On the Topology of Wireless Sensor Networks

Sen Yang, Xinbing Wang, Luoyi Fu
Dept. of Electronic Engineering
Shanghai Jiao Tong University, China
Email: {twood,xwang8,fly}@sjtu.edu.cn

***Abstract*—In this paper, we explore methods to generate optimal network topologies for wireless sensor networks (WSNs) with and without obstacles. Specifically, we investigate a dense network with n sensor nodes and $m = n^b$ $(0 < b < 1)$ helping nodes, and evaluate the impact of topology on its throughput capacity. For networks without obstacles, we find that uniformly distributed sensor nodes and regularly distributed helping nodes have some advantages in improving the throughput capacity. We also explore properties of networks composed of some isomorphic sub-networks. For networks with obstacles, we assume there are $M = \Theta(n^v)$ $(0 < v \leq 1)$ arbitrarily or randomly distributed obstacles, which block cells they are located in, i.e., sensor nodes cannot be placed in these cells and nodes' communication cannot cross them directly. We find that the overall throughput capacity is bounded by the transmission burden in areas around these blocked cells and introduce a novel algorithm of complexity $O(M)$ to generate optimal sensor nodes' topologies for any given obstacles' distributions.**

I. Introduction

Capacity is a fundamental issue in wireless sensor networks (WSNs). A typical WSN involves little or no infrastructure and sensor nodes may communicate in an ad hoc manner. In Gupta and Kumar's seminal work, they adopt Protocol and Physical Model to describe a successful transmission and show that the per-node throughput capacity scales as $\Theta(1/\sqrt{n \log n})$ in random networks, and the per-node transport capacity scales as $\Theta(1/\sqrt{n})$ in arbitrary networks, respectively [1]. These results provide us not only a theoretical bound but also a foundation in the network optimization and protocol design. Following their work, extensive research are conduced to understand the scaling laws in wireless sensor networks better. On the other hand, in some applications, helping nodes are introduced to improve the performance, which results in a heterogeneous network. In heterogeneous networks, access control is studied in [4], routing protocol is studied in [3], N. Li et al. studied topology control in [5], P. Li et al. studied the throughput capacity of networks with rectangular areas in [6]. Many other schemes such as mobility [7], motioncast [8], multi-channel multi-radio (MRMC) [9], network coding [10] and adding infrastructure [11] are also explored in literatures to improve the network capacity.

However, most of works above are for networks with regularly or uniformly distributed sensor nodes. While in practice, sensor nodes may not be placed uniformly, which could have a huge impact on network properties, including the capacity. For example, if lots of nodes are confined in a small region, it would lead to great interference and deteriorates the capacity. Also, if nodes are too sparse in a particular area, communication might get difficult, which also harms networks' performance. To the best of our knowledge, only a few works have dealt with the capacity of networks with inhomogeneous node density. In [12], [13], [14], [15], [16], capacity of inhomogeneous clustered networks are analyzed. Corresponding scheduling and routing schemes to approach the upper bounds are discussed in [17].

On the other hand, almost all the previous works on capacity dealt with flat network region. While in practice, sensor networks are often deployed in complex environments, such as battle fields or mountainous areas, and there are often many obstacles distributed in these regions. These obstacles may constrain the distribution of sensor nodes and the transmission of packets. For example, in a building monitoring WSN, electromagnetic wave signal can be attenuated significantly when passing through furniture, walls or floors, which could have a great impact on network performance. Another example is WSNs deployed in a mountainous area, in which both routing strategy design and deployment of sensor nodes should consider the constraint of the complex landform. Routing strategy for network with obstacles (holes) is discussed in [18], [19], [20]. However, the capacity problem has not been well studied before. Generally, obstacles have a negative impact on the network capacity. But if we design the network topology appropriately, it could lead to a noneligible improvement. For example, in building monitoring WSNs, capacity can be improved if we place less nodes in areas shadowed by obstacles. Also, in a mountainous region, if we deploy more nodes in open areas, network capacity can be much larger than that we put most of them in valleys or laps.

These motivate us to explore better network topologies for given network regions, especially for networks with obstacles. In this paper, we investigate how node distributions influence the throughput capacity and explore the optimal nodes distribution on given conditions. We obtain some useful conclusions on generating the optimal topology for flat network areas. For networks with obstacles, it's difficult to derive a general solution for various obstacle distributions. However, a feasible algorithm with linear complexity can be proposed by dividing the whole network region into some small pieces and dealing with them respectively.

Our main contributions are as follows:

- For networks which consist of many isomorphic sub-networks, compared with the topology of sub-networks, the overall network's topology results in a larger throughput capacity.

- For networks without helping nodes, uniform sensor nodes' distribution is order optimal on maximizing throughput capacity.
- For networks with uniformly distributed sensor nodes, we find that regularly distributed helping nodes are optimal to maximize the network throughput capacity.
- For networks with non-uniformly distributed sensor nodes, though regularly distributed helping nodes are no longer optimal, any improvement on the distribution of helping nodes cannot make a significant change on the throughput capacity in the sense of scaling law.
- For networks with obstacles, we introduce a novel algorithm of complexity $O(M)$ to generate the optimal sensor nodes' topology for any given obstacle distributions.

The rest of the paper is organized as follows. Section II gives the network model. In Section III, we derive the throughput capacity of networks with different topologies. In Section IV and V, we explore some general properties of network topologies. In Section VI, we study the throughput capacity of networks with obstacles and introduce an algorithm to generate the optimal sensor nodes' topology for any given obstacle distributions. We finally conclude this paper in section VII.

II. Network Model

In this section, we introduce the heterogeneous wireless sensor network model, definition of obstacles and routing strategies.

A. Network Components

We consider dense networks with n sensor nodes and $m = n^b$ $(0 < b < 1)$ helping nodes. We assume that the network has asymmetric traffic, i.e., all the n sensor nodes are sources while only n^d $(0 < d < 1)$ sensor nodes are randomly chosen as destinations. Also, sensor nodes can serve as relays if needed. Differently, helping nodes do not have information to transmit or receive. They only help relay packets. Similar to [6], according to whether packets are forwarded by helping nodes, we divide network traffic into two modes, namely, normal mode and helping mode. In normal mode, packets are forwarded by only sensor nodes. While in helping mode, packets are firstly sent to the nearest helping node, and then forwarded to their destinations by helping network. We split the total bandwidth of sensor nodes into three parts: bandwidth for ad hoc transmissions in normal mode, uplink bandwidth in helping mode, and downlink bandwidth in helping mode, denoted by W_1, W_2 and W_3, respectively. We also assume that helping nodes have an additional bandwidth, denoted by W_4, to forward packets among themselves.

