

# Theory and Algorithms for Hop-Count-Based Localization with Random Geometric Graph Models of Dense Sensor Networks

Swaprava Nath, Venkatesan N. E., Anurag Kumar and P. Vijay Kumar  
Department of Electrical Communication Engineering, Indian Institute of Science

---

Wireless sensor networks can often be viewed in terms of a uniform deployment of a large number of nodes in a region of Euclidean space. Following deployment, the nodes self-organize into a mesh topology with a key aspect being *self-localization*. Having obtained a mesh topology in a dense, homogeneous deployment, a frequently used approximation is to take the hop distance between nodes to be proportional to the Euclidean distance between them. In this work, we analyze this approximation through two complementary analyses. We assume that the mesh topology is a random geometric graph on the nodes; and that some nodes are designated as *anchors* with known locations. First, we obtain high probability bounds on the Euclidean distances of *all* nodes that are  $h$  hops away from a fixed anchor node. In the second analysis, we provide a heuristic argument that leads to a direct approximation for the density function of the Euclidean distance between two nodes that are separated by a hop distance  $h$ . This approximation is shown, through simulation, to very closely match the true density function.

Localization algorithms that draw upon the above analyses are then proposed and shown to perform better than some of the well-known algorithms present in the literature. Belief-propagation-based message-passing is then used to further enhance the performance of the proposed localization algorithms. To our knowledge, this is the first usage of message-passing for hop-count-based self-localization.

Categories and Subject Descriptors: F.2.2 [Analysis of Algorithms and Problem Complexity]: Nonnumerical Algorithms and Problems; G.3 [Probability and Statistics]: Stochastic processes

General Terms: Theory, Algorithm, Performance

Additional Key Words and Phrases: Random Geometric Graph, Localization, Belief Propagation

---

## 1. INTRODUCTION

We consider a wireless sensor network comprising a large number of nodes,  $n$ , distributed over a fixed (constant area) region in 2-dimensional Euclidean space, e.g.,

---

A preliminary version of a part of this work has appeared in Proceedings of the 3rd International Conference on Performance Evaluation Methodologies and Tools (Valuetools 2008).

This work was supported by the Defence Research & Development Organisation (DRDO), Ministry of Defence, Government of India under a research grant on wireless sensor networks (DRDO 571, IISc).

Authors' address: Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560012, India

Permission to make digital/hard copy of all or part of this material without fee for personal or classroom use provided that the copies are not made or distributed for profit or commercial advantage, the ACM copyright/server notice, the title of the publication, and its date appear, and notice is given that copying is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or a fee.

© 20YY ACM 1529-3785/20YY/0700-0111 \$5.00

the unit square. If each node is able to communicate perfectly with all the nodes that are at a distance  $\leq r$  from it, and only with these nodes, the communication topology becomes a geometric graph. If the node deployment is random, e.g., uniform i.i.d. deployment (this means that each node is deployed uniformly over the region independently of all other nodes), then the network topology becomes a random geometric graph (RGG) (see, e.g., [Penrose 2003]). Given a deployment of nodes, and a topology over them, a frequently used approximation is to take the minimum number of hops between nodes (i.e., the *hop distance*) as a measure of the Euclidean distance between them. [Niculescu and Nath 2001], [Nagpal et al. 2003] and [Yang et al. 2007] have used this approximation to develop techniques for GPS-free localization in dense wireless sensor networks. Yang *et al.* [Yang et al. 2007], in particular, make the key assumption that the ratio of the Euclidean distance (ED) between a node and two anchor nodes is well approximated by the ratio of the corresponding hop distances<sup>1</sup> (HD). Such an approximation of proportionality between Euclidean and Hop-distance in a random geometric graph still lacks theoretical formalization. This has been pointed out in a recent paper by [Li and Liu 2007], where the authors consider this as a heuristic even for an isotropic network.

