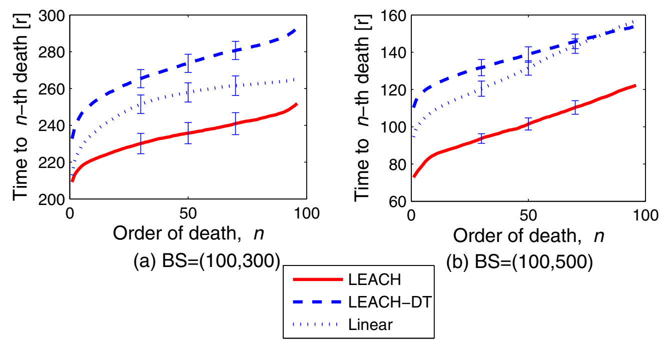

Fig. 4. Mean time to the n-th death as a function of the order of death n.

TABLE I
FIRST NODE DEATH AND MEAN NODE LIFESPAN

	Time to First Node Death [rounds]		
BS	LEACH	LEACH-DT	Linear
(100,300)	210	234	211
(100,500)	72	110	96
	Mean Node Lifespan [rounds]		
BS	LEACH	LEACH-DT	Linear
(100,300)	234	273	254
(100,500)	102	137	132

VI. CONCLUSION

Investigating energy depletion of a node as a CH node and a non-CH node, we propose a new distributed CH selection algorithm LEACH-DT for sensor networks based on the node distance to the BS, in order to balance the energy consumption among the nodes. Simulations show that LEACH-DT outperforms the original LEACH with improved network lifespan over 10%.

REFERENCES

[1] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, “An application-specific protocol architecture for wireless microsensor networks,” *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 660–670, Oct. 2002.

[2] M. Saadat, R. Saadat, and G. Mirjalily, “Improveing threshold assignment for cluster head selection in hierarchical wireless sensor networks,” in *Proc. 2010 Int’l Symposium on Telecommunications*, pp. 409–414.

[3] Y. Liu, K. Xu, Z. Luo, and L. Chen, “A reliable clustering algorithm based on LEACH protocol in wireless mobile sensor networks,” in *Proc. 2010 IEEE ICMET*, pp. 692–696.

[4] P. Ren, J. Qian, L. Li, Z. Zhao, and X. Li, “Unequal clustering scheme based LEACH for wireless sensor networks,” in *Proc. 2010 Int’l Conf. on Genetic and Evolutionary Computing*, pp. 90–93.

[5] V. Katiyar, N. Chand, G. C. Gautam, and A. Kumar, “Improvement in LEACH protocol for large-scale wireless sensor networks,” in *Proc. 2011 ICETECT*, pp. 1070–1075.

[6] http://www.ece.rochester.edu/∼wheinzel/research.html#code

[7] M. G. C. Torres, “Energy consumption in wireless sensor networks using GSP,” M.S. dissertation, Universiy of Pittsburgh, 2006.