B. Definition of Obstacles

To describe networks with obstacles, we assume the network area is partitioned into $\mathcal{K} = \Theta(n^w)$ $(0 < w \leq 1)$ cells. When there is no sensor node distributed in a cell, we assume that at the cell's center there is a relay working in the same bandwidth as sensor nodes, which keeps the network's connectivity. We assume there are $M = \Theta(n^v)$ $(0 < v \leq w)$ number of obstacle nodes in the network area, which can be arbitrarily or randomly distributed. Cells are blocked when there are obstacle nodes in them. Here, "blocked" has two implications: no sensor node can be distributed in these cells and nodes' communication cannot cross them directly. If several blocked cells are adjacent with each other, they form a polygon, which can be either convex or concave, we call it "obstacle polygons".

C. Interference Model

Assume that the network is an unit square and we divide it into non-overlapping cells with equal size. Nodes can communicate with each other only when they are in the neighboring cells. To bound the interference among simultaneous transmissions, we apply a TDMA rotating scheduling scheme as that described in [6]. According to the power propagation model introduced in [21], the reception power at node X_j of the signal from node X_i is

$$P_{ij} = C\frac{P_i}{d_{ij}^{\gamma}}$$

where d_{ij} is the distance between X_i and X_j, and P_i is the power emitted by node X_i. Following Shannon's Theorem, the achievable rate R_{ij} of transmission from node X_i to node X_j is:

$$R_{ij} = W\log(1 + SINR_{ij})$$

where W is the channel bandwidth, and $SINR_{ij} = \frac{C\frac{P_i}{d_{ij}^{\gamma}}}{N+\sum_{k\neq i} C\frac{P_k}{d_{kj}^{\gamma}}}$ is the Signal-to-Interference and Noise Ratio at node X_j. In this paper we assume that sensor nodes use the same transmission power $\mathcal{P}_s$ and helping nodes use the same transmission power $\mathcal{P}_h$, respectively. As derived in [6], we have the following lemma.

***Lemma* 1.** *[6] Each cell in the network can work at a transmission rate c_1W, where c_1 is a deterministic positive constant relevant to the cells' scale and W is the channel bandwidth.*

D. Routing Strategy

As sensor nodes can only communicate with nodes in neighboring cells, packets need to be forwarded through multi-hop transmissions to reach destination nodes. Thus, for networks with and without obstacles, we adopt following routing strategies, respectively.

Routing Strategy I - for networks without obstacles: Suppose a source node is located in cell S_i and its destination is located in cell S_j. Packet sent from the source node is forwarded along cells in the same vertical line of cell S_i until it gets the cell in the same horizontal line of cell S_j, then the packet is forwarded along the cells in the same horizontal line of cell S_j until it reaches the destination node.

Routing Strategy II - for networks with obstacles:

1) If packet sent from the source node can be relayed to its destination by Routing Strategy I, do it.

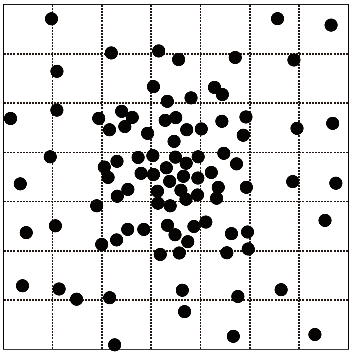

Fig. 1. Centralized networks

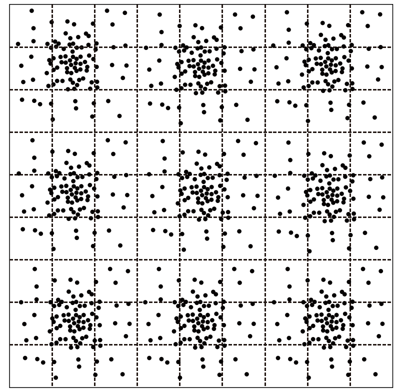

Fig. 2. Multi-centralized networks

2) Otherwise, if there are only convex obstacle polygons, firstly forward the packet along the routing path generated by Routing Strategy I. When it can no longer be forwarded in current direction (vertical or horizontal), change the forwarding direction to another one (horizontal or vertical). Repeat this until it arrives at the destination node.
3) If there exist concave polygons obstacles and neither of the source node and the destination node are in the groove of a concave obstacles polygon, replace the concave obstacle polygons by their convex hulls, respectively, and then following Step 1) and 2) to forward the packet.
4) If source and destination nodes (or either of them) are in the grooves of concave obstacle polygons, we can find cells outside the corresponding convex hulls and nearest to source and destination nodes, respectively. Denote them by S_A and S_B, respectively. First we transmit the packet from the source node to cell S_A, then following Step 1), 2) and 3) to forward this packet from cell S_A to cell S_B, and finally we forward the packet from cell S_B to the destination node.

E. Network Topology

In this paper, we investigate throughput capacity of networks with the following topologies and then generalize the results to get some useful conclusions.

1) Uniform Distribution: For networks with uniformly distributed nodes, the probability density function (PDF) of the distribution is

$$\begin{cases} f(x) = 1 \quad (-\frac{1}{2} < x < \frac{1}{2}) \\ f(y) = 1 \quad (-\frac{1}{2} < y < \frac{1}{2}) \end{cases} \tag{1}$$

2) Centralized Distribution: We consider a simple case of the non-uniform distribution first. As shown in Figure 1, nodes density is large at the center of the network and small at the edge. We call it "centralized distribution". One of its possible PDFs can be described as follows:

$$\begin{cases} f(x) = (4a-4)x + 2 - a \quad (-\frac{1}{2} < x < \frac{1}{2}) \\ f(y) = (4a-4)y + 2 - a \quad (-\frac{1}{2} < y < \frac{1}{2}) \end{cases} \tag{2}$$

where a $(0 \leq a \leq 1)$ is the centralization coefficient, which determines the extension that nodes aggregate to the center. The larger a is, the more uniform the nodes are, or vice versa. Particularly, when $a = 1$ nodes are distributed uniformly; when $a = 0$ probability that nodes distributed at the edge of the network is 0.