For geometric graphs over arbitrary node placements, it is easy to see that HD does not provide a useful measure of ED. Thus, the formal study of the ED-HD relationship on a random geometric graph (RGG) becomes interesting. We take a random deployment of  $n$  nodes on a unit square, and consider the geometric graph on these nodes with radius  $r$ . There are several points on the plane,  $b_l, 1 \leq l \leq L$ , with known locations, designated as *anchors*. There are several nodes that are at a certain HD, say  $h$ , from a fixed anchor node  $b_l$ . Let us denote the random vector of the EDs of these nodes by  $D_{l,i_1}, D_{l,i_2}, \dots, D_{l,i_{M_l}}$ , where  $M_l$  is the random number of such nodes. In Section 4, we are concerned with characterizing the *support* of the *joint distribution* of  $D_{l,i_1}, D_{l,i_2}, \dots, D_{l,i_{M_l}}$ . In particular, we show that, for  $n$  nodes deployed in a given region with fixed area, as  $n \rightarrow \infty$  with probability approaching 1, the support is the intersection of an annulus of width approximately  $r$ , and the unit square.<sup>2</sup> We show that the result holds for critically scaled  $r$  ([Gupta and Kumar 1998]) and for fixed  $r$ , and for different node deployments, e.g., uniform i.i.d. and randomized lattice. To our knowledge this is the first attempt to provide high probability (with probability approaching 1 as  $n \rightarrow \infty$ ) bounds on the Euclidean distances of *all* nodes, that are at given hop distances from beacons. We provide simulation results that illustrate the theory, and serve to show how large  $n$  needs to be for the asymptotics to be useful.

We then turn to studying the *marginal* distribution of the distance of a node from an anchor, given that it is at an HD  $h$  from that anchor. A detailed overview of the existing results on the distribution of Euclidean distance traversed for a given number of hops is provided in Section 5. Then, in this section, we provide a heuristic derivation that replaces the problem of determining the density of the ED from the knowledge of the hop count by a task that is more amenable to

<sup>1</sup>The terms hop count and hop distance will be used interchangeably in this paper.

<sup>2</sup>In the recent literature, [Ozgun et al. 2007], such a network has been called *dense* since the node density increases as  $n \rightarrow \infty$ , in contrast to constant node density networks.

analysis. The resulting approximate derivation of the density function is shown through simulation to fit well with the true density function. The assumptions that we make help us to use the central limit theorem, which explains the reason why the distribution is Gaussian-like.

Next, two localization algorithms are presented, based, respectively, on the high probability bounds on ED, and also on the derivation of the approximate marginal density function discussed above. Through simulation, the algorithms are shown to outperform some of the better-known hop-count-based localization algorithms that have appeared in the literature. Finally, we show how our algorithms can be further improved using *belief-propagation*-based message-passing techniques wherein the messages passed correspond to beliefs based on probability. To the best of our knowledge, this is *the first use of belief-propagation for hop-count based localization techniques*.

The paper is organized as follows. Next in this section, we explain the setting of our problem with a motivating example which also reasons for the assumptions and analyses. Section 2 introduces notation and the geometric-graph setting. In Section 3 we briefly mention that in an arbitrary two-dimensional geometric graph, the HD is not in general a good measure of ED. In Section 4, we present a lower bound to ED given HD that holds with high probability in different settings. The approximate derivation of the density function of ED given HD is presented in Section 5. Localization algorithms based on the theory developed in the previous sections are provided in Section 6. Belief-propagation-based message passing is then used to derive algorithms with improved performance in Section 7. Section 8 draws conclusions.

### 1.1 Motivation for Hop-Distance based Localization

Sensor network localization algorithms can be broadly classified into two categories.

- (i) *Range based localization*: Algorithms under this category assume the availability of noisy distance measurements between the nodes. These require the nodes to be equipped with hardware that enable them to use Received Signal Strength (RSS), Time of Arrival (ToA), Time Difference of Arrival (TDoA), Angle of Arrival (AoA) etc. to obtain an estimate of the true distance between the nodes. There is a plethora of literature that addresses this class of algorithms. Some of the more recent work includes popular multidimensional scaling algorithm [Costa et al. 2006], belief propagation based algorithms [Ihler et al. 2005], iterative algorithms [Khan et al. 2009] etc. We refer the reader to [Langendoen and Reijers 2003] and to [Guvenc and Chong 2009] for a survey of these algorithms.
- (ii) *Range-free localization*: These algorithms do not assume the availability of the necessary hardware to obtain distance measurements. Sensor nodes are only aware of their connectivity to their neighbors, which can be communicated across the network. Hence a sensor node does not have any Euclidean distance estimate and could only possibly know how many hops away it is from another node. An important class of algorithms in this category is the hop-count based algorithms which attempt to estimate the position of the nodes based on only the number of hops in the shortest path between pairs of nodes. Existing literature in this area includes [De et al. 2006], [Niculescu and Nath 2003], [Li and Liu 2007] etc.

The motivation for considering either class of algorithms strongly depends on the application scenario and the capability of the sensor nodes. Range based techniques require that the sensor nodes have additional hardware to estimate the Euclidean distances and need to expend some energy in the estimation process. Proponents of *range free localization*, such as [He et al. 2003], [Hu and Evans 2004], argue that the higher cost and power consumption of the hardware on the inexpensive nodes and irregular signal transmission characteristics of RSSI prohibit the use of range based techniques. Availability of additional distance information would only help, and clearly range based techniques would outperform range free techniques. However we consider scenarios of large scale deployments as in military applications, environmental monitoring, etc., where sensor capabilities would possibly be limited to only communication and physical sensing, and might not include hardware for making distance measurements. In such situations, *range-free localization* would be essential and in this paper we focus on the hop-count based techniques.

## 2. THE GEOMETRIC GRAPH SETTING & NOTATION

In this section we describe the basic setting for our results, and also develop the main notation.

### Setting:

- $n$  nodes are deployed on a unit 2-dimensional area  $\mathcal{A}$ . The node locations are denoted by the vector  $\mathbf{v} = [v_1, v_2, \dots, v_n] \in \mathcal{A}^n$ , where  $v_i$  is the location of the  $i^{\text{th}}$  node.
- We form the geometric graph  $\mathcal{G}(\mathbf{v}, r)$  by connecting nodes that are within the distance  $r$  of each other. Then  $r$  is called the radius of the geometric graph.

We define *anchors* as nodes whose locations are known *a priori*, e.g., in Figure 1, we have shown 4 anchors  $b_1, b_2, b_3$  and  $b_4$ , with their positions fixed at the 4 corners of the unit square  $\mathcal{A}$ . The figure also shows the neighbors of a particular node in geometric graph with radius  $r$ .

### Notation:

- $\mathcal{N} = [n] = \{1, 2, \dots, n\}$ , the index set of the nodes, i.e., node  $i \in \mathcal{N}$  has a location  $v_i$  on  $\mathcal{A}$ .
- $b_l$  = Location of the  $l^{\text{th}}$  anchor node,  $l = 1, \dots, L$ , e.g., in Figure 1,  $L = 4$ .
- $\mathcal{N}_j$  : The neighbor set of node  $j$  in  $\mathcal{G}(\mathbf{v}, r)$ . Mathematically,  $\forall j \in \mathcal{N}, \mathcal{N}_j = \{k \in \mathcal{N} : \|v_j - v_k\| \leq r, k \neq j\}$ .
- $H_{l,i}(\mathbf{v})$  = Minimum number of hops of node  $i$  from anchor  $b_l$  on the graph  $\mathcal{G}(\mathbf{v}, r)$  for the deployment  $\mathbf{v}$ .
- $D_{l,i}(\mathbf{v})$  = Euclidean distance of node  $i$  from anchor  $b_l$  for the deployment  $\mathbf{v}$ .

$$\overline{D}_l(\mathbf{v}, h_l) := \max_{\{i \in \mathcal{N} : H_{l,i}(\mathbf{v}) = h_l\}} D_{l,i}(\mathbf{v})$$

$$\underline{D}_l(\mathbf{v}, h_l) := \min_{\{i \in \mathcal{N} : H_{l,i}(\mathbf{v}) = h_l\}} D_{l,i}(\mathbf{v})$$

Given  $\mathbf{v}$ , these are respectively the maximum and minimum Euclidean distance for a hop-distance  $h_l$  from anchor  $l$ . A graphical illustration of the maximum and minimum EDs is given in Figure 2.