3) Multi-centralized Distribution: In practice, network nodes often clustered in several locations of the network, not just the center of the network. We call it "multi-centralized distribution". As shown in Figure 2, We can divide the whole network into many small sub-networks and each sub-network has similar network topology. In this paper, we assume that all the sub-network is a small centralized network, i.e., it satisfied the probability density function given in (2).

III. The Throughput Capacity of Heterogeneous Wireless Sensor Networks without Obstacles

In this section, we study the throughput capacity of heterogeneous WSNs without obstacles by deriving an achievable per-node throughput. Communications in helping mode can be divided into three phases: Firstly, packet is sent from source node to the nearest helping node, then forwarded in the helping-network until it reaches the helping node nearest to the destination, and in the final phase packet is transmitted from that helping node to its destination [6]. We analyze the throughput capacity in normal mode and the three phases of helping mode, respectively. Denote achievable per-node throughput in normal mode and helping mode by T_n and T_h, respectively. Thus, the achievable per-node throughput of the heterogeneous WSN, denoted by T, can be given as follows:

$$T = \max\{T_n, T_h\} \tag{3}$$

where

$$T_h = \min\{T_{h1}, T_{h2}, T_{h3}\} \tag{4}$$

Here, T_{h1}, T_{h2}, T_{h3} are achievable per-node throughput in the three phases of helping mode, respectively.

According to [6], for the network with n source nodes and n^d destination nodes, we have

Lemma 2. *[6] For each destination node, with high probability (w.h.p.), there are at most $2n^{1-d}$ data flows destined to it.*

Thus, using similar techniques in [6], we can calculate networks' throughput capacity as follows.

1) In normal mode, let $N_{x,max}$ and $N_{y,max}$ denote the maximal number of source nodes located in one column and the maximal number of destination nodes located in one row, respectively. Since each source node generate only one data flow and each destination node has at most $2n^{1-d}$ flows destined to it w.h.p., the maximal number of flows crossing a cell, denoted by $F_{ij,max}$, is

$$F_{ij,max} \leq N_{x,max} + 2n^{1-d} N_{y,max} \tag{5}$$

According to Lemma 1, each cell can achieve a constant transmission rate. Thus, the achievable throughput in normal mode is

$$T_n = \Omega\left(\frac{W_1}{F_{ij,max}}\right) \tag{6}$$

2) In helping mode, let C_{max}, D_{max}, $N'_{x,max}$ and $N'_{y,max}$ denote the maximal number of source nodes in one cell, the maximal number of destination nodes in one cell, the maximal number of source nodes in one column and the maximal number of destination nodes in one row, respectively. Denote the maximal number of flows crossing a cell in the second phase of helping mode by $F'_{ij,max}$, we have

- In the first phase

$$T_{h1} = \Omega\left(\frac{W_2}{C_{max}}\right) \tag{7}$$

- In the second phase

$$F'_{ij,max} \leq N'_{x,max} + 2n^{1-d} N'_{y,max} \tag{8}$$

$$T_{h2} = \Omega\left(\frac{W_4}{F'_{ij,max}}\right) \tag{9}$$

- In the third phase

$$T_{h3} = \Omega\left(\frac{W_3}{n^{1-d} D_{max}}\right) \tag{10}$$

In this section, we assume that all the helping nodes are placed regularly and only investigate the impact of the sensor nodes' topology. Impacts of helping nodes' topology are studied in the following sections. Following this trace of derivation, we can obtain the following theorems.

A. Uniform Network

The uniform network is a special case of [6], in which network region is assumed to be rectangular. Hence the following theorem can be derived from [6] directly.

Theorem 1. *[6] An achievable throughput in uniform networks, denoted by $T^{uniform}$, is*

$$T^{uniform} = \Omega\left(\max\left\{\min\left\{\frac{1}{\sqrt{n\log n}}, n^{d-1}\right\}, \min\left\{n^{\frac{b}{2}-1}, n^{d-1}\right\}\right\}\right) \tag{11}$$

B. Centralized Network

To facilitate the calculation, here we will only consider the case that centralization coefficient is 0, i.e., nodes density at the center goes to the maximum. Results of such extreme case are also easier for us to compare with that of the uniform network.

1) Achievable Throughput in Normal Mode: Let N_x^i and N_y^j denote the number of source nodes located in the ith column and the number of destination nodes located in the jth row, respectively. We have

$$E[N_x^i] = 2n(-2i+1)\left(\tfrac{c_2\log n}{n}\right)^{\frac{1}{2}} + 2n\left(\tfrac{c_2\log n}{n}\right)^{\frac{1}{4}}$$
$$E[N_y^j] = 2n^d(-2j+1)\left(\tfrac{c_2\log n}{n}\right)^{\frac{1}{2}} + 2n^d\left(\tfrac{c_2\log n}{n}\right)^{\frac{1}{4}} \tag{12}$$

Apply Chernoff bounds to equation (12), we can obtain the following lemma. The detailed proof is omitted due to the space limitation.

***Lemma* 3.** *For each cell, w.h.p.,*

1) *The number of source nodes which are located in the same column is at most $4n(-2i+1)\left(\frac{c_2\log n}{n}\right)^{\frac{1}{2}} + 4n\left(\frac{c_2\log n}{n}\right)^{\frac{1}{4}}$.*
2) *The number of destination nodes which are located in the same row is at most $4n^d(-2j+1)\left(\frac{c_2\log n}{n}\right)^{\frac{1}{2}} + 4n^d\left(\frac{c_2\log n}{n}\right)^{\frac{1}{4}}$ when $1/4 < d < 1$, and at most c_3 when $0 < d < 1/4$, where c_3 is a constant and $c_3 > \frac{2}{1-4d}$.*

Let F_k^{ij} denote the number of data flows crossing cell $S(i,j)$. For each cell, we have

$$\begin{aligned} F_k^{ij} &\leq N_x^i + 2n^{1-d} N_y^j \\ &\leq \begin{cases} 4n(-2i+1)\left(\frac{c_2\log n}{n}\right)^{\frac{1}{2}} + 4n\left(\frac{c_2\log n}{n}\right)^{\frac{1}{4}} \\ +2n^{1-d}\left[4n^d(-2j+1)\left(\frac{c_2\log n}{n}\right)^{\frac{1}{2}}\right. \\ \left.+4n^d\left(\frac{c_2\log n}{n}\right)^{\frac{1}{4}}\right] \quad \text{when } \frac{1}{4} < d < 1 \\ 4n(-2i+1)\left(\frac{c_2\log n}{n}\right)^{\frac{1}{2}} + 4n\left(\frac{c_2\log n}{n}\right)^{\frac{1}{4}} \\ +2c_3 n^{1-d} \quad \text{when } 0 < d < \frac{1}{4} \end{cases} \end{aligned} \tag{13}$$