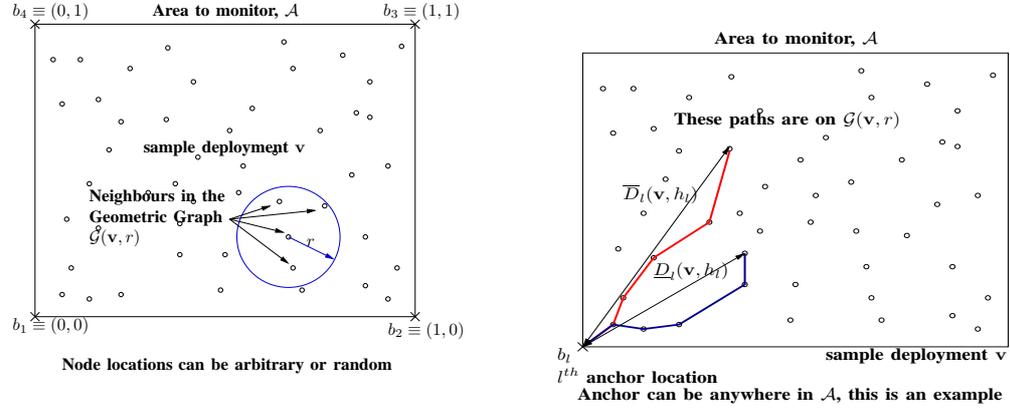


Fig. 1. An example deployment of nodes in a square region  $\mathcal{A}$  with 4 anchors, one at each corner. The neighbors of a particular node in  $\mathcal{G}(\mathbf{v}, r)$  with  $h_l = 5$  are also shown. Fig. 2. Graphical illustration of  $\overline{D}_l(\mathbf{v}, h_l)$  and  $\underline{D}_l(\mathbf{v}, h_l)$  with  $h_l = 5$ . The paths are on  $\mathcal{G}(\mathbf{v}, r)$ .

With this setting, given the hop distance  $h_l$  on  $\mathcal{G}(\mathbf{v}, r)$  between a node and an anchor, we wish to obtain constraints on the Euclidean distance of the node from anchor  $b_l$ .

### 3. HD-ED RELATIONSHIP IN ARBITRARY GEOMETRIC GRAPHS

In this section, we briefly review a result on the HD-ED relation in an *arbitrary* geometric graph with radius  $r$ , where by “arbitrary” we mean that the node locations are arbitrary. It has been shown in [Nath and Kumar 2008] that the proportionality approximation can be arbitrarily coarse in this scenario. The authors have proved the following lemma.

**LEMMA 1.** *For arbitrary  $\mathbf{v}$  and  $h_l \geq 2$ ,  $r < \underline{D}_l(\mathbf{v}, h_l) \leq \overline{D}_l(\mathbf{v}, h_l) \leq h_l r$  and both bounds are sharp.*

Where ‘sharp’ means both bounds are achievable. In [Nath and Kumar 2008], examples are provided that validate the sharpness of the bounds in the above lemma.

However, hop-distance serves as a good measure of the Euclidean distance when the distribution of nodes has positive density over all points on  $\mathcal{A}$ , e.g., the node distribution is uniform i.i.d. or randomized lattice. We will provide support for this assertion via rigorous analyses in the following sections.

### 4. HIGH PROBABILITY BOUNDS ON ED GIVEN THE HD

In this section, we will consider random geometric graphs (RGG) and address the following problem. Given an HD value, say  $h$ , from an anchor node, in general, there are several nodes that are at HD  $h$  from that anchor. Can we provide an interesting characterization of a subregion of the deployment region in which all nodes with an HD of  $h$  from that anchor will lie, with a probability approaching 1, as the number of deployed nodes,  $n$ , increases to  $\infty$ ? Such a characterization is obviously useful for localization.

#### 4.1 Random Geometric Graphs with Critical Scaling of Radius

We now specialize to the following setting.

##### Setting:

- $n$  nodes are deployed on a unit area  $\mathcal{A}$  in the uniform i.i.d. fashion. The difference with this setup from the previous section is that the node locations are random, and are denoted by the random vector  $\mathbf{V} \in \mathcal{A}^n$ , with a particular realization being denoted by  $\mathbf{v}$ . We denote by  $\mathbb{P}^n(\cdot)$  the probability measure on  $\mathcal{A}^n$  so obtained.
- We form the RGG  $\mathcal{G}(\mathbf{v}, r(n))$  by connecting the nodes that are within the radius  $r(n)$  of each other, where  $r(n)$ , the radius of the geometric graph is chosen so that the network remains asymptotically connected. We take  $r(n) = c\sqrt{\frac{\ln n}{n}}$ ,  $c > \frac{1}{\sqrt{\pi}}$ , a constant; this ensures asymptotic connectivity (see [Gupta and Kumar 1998]). For a finite number of nodes, the radius which ensures connectivity depends on the node placements. For a given node placement, we will call this radius as *critical radius* and the graph so obtained as *critical graph*.

In Section 4.1.1, we analyze the distribution of distance from one anchor node and in Section 4.1.2, we generalize it for  $L$  anchors.

The choice of the radius,  $r(n) = c\sqrt{\frac{\ln n}{n}}$ ,  $c > \frac{1}{\sqrt{\pi}}$ , does not only guarantee asymptotic connectivity among the nodes, but also ensures connectivity of the nodes with all the anchors. The following lemma states that there will be at least a node within a distance  $r(n)$  of each anchor  $b_l, l = 1, \dots, L$  w.h.p. and so the nodes are connected to all the anchors for large  $n$ . Define,  $B_l = \{\mathbf{v} : \exists \text{ at least one node within a radius of } r(n) \text{ from } b_l\}$ ,  $l = 1, \dots, L$ .

LEMMA 2.  $\lim_{n \rightarrow \infty} \mathbb{P}^n(\cap_{l=1}^L B_l) = 1$

**Proof:**  $\mathbb{P}^n(\cap_{l=1}^L B_l) = 1 - \mathbb{P}^n(\cup_{l=1}^L B_l^c) \geq 1 - \sum_{l=1}^L \mathbb{P}^n\{B_l^c\} = 1 - \sum_{l=1}^L (1 - \pi r^2(n))^n \geq 1 - Le^{-n\pi r^2(n)} \xrightarrow{n \rightarrow \infty} 1$ , since  $r(n) = c\sqrt{\frac{\ln n}{n}}$  and  $1 - x \leq e^{-x}$ . ■

4.1.1 *High probability ED bound given HD from an anchor  $b_l$ : Uniform i.i.d. deployment.* We make the construction as shown in Figure 3. From  $b_l$  (without loss of generality, we can choose  $l = 1$ ), we draw a circle of radius  $h_l r(n)$  centered at  $b_l$ , this is the maximum distance reachable in  $h_l$  hops, by triangle inequality, since each hop can be of maximum length  $r(n)$ . All the nodes  $\{i \in \mathcal{N} : H_{l,i}(\mathbf{v}) = h_l\}$  lie within this disk. So,  $\overline{D}_l(\mathbf{v}, h_l) \leq h_l r(n)$  for all  $\mathbf{v}$ . To obtain a lower bound on  $\underline{D}_l(\mathbf{v}, h_l)$ , we construct blades as shown in Figure 3. We start with one blade. It will cover some portion of the circumference of the circle of radius  $h_l r(n)$ ; see Figure 3. Construct the next blade so that it covers the adjacent portion of the circumference that has not been covered by the previous blade. We go on constructing these blades until the entire portion of the circle lying inside the unit square  $\mathcal{A}$  is covered (see Figure 3). Let us define,

- $J(n)$  : Number of blades required to cover the part of the circle within  $\mathcal{A}$ .
- $\mathcal{B}_j^l$  :  $j^{\text{th}}$  blade drawn from the point  $b_l$  as shown in Figure 3,  $1 \leq j \leq J(n)$ .