Notice that the left part of (13) are monotonically decreasing functions of i and j, we have

$$\begin{aligned} F_{k,max}^{ij} &= F_k^{11} \\ &= \begin{cases} 9n\left[\left(\frac{c_2\log n}{n}\right)^{\frac{1}{4}} - \left(\frac{c_2\log n}{n}\right)^{\frac{1}{2}}\right] \\ \quad \text{when } \frac{1}{4} < d < 1 \\ 4n\left[\left(\frac{c_2\log n}{n}\right)^{\frac{1}{4}} - \left(\frac{c_2\log n}{n}\right)^{\frac{1}{2}}\right] + 2c_3 n^{1-d} \\ \quad \text{when } 0 < d < \frac{1}{4} \end{cases} \\ &= O\left(\max\left\{n\cdot\left(\frac{\log n}{n}\right)^{\frac{1}{4}}, n^{1-d}\right\}\right) \end{aligned} \tag{14}$$

Denote the achievable throughput in normal mode by $T_n^{central}$, from (14), we can obtain that

$$T_n^{central} = \Omega\left(\min\left\{n^{-1}\cdot\left(\frac{n}{\log n}\right)^{\frac{1}{4}}, n^{d-1}\right\}\right) \quad (15)$$

2) Achievable Throughput in Helping Mode: Since helping nodes are regularly placed, we can divide the network into m cells of length $l' = \sqrt{1/m} = n^{-\frac{b}{2}}$. Let $T_{h1}^{central}$, $T_{h2}^{central}$ and $T_{h3}^{central}$ denote the achievable throughput in phases I, II and III, respectively. Employing similar techniques, we can obtain that

$$T_{h1}^{central} = \Omega\left(\frac{W_2}{8n\left(-n^{-b}+n^{-\frac{b}{2}}\right)^2}\right) = \Omega(n^{b-1})$$

$$T_{h2}^{central} = \Omega\left(\max\left\{n^{\frac{b}{2}-1}, n^{d-1}\right\}\right)$$

$$T_{h3}^{central} = \begin{cases} \Omega(n^{b-1}) & \text{when } 0<b<d<1 \\ \Omega(n^{d-1}) & \text{when } 0<b<d<1 \end{cases}$$

Thus, we have

$$\begin{aligned} T_h^{central} &= \min\{T_{h1}^{central}, T_{h2}^{central}, T_{h3}^{central}\} \\ &= \Omega\left(\min\left\{n^{\frac{b}{2}-1}, n^{d-1}\right\}\right) \end{aligned} \quad (16)$$

Substituting (15) and (16) into (3), we can obtain the following theorem

Theorem 2. *An achievable throughput in centralized networks, denoted by $T^{central}$, is*

$$T^{central} = \Omega\left(\max\left\{\min\left\{n^{-1}\cdot\left(\frac{n}{\log n}\right)^{\frac{1}{4}}, n^{d-1}\right\},\right.\right. \left.\left.\min\left\{n^{\frac{b}{2}-1}, n^{d-1}\right\}\right\}\right) \quad (17)$$

C. Multi-centralized Network

Theorem 3. *Denote the achievable throughput in multi-centralized networks denoted by T^{multi}, we have*

$$T^{multi} = \Omega\left(\max\left\{\min\left\{\frac{k}{n}\cdot\left(\frac{\frac{n}{k^2}}{\log\frac{n}{k^2}}\right)^{\frac{1}{4}}, n^{d-1}\right\},\right.\right. \left.\left.\min\left\{kn^{\frac{b}{2}-1}, n^{d-1}\right\}\right\}\right) \quad (18)$$

Proof: Similar to that in centralized network, we have

1) *In normal mode*

For multi-centralized network, in every sub-network, there are n/k^2 nodes. Thus, in normal mode, cells' length is $\frac{1}{k}\cdot\sqrt[4]{\frac{c_4\frac{n}{k^2}}{\log\frac{n}{k^2}}}$. Similar to that in section III-B, we can obtain that

$$T_n^{multi} = \min\left\{\frac{k}{n}\cdot\left(\frac{\frac{n}{k^2}}{\log\frac{n}{k^2}}\right)^{\frac{1}{4}}, n^{d-1}\right\}$$

2) *In helping mode*

The edge of cell in helping mode is still $\sqrt{\frac{1}{m}}$ while the area of each centralized sub-network is only $\frac{1}{k^2}$. Thus, we can obtain that

$$\begin{aligned} T_{h1}^{multi} &= \Omega\left(\frac{W_2}{8\frac{n}{k^2}\left(-k^2\cdot n^{-b}+k\cdot n^{-\frac{b}{2}}\right)^2}\right) \\ &= \Omega\left(n^{b-1}\right) \\ T_{h2}^{multi} &= \Omega\left(\min\left\{\frac{W_4}{9k\cdot\frac{n}{k^2}\left(-n^{-b}+n^{-\frac{2}{b}}\right)}, n^{d-1}\right\}\right) \\ &= \Omega\left(\min\left\{kn^{\frac{b}{2}-1}, n^{d-1}\right\}\right) \\ T_{h3}^{multi} &= \begin{cases} \Omega(n^{b-1}) & \text{when } 0<b<d<1 \\ \Omega(n^{d-1}) & \text{when } 0<b<d<1 \end{cases} \end{aligned}$$

Equation (18) can be obtained by substituting results above into (3) and (4). ■

IV. General Properties of "Combined Networks"

From results in section III, we can see that compared with centralized network, the multi-centralized network has a larger throughput capacity. It seems that for network consisting some small isomorphic sub-networks (we call it "combined network"), if we extend the sub-network to the same scale[1] as the overall network, its throughput capacity is still smaller than the combine network. Thus, we can obtain the following Theorem.

Theorem 4. *For network composed of some isomorphic sub-networks, the throughput capacity of the overall network, denoted by $\widetilde{T}$, and the throughput capacity of sub-network of same network scales, denoted by $\widetilde{\widetilde{T}}$, have the following relationship.*

$$\widetilde{T} \geq \widetilde{\widetilde{T}} \quad (19)$$

Proof: From Section III, we can see that the achievable throughput of a network without obstacles is determined by cells and rows of the largest nodes density, or equivalently, by variables $N_{x,max}$ $N_{y,max}$, C_{max}, D_{max}, $N'_{x,max}$ and $N'_{y,max}$. In the combined network which consists of $k\times k$ sub-networks, denote corresponding variables relevant to the overall network by $\widetilde{(\cdot)}$ and corresponding variables relevant to one particular sub-network (has not be extended) by $(\cdot)^{(s)}$. We have

$$\begin{cases} \widetilde{C}_{max} = C_{max}^{(s)} \\ \widetilde{D}_{max} = D_{max}^{(s)} \\ \widetilde{N}_{x,max} = kN_{x,max}^{(s)} \\ \widetilde{N}_{y,max} = kN_{y,max}^{(s)} \\ \widetilde{N}'_{x,max} = kN_{x,max}^{(s)'} \\ \widetilde{N}'_{y,max} = kN_{y,max}^{(s)'} \end{cases} \quad (20)$$

[1]Here, "scale" means the size of network area, number of nodes and size of cells.

Furthermore, we have

$$\begin{aligned}\widetilde{F}_{ij,max} &= \widetilde{N}_{x,max} + 2n^{1-d}\widetilde{N}_{y,max} \\ &= kN^{(s)}{}_{x,max} + 2n^{1-d}kN^{(s)}_{y,max} \\ &= kF^{(s)}_{ij,max} \end{aligned} \tag{21}$$

Similarly, we have

$$\widetilde{F}'_{ij,max} = kF^{(s)'}_{ij,max} \tag{22}$$

Substituting (20), (21) and (22) into (6), (7), (9) and (10), we can obtain that

$$\widetilde{T}_n = \Omega\left(\frac{W_1}{\widetilde{F}_{ij,max}}\right) = \Omega\left(\frac{W_1}{kF^{(s)}_{ij,max}}\right)$$

$$\widetilde{T}_{h1} = \Omega\left(\frac{W_2}{\widetilde{C}_{max}}\right) = \Omega\left(\frac{W_2}{C^{(s)}_{max}}\right)$$

$$\widetilde{T}_{h2} = \Omega\left(\frac{W_4}{\widetilde{F}'_{ij,max}}\right) = \Omega\left(\frac{W_4}{kF^{(s)'}_{ij,max}}\right)$$

$$\widetilde{T}_{h3} = \Omega\left(\frac{W_3}{\widetilde{D}_{max}}\right) = \Omega\left(\frac{W_3}{D^{(s)}_{max}}\right)$$

On the other hand, if we extend this sub-network to the same scale as the combined network, the number of nodes and the number of cells in this extended sub-network are both k^2 times larger than before. However, if the nodes distribution is not uniform, the maximal node density in one cell or in one row is larger than before. Denote the corresponding variables relevant to this extended sub-network by $\widetilde{\widetilde{(\cdot)}}$. We have

$$\widetilde{\widetilde{T}}_n = \Omega\left(\frac{W_1}{\widetilde{\widetilde{F}}_{ij,max}}\right) \leq \Omega\left(\frac{W_1}{k^2 \cdot \frac{1}{k} \cdot F^{(s)}_{ij,max}}\right) = \widetilde{T}_n$$

$$\widetilde{\widetilde{T}}_{h1} = \Omega\left(\frac{W_2}{\widetilde{\widetilde{C}}_{max}}\right) \leq \Omega\left(\frac{W_2}{k^2 \cdot \frac{1}{k^2} \cdot C^{(s)}_{max}}\right) = \widetilde{T}_{h1}$$

$$\widetilde{\widetilde{T}}_{h2} = \Omega\left(\frac{W_4}{\widetilde{\widetilde{F}}'_{ij,max}}\right) \leq \Omega\left(\frac{W_4}{k^2 \cdot \frac{1}{k} \cdot F^{(s)'}_{ij,max}}\right) = \widetilde{T}_{h2}$$

$$\widetilde{\widetilde{T}}_{h3} = \Omega\left(\frac{W_3}{\widetilde{\widetilde{D}}_{max}}\right) \leq \Omega\left(\frac{W_3}{k^2 \cdot \frac{1}{k^2} \cdot D^{(s)}_{max}}\right) = \widetilde{T}_{h3}$$

Thus, we can obtain that

$$\begin{aligned}\widetilde{T} &= \max\{\widetilde{T}_n, \widetilde{T}_h\} \\ &= \max\{\widetilde{T}_n, \min\{\widetilde{T}_{h1}, \widetilde{T}_{h2}, \widetilde{T}_{h3}\}\} \\ &\geq \max\{\widetilde{\widetilde{T}}_n, \min\{\widetilde{\widetilde{T}}_{h1}, \widetilde{\widetilde{T}}_{h2}, \widetilde{\widetilde{T}}_{h3}\}\} \\ &= \widetilde{\widetilde{T}} \end{aligned} \tag{23}$$

■

Conclusion in Theorem 4 means that compared with the topology of sub-networks, the overall network's topology results in a larger throughput capacity, which provide us a method to generate a better topology to achieve a larger throughput capacity. Intuitively, the overall network has a more uniform distribution than the sub-networks. As discussed in the following section, uniform distribution tend to result in a larger capacity, which also demonstrates the conclusion in this section.

V. Impact of Network Topology on Throughput Capacity

A. Impact of Sensor Nodes' Topology

Comparing the results in section III with each other, we can find that they have similar scales (take k as a constant). In general, we have the following theorem.

Theorem 5. *For the topology of sensor nodes, if the value range of nodes distribution's PDF is limited, the gap in achievable throughput of non-uniform networks and uniform networks is at most a constant time.*

Proof: Firstly, from the analysis in section IV, we can see that if size of cells stay unchanged, interference can not be changed by network topologies. Secondly, in the proof of Theorem 4, we have concluded that in networks without obstacles the achievable throughput is determined by the busiest cells, i.e., by variables $N_{x,max}$ $N_{y,max}$, C_{max}, D_{max}, $N'_{x,max}$ and $N'_{y,max}$. Let $\overline{N}_{x,max}$ and $\widehat{N}_{x,max}$ denote the the maximal number of source nodes located in the same column in uniform and non-uniform network, respectively. Since the value range of the nodes distribution's PDF is limited, i.e., $\exists M \in R^+$, for $\forall\ x, y$ we have $|f(x)| \leq M, |f(y)| \leq M$. Thus, we have

$$E[\overline{N}_x] = n \cdot \frac{l}{1} \tag{24}$$

$$\begin{aligned} E[\widehat{N}^i_x] &= n \cdot \int_{(i-1)l}^{il} f(y)dy \\ &\leq n \cdot \int_{(i-1)l}^{il} Mdy \\ &= Mnl \\ &= ME[\overline{N}_x] \end{aligned}$$