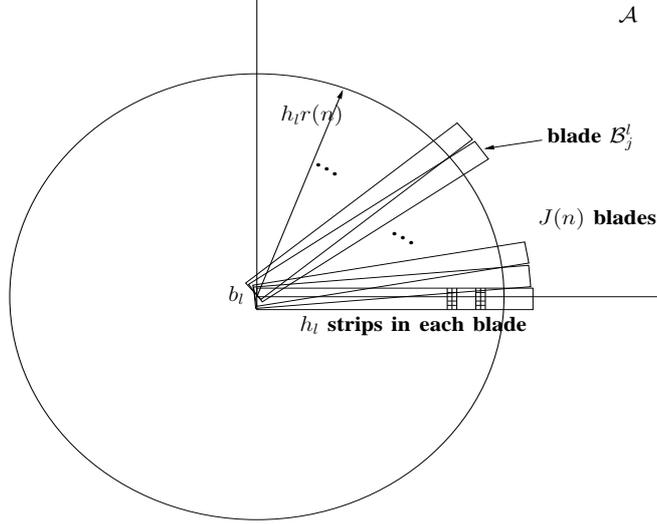


Fig. 3. Construction using the blades cutting the circumference of the circle of radius  $h_l r(n)$ .

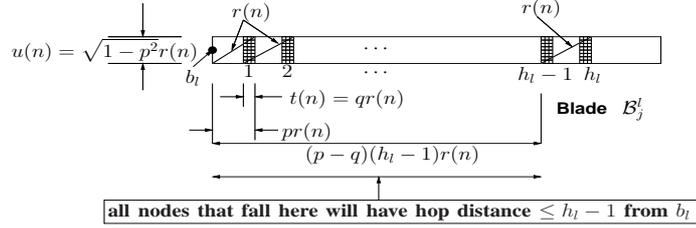


Fig. 4. The construction with  $h_l$  hops.

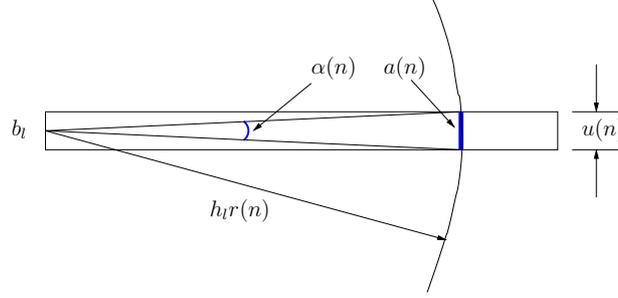
On each of these blades, we construct  $h_l$  strips<sup>3</sup>, shown shaded in Figure 4,  $u(n)$  being the width of the blade and  $t(n)$  the width of the strip. We define the following event.

$$-A_{i,j}^l = \{\mathbf{v} : \exists \text{ at least one node in the } i^{\text{th}} \text{ strip of } \mathcal{B}_j^l\}$$

If a  $\mathbf{v} \in A_{i,j}^l$ ,  $\forall 1 \leq i \leq h_l - 1, 1 \leq j \leq J(n)$ , i.e., there exists at least one node in each of the  $h_l - 1$  strips (see Figure 4) for all the blades  $\mathcal{B}_j^l$ , then for that  $\mathbf{v}$ , all nodes at a distance  $< (p - q)(h_l - 1)r(n)$  from  $b_l$  are reachable in at most  $h_l - 1$  hops, hence will have a hop distance  $\leq h_l - 1 < h_l$ . So, we have  $\underline{D}_l(\mathbf{v}, h_l) \geq (p - q)(h_l - 1)r(n)$ , for such a deployment  $\mathbf{v}$ ; see Figure 4. Hence,

$$\begin{aligned} & \{\cap_{j=1}^{J(n)} \cap_{i=1}^{h_l-1} A_{i,j}^l\} \\ & \subseteq \{\mathbf{v} : (p - q)(h_l - 1)r(n) \leq \underline{D}_l(\mathbf{v}, h_l) \leq \overline{D}_l(\mathbf{v}, h_l) \leq h_l r(n)\} \end{aligned} \quad (1)$$

<sup>3</sup>A construction with improved convergence rate based on lens-shaped areas rather than rectangular strips is presented in Appendix A.

Fig. 5. Construction to find  $J(n)$ .

Since  $1 > p > q > 0$ , we can choose  $p - q$  to be equal to  $1 - \epsilon$ , for the given  $\epsilon > 0$ . So the lower bound in Equation 1 becomes,  $(p - q)(h_l - 1)r(n) = (1 - \epsilon)(h_l - 1)r(n)$ . To find the value of  $J(n)$ , we need to define the following.

- $a(n)$  is the length of the arc of radius  $h_l r(n)$  that lies within a blade, drawn taking  $b_l$  as center, as shown in Figure 5.
- $\alpha(n)$  : angle subtended by  $a(n)$  at  $b_l$ , see Figure 5.

Now from Figure 3, we have,  $J(n) = \left\lceil \frac{\pi}{2\alpha(n)} \right\rceil$ . We also have from Figure 5,  $h_l r(n)\alpha(n) = a(n) \geq u(n) = \sqrt{1 - p^2}r(n)$ . Hence,  $\alpha(n) \geq \frac{\sqrt{1 - p^2}}{h_l}$ . So,  $J(n) \leq \left\lceil \frac{\pi h_l}{2\sqrt{1 - p^2}} \right\rceil$ . Now we compute,

$$\begin{aligned}
& \mathbb{P}^n \left( \bigcap_{j=1}^{J(n)} \bigcap_{i=1}^{h_l-1} A_{i,j}^l \right) \\
&= 1 - \mathbb{P}^n \left( \bigcup_{j=1}^{J(n)} \bigcup_{i=1}^{h_l-1} A_{i,j}^{l,c} \right) \\
&\geq 1 - \sum_{j=1}^{J(n)} \sum_{i=1}^{h_l-1} \mathbb{P}^n \left( A_{i,j}^{l,c} \right) \\
&\geq 1 - \left\lceil \frac{\pi h_l}{2\sqrt{1 - p^2}} \right\rceil (h_l - 1)(1 - u(n)t(n))^n \\
&\geq 1 - \left\lceil \frac{\pi h_l}{2\sqrt{1 - p^2}} \right\rceil (h_l - 1)e^{-nu(n)t(n)} \\
&= 1 - \left\lceil \frac{\pi h_l}{2\sqrt{1 - p^2}} \right\rceil (h_l - 1)e^{-nq\sqrt{1 - p^2}r^2(n)} \\
&\xrightarrow{n \rightarrow \infty} 1
\end{aligned} \tag{2}$$

The first inequality comes from the union bound, the second inequality, from the upper bound on  $J(n)$ . The third inequality uses the result  $1 - x \leq e^{-x}$ .

Let us define,  $E_{h_l}(n) = \{\mathbf{v} : (1-\epsilon)(h_l-1)r(n) \leq \underline{D}_l(\mathbf{v}, h_l) \leq \overline{D}_l(\mathbf{v}, h_l) \leq h_l r(n)\}$ . So, we have, for the given  $\epsilon > 0$ , and using Equations 1 and 2,

$$\begin{aligned} 1 &\geq \mathbb{P}^n(E_{h_l}(n)) \\ &\geq \mathbb{P}^n\left(\bigcap_{j=1}^{J(n)} \bigcap_{i=1}^{h_l-1} A_{i,j}^l\right) \\ &\geq 1 - (h_l - 1) \left[ \frac{\pi h_l}{2\sqrt{1-p^2}} \right] e^{-nq\sqrt{1-p^2}c^2 \frac{\ln n}{n}} \end{aligned} \quad (3)$$

which implies,

$$0 \leq 1 - \mathbb{P}^n(E_{h_l}(n)) \leq (h_l - 1) \left[ \frac{\pi h_l}{2\sqrt{1-p^2}} \right] e^{-q\sqrt{1-p^2}c^2 \ln n} \quad (4)$$

And as  $n \rightarrow \infty$ ,

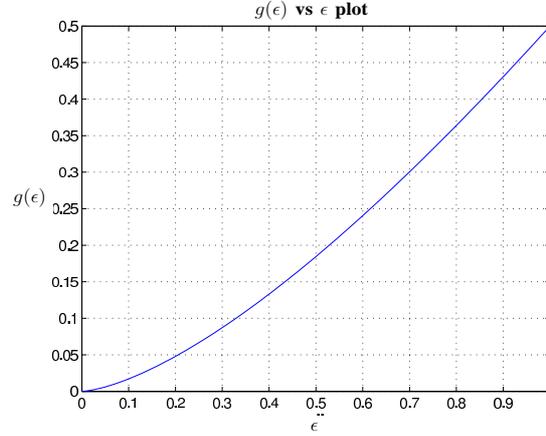
$$\begin{aligned} 1 - \mathbb{P}^n(E_{h_l}(n)) &= \mathcal{O}\left(e^{-q\sqrt{1-p^2}c^2 \ln n}\right) \\ &= \mathcal{O}\left(\frac{1}{n^{q\sqrt{1-p^2}c^2}}\right) \end{aligned} \quad (5)$$