Using Chernoff Bounds, we can prove that w.h.p. $\overline{N}_x \leq 2E[\overline{N}_x]$ and $\widehat{N}^i_x \leq 2E[\widehat{N}^i_x]$. Thus, we can obtain that

$$\widehat{N}_{x,max} \leq M\overline{N}_{x,max} \tag{25}$$

Similar results can be proved for $N_{y,max}$, C_{max}, D_{max}, $N'_{x,max}$ and $N'_{y,max}$. Denoted the achievable throughput in uniform and non-uniform network by $\overline{T}$ and $\widehat{T}$, respectively. Following similar trace of derivation in Theorem 4' proof, we can conclude that

$$\overline{T} \leq C(M)\widehat{T} \tag{26}$$

where $C(M)$ is a constant relative to M. ■

B. Impact of Helping Nodes' Topology

In analyses above, we have only considered regularly distributed helping nodes. In this subsection, we investigate the impact of variability in helping nodes' distribution. Generally, we have the following theorem.

Theorem 6. *For networks with uniformly distributed sensor nodes, regularly distributed helping nodes are optimal to maximize the network throughput capacity.*

Proof: Firstly, we analyze the impact on interference. In network with non-uniformly distributed helping nodes, we have to divide network area into cells of different lengths. Let $l^{(r)}$ denote length of cells in network with regularly distributed helping nodes, and $l_{max}^{(n)}$ denote the maximal cells' length in network with non-uniformly distributed helping nodes, respectively. We can easily obtain that $l_{max}^{(n)} \geq l^{(r)}$. Denote the achievable rate per cell in networks with regularly and non-uniformly distributed helping nodes by $W^{(r)}$ and $W^{(n)}$, respectively. From the derivation of Lemma 1, we can obtain that $W^{(r)} \geq W^{(n)}$.

Secondly, similar to the proof of Theorem 4, let $C_{max}^{(r)}$, $D_{max}^{(r)}$, ${N_{x,max}^{(r)}}'$ and ${N_{y,max}^{(r)}}'$ denote corresponding variables in networks with regularly distributed helping nodes, and $C_{max}^{(n)}$, $D_{max}^{(n)}$, ${N_{x,max}^{(n)}}'$ and ${N_{y,max}^{(n)}}'$ denote corresponding variables in networks with non-uniformly distributed helping nodes, respectively. In networks with non-uniformly distributed helping nodes, there must be cells of length larger than the average value. Thus, we can obtain that

$$\begin{cases} C_{max}^{(n)} > C_{max}^{(r)} \\ D_{max}^{(n)} > D_{max}^{(r)} \\ {N_{x,max}^{(n)}}' > {N_{x,max}^{(r)}}' \\ {N_{y,max}^{(n)}}' > {N_{y,max}^{(r)}}' \end{cases} \quad (27)$$

Denote the achievable throughput capacity of network with regularly and non-uniformly distributed network by $T^{(r)}$ and $T^{(n)}$, respectively. Since that throughput capacity in helping mode is inversely proportional to variables above, we can conclude that

$$T^{(r)} \geq T^{(n)} \quad (28)$$

■

Theorem 7. *For networks with non-uniformly distributed sensor nodes, though regularly distributed helping nodes are no longer optimal, any improvement on the helping nodes' topology cannot change the scale of network throughput capacity.*

Proof: Firstly, consider the interference. According to the proof of Theorem 6, non-uniformly distributed helping nodes can only increase the interference and thus decrease the achievable per-cell transmission rate. Thus, it has a negative impact on the network throughput capacity.

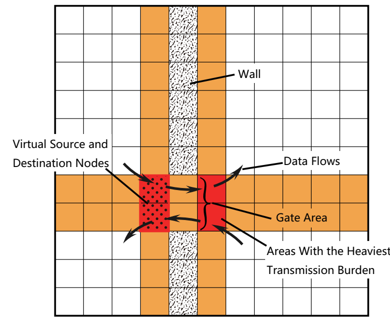

Fig. 3. A wall with gate in the network area

Secondly, if we do not consider the change of interference, network throughput capacity in helping mode is determined by variables C_{max}, D_{max}, $N'_{x,max}$ and $N'_{y,max}$. Thus, if we change the helping nodes' topology, network throughput capacity achieves the maximum when each cell, each column and each row has the same number of nodes, respectively. However, similar to that in the proof of Theorem 5, we can easily show that this improvement is not larger than a constant time.

Combining conclusions above together, we can conclude that improvement on helping nodes' topology cannot change the network throughput capacity in the sense of scaling law. ■

VI. Optimal Topology for Networks with Obstacles

In this section, we introduce a novel algorithm to generate the optimal network topology for any given obstacle distributions and analyze its performance.

A. Algorithm to Obtain the Optimal Network Topology

To design the optimization algorithm, we consider a simple scenario first. As shown in Figure 3, assume that there is a "wall" with a "gate" in the network, which divides the network area into two parts. In this case, area around the gate is the communication bottleneck since any data flow transmitting from one side of the wall to another side has to pass through the gate. To maximize the throughput capacity, we can reduce the transmission burden in this area by the following algorithm.

Algorithm - "Wall with Gate":

1) Assume that there are $\widehat{n}$, n_1 and n_2 number of nodes in the gate area, the left and the right part of the network, respectively, where $\widehat{n}+n_1+n_2 = n$. The expect number of data flows passing through the gate is $u = f(\widehat{n}, n_1, n_2)$, where function $f(\cdot)$ can be decided using techniques given in Section III. Thus the transmission burden of the gate area is $B_0 = u/k_0$, where k_0 is the number of cells in the gate area (nodes' distribution in the gate area is assumed to be uniform since this area is relatively small).
2) Ignore the wall and the right part of the network. Put $\varphi_1 = g_s(\widehat{n}, n_1, n_2)$ number of virtual source nodes and $\phi_1 = g_d(\widehat{n}, n_1, n_2)$ number of virtual destination nodes

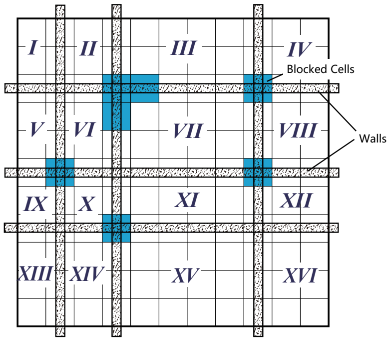

Fig. 4. Network dividing method I

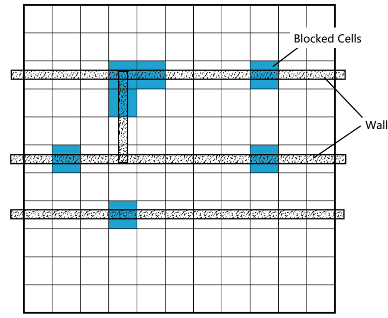

Fig. 5. Network dividing method II

uniformly in front of the gate (i.e., the area illustrated in Figure 3) to replace the ignored sensor nodes. Virtual source and destination nodes work as sources and destinations, respectively, generating virtual date flows. Functions $g_s(\cdot)$ and $g_d(\cdot)$ are designed according to the routing strategy so that these numbers of virtual nodes have the same influence on the left part of the network as the ignored parts.[2] Then we obtain a degraded sub-network without any obstacles.

3) For the degraded sub-network, use techniques and conclusions given in Sections III - V to generate an optimal sub-network topology $\mathfrak{T}_1 = T_1(\widehat{n}, n_1, n_2)$. The basic idea is to reduce the transmission burden of the shadowed areas and distribute it uniformly to the whole sub-network, or equivalently, to minimize the maximal transmission burden $B_1 = \max{(N_x^i + 2n^{1-d}N_y^i)} = h_1(\widehat{n}, n_1, n_2)$.
4) Repeat Step 2 and 3 for the right part of the network, respectively. Generate an optimal subnetwork topology $\mathfrak{T}_2 = T_2(\widehat{n}, n_1, n_2)$ and calculate the minimized sub-network transmission burden $B_2 = h_2(\widehat{n}, n_1, n_2)$.
5) Use appropriate optimization methods to minimize the burden function $B = \max(B_0, B_1, B_2)$. Calculate corresponding $\widehat{n}$, n_1 and n_2. Thus, the optimal topology for the whole network, denoted by $\mathfrak{T}$, can be obtained by combining $\mathfrak{T}_1$ and $\mathfrak{T}_2$, i.e.,

$$\mathfrak{T} = T(\widehat{n}, n_1, n_2) = T'(\mathfrak{T}_1, \mathfrak{T}_2)$$

where $T(\cdot)$ and $T'(\cdot)$ are deterministic functions.

This "Wall with Gate" algorithm can be generalized to obtain optimal topologies for any given networks with obstacles. Firstly, divide the network region into pieces by the following method.

Divide the network by walls - Method I: As shown in Figure 4, take blocked cells in a row (either vertical or horizontal) as a wall and cells without obstacles in this row as gates. Thus, the network region is divided into some sub-networks by these walls.

The optimal topology for this network region can be obtained by applying "Wall with Gate" algorithm to each of these sub-networks and gate areas. Note that since the gate areas here might be relatively large, nodes distribution in these areas can no longer be assumed to be uniform and Step 2 - 3 must be applied to these gate areas. Assuming that there are S sub-networks and R gates, we have

$$\begin{aligned}\mathfrak{T} &= T(n_1, n_2, \ldots, n_S,\ \widehat{n}_1, \widehat{n}_2, \cdots, \widehat{n}_R) \\ &= T'(\mathfrak{T}_1, \mathfrak{T}_2, \ldots, \mathfrak{T}_S,\ \widehat{\mathfrak{T}}_1, \widehat{\mathfrak{T}}_2, \ldots, \widehat{\mathfrak{T}}_R)\end{aligned}$$

where n_i $(i = 1, 2, \ldots S)$ is the number of sensor nodes in the ith sub-network, $\widehat{n}_j$ $(j = 1, 2, \ldots R)$ is the number of sensor nodes in the jth gate area, $\mathfrak{T}_k$ $(k = 1, 2, \ldots S)$ is the optimal topology of the kth sub-network and $\widehat{\mathfrak{T}}_l$ $(l = 1, 2, \ldots R)$ is the optimal topology of the lth gate area.

B. Complexity of the Algorithm

Noticing that the algorithm complexity is proportional to the number of sub-networks S and number of gates R, to reduce the algorithm complexity, we can simplify the division of the network area. Note that in Figure 4, sub-networks I - IV can be combined into a larger sub-network, so do sub-networks V - VI, VII - $VIII$, IX - XII and $XIII$ - XVI. We can modify the network dividing method as follows.

Divide the network by walls - Method II: As shown in Figure 5, firstly construct a wall in the row (either vertical or horizontal) with the most number of blocked cells, dividing the network area into two parts. For each part, repeat this step iteratively until all the blocked cells are crossed by at least one wall.

The complexity of the algorithm is given by the following lemma.

***Lemma* 4.** *The algorithm complexity is $O(M^2)$ when using network dividing method I and is $O(M)$ when using method II.*

Proof: Here, we consider the worst cases, i.e., cases that all blocked cells are not collinear. When using method I to divide the network area, there are $2M$ walls. Denote the number of sub-networks and number of gates by S_I and R_I, respectively. In the worst case, these $2M$ walls divide the network into M^2 areas, i.e., $S_I = M^2$. Since there is a gate between any neighboring sub-network areas, there are at least

[2]For the routing strategy given in Section II, we let $g_s(\widehat{n}, n_1, n_2) = n_1 n_2/n$ and $g_d(\widehat{n}, n_1, n_2) = n_1 n_2/n$, respectively.

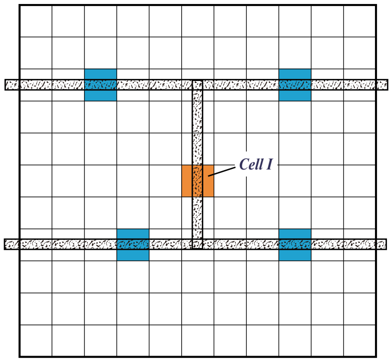

Fig. 6. Cell I add four gates to the divided network

$4*M^2/2$ gates, i.e., $R_I = 2M^2$. So the algorithm complexity is

$$\eta_I = O(S_I + R_I) = O(M^2) \quad (29)$$

When using method II, each cell with obstacles generate at most one wall, so there are at most M walls. Denote the number of sub-networks and number of gates by S_{II} and R_{II}, respectively. These walls divide the network into $M+1$ areas, thus, $S_{II} = M + 1$. Furthermore, as shown in Figure 6, each cell with obstacles can generate at most four additional gates to the network, i.e., $R_{II} \leq 4M$. So the algorithm complexity is

$$\eta_{II} = O(S_{II} + R_{II}) = O(M) \quad (30)$$

■

Although method I generate more sub-networks, each subnetwork is relatively simple and easy to analyze. In some particular cases, for example, in the case that the obstacle distribution has some symmetry properties, using method I might result in a smaller algorithm complexity.

VII. Conclusion

In this paper, we investigate the throughput capacity of heterogeneous WSNs with different network topologies and analyze the impact of topologies on network properties. We find that compared to sub-networks of the same scale, combined networks have a larger overall throughput capacity. We also find that uniformly distributed sensor nodes and regularly distributed helping nodes have some advantages in improving the network capacity. Compared to regularly distributed helping nodes, any change of helping nodes' topology cannot improve the network throughput capacity in the sense of scaling law. We further investigate the impact of obstacles and introduce an algorithm to generate the optimal sensor nodes distribution for any given network areas with obstacles.

VIII. Acknowledgment

This work is supported by National Fundamental research grant (2011CB302701), NSF China (No. 60832005); China Ministry of Education New Century Excellent Talent (No. NCET-10-0580); China Ministry of Education Fok Ying Tung Fund (No. 122002); Qualcomm Research Grant; Shanghai Basic Research Key Project (No.11JC1405100); National Key Project of China (2010ZX03003-001-01).

References

[1] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transaction on Information Theory*, vol. 46, no. 2, pp. 388-404, Mar. 2000.

[2] A. Keshavarz-Haddad, V. Ribeiro, and R. Riedi, "Broadcast capacity in multihop wireless networks," in *Proc. ACM MobiCom*, Sept. 2006.

[3] K. Yang, Y. Wu, and H.-H. Chen, "Qos-aware routing in emerging heterogeneous wireless networks," *IEEE Communications Magazine*, 45(2):74-80, 2007.

[4] P. Li, X. Geng, and Y. Fang, "An adaptive power controlled mac protocol for wireless ad hoc networks," *IEEE Transactions on Wireless Communications*, 8(1):226-233, Jan. 2009.

[5] N. Li and J. Hou, "Topology control in heterogeneous wireless networks: Problems and solutions," in *Proc. IEEE INFOCOM*, Hong Kong, China, Mar. 2004.

[6] P. Li and Y. Fang, The "Capacity of heterogeneous wireless networks," in *Proc. IEEE INFOCOM*, 2010.

[7] J. Luo, X. Liu, D. Ye, "Research on Multicast Routing Protocols for Mobile Ad-hoc Networks," in *Elsevier Computer Networks Journal*, Vol. 52, No. 5, pp. 988-997, Apr. 2008.

[8] X. Wang, W. Huang, S. Wang, J. Zhang, C. Hu, "Delay and Capacity Tradeoff Analysis for MotionCast," *IEEE/ACM Transactions on Networking*, Vol. 19, no. 5, pp. 1354-1367, Oct. 2011.

[9] D. Shila, Y. Cheng, T. Anjali, and P. Wan, "Extracting more capacity from multi-channel multi-radio wireless networks by exploiting power", in *Proc. IEEE ICDCS*, Genoa, Italy, Jun. 2010.

[10] H. Li, X. Liu, W. He, J. Li, W. Dou, "End-to-End Delay Analysis in Wireless Network Coding: A Network Calculus-Based Approach," in *Proc. of ICDCS*, Minneapolis, Minnesota, 2011.

[11] D. Shila, Y. Cheng, and T. Anjali, "Throughput and delay analysis of hybrid wireless networks with multi-hop uplinks," in *Proc. IEEE INFOCOM*, Shanghai, China, Apr. 2011.

[12] S. R. Kulkarni, P.Viswanath, "Deterministic Approach to Throughput Scaling in Wireless Networks", *IEEE Transaction on Information Theory*, vol. 50(6), pp. 1041-049, Jun. 2004.

[13] E. Perevalov, R. S. Blum, D. Safi, "Capacity of Clustered Ad Hoc Networks: How Large Is Large?" *IEEE Transaction on Communication*, Vol. 54, No. 9, pp. 1672-1681, Sept. 2006.

[14] R. K. Ganti and M. Haenggi, "Interference and Outage in Clustered Wireless Ad Hoc Networks," *IEEE Transaction on Information Theory*. Available at http://arxiv.org/abs/0706.2434v1

[15] A. Keshavarz-Haddad and R. H. Riedi, "Bounds for the capacity of wireless multihop networks imposed by topology and demand," in *Proc. ACM MobiHoc*, pp. 256-265, Montreal, Canada, Sept. 2007.

[16] G. Alfano, M. Garetto, E. Leonardi, "Capacity Scaling of Wireless Networks with Inhomogeneous Node Density: Upper Bounds," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 7, Sept. 2009.

[17] G. Alfano, M. Garetto, E. Leonardi, "Capacity Scaling of Wireless Networks with Inhomogeneous Node Density: Lower Bounds," in *Proc. INFOCOM*, Rio de Janeiro, Brazil, Apr. 2009.

[18] M. Li and Y. Liu, "Rendered Path: Range-Free Localization in Anisotropic Sensor Networks with Holes," in *Proc. ACM MobiCom*, Montreal, Quebec, Canada, Sept. 2007.

[19] J. Lian, Y. Liu, K. Naik, and L. Chen, "Virtual Surrounding Face Geocasting with Guaranteed Message Delivery for Ad Hoc and Sensor Networks," *IEEE/ACM Transactions on Networking (TON)*, vol. 17, no. 1, Feb. 2009, pp. 200-211.

[20] Q. Fang, J. Gao, and L. Guibas. "Locating and bypassing routing holes in sensor networks," in *Proc. INFOCOM*, vol. 23, pp. 2458-2468, Mar. 2004.

[21] T. Rappaport, "Wireless Communications: Principles and Practice (Second Edition)," *Prentice-Hall PTR*, 2002.

[22] A. E. Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Optimal Throughput-Delay Scaling in Wireless Networks-Part I The Fluid Model," Dept. Elect. Eng., Stanford Univ., Stanford, CA, 2005 [Online]. Available: http://www.standford.edu/ jmammen/