This result is true for any  $p$  and  $q$ . But we can choose these constants so that the convergence  $\rightarrow 0$  of the bound in Equation 5 is the most rapid, i.e.,  $p$  and  $q$  are chosen so as to maximize  $q\sqrt{1-p^2}$ , thus making the upper bound to reduce at the fastest rate. For the given  $\epsilon > 0$ ,  $p - q = 1 - \epsilon \Rightarrow q = p - (1 - \epsilon)$ . We can show that,  $p = \arg \max_p (p - (1 - \epsilon))\sqrt{1-p^2} = \frac{1-\epsilon+\sqrt{(1-\epsilon)^2+8}}{4}$ ,  $q = \frac{-3(1-\epsilon)+\sqrt{(1-\epsilon)^2+8}}{4}$ . Then writing,  $g(\epsilon) = q(\epsilon)\sqrt{1-p^2(\epsilon)}$ , we obtain the following theorem,

**THEOREM 1.** For a given  $1 > \epsilon > 0$ , and  $r(n) = c\sqrt{\frac{\ln n}{n}}$ ,  $c > \frac{1}{\sqrt{\pi}}$ ,  $\mathbb{P}^n(E_{h_l}(n)) = 1 - \mathcal{O}\left(\frac{1}{n^{g(\epsilon)c^2}}\right)$ ,  
 where  $g(\epsilon) = q(\epsilon)\sqrt{1-p^2(\epsilon)}$ ,  
 $p(\epsilon) = \frac{1-\epsilon+\sqrt{(1-\epsilon)^2+8}}{4}$ ,  $q(\epsilon) = \frac{-3(1-\epsilon)+\sqrt{(1-\epsilon)^2+8}}{4}$ .

**Remark 1:** *Convergence rate in the above result:* A plot of  $g(\epsilon)$  vs  $\epsilon$  is given in Figure 6. We see that  $g(\epsilon) \downarrow 0$  as  $\epsilon \downarrow 0$ . Hence Theorem 1 says that  $\lim_{n \rightarrow \infty} \mathbb{P}^n(E_{h_l}(n)) = 1$ , for any  $1 > \epsilon > 0$ , so we can expect a node having a HD of  $h_l$  from anchor  $b_l$  to be within a distance  $[(1 - \epsilon)(h_l - 1)r(n), h_l r(n)]$  from  $b_l$  in a dense network. We notice that the width of this band of uncertainty is roughly  $r(n)$ , which is the unit of distance measurement on  $\mathcal{G}(\mathbf{v}, r(n))$ . The theorem also says that the rate of convergence is governed by the  $\epsilon$  chosen, i.e., the smaller the  $\epsilon$ , the slower the rate of convergence.

**Remark 2:** *An alternate construction:* The rate of convergence of  $\mathbb{P}^n(E_{h_l}(n))$  to unity can be made more rapid through an improved construction that replaces the strips used in the discussion here with lens-shaped areas derived from the intersection of circles as shown in Appendix A. While the appendix deals with lower

Fig. 6.  $g(\epsilon)$  vs  $\epsilon$  plot.

bounds to the ED of a single node based on its known HD, the derivation can be extended to provide a global lower bound to the ED between all pairs of nodes in a deployment that are separated by the same HD.

**Remark 3:** *Non-homogeneous node deployment:* If the node distribution was non-homogeneous with positive density over all points in  $\mathcal{A}$ , the term  $(1-u(n)t(n))^n$  in Equation 2 could have been replaced by  $(1-f_{min}u(n)t(n))^n$ , where  $f_{min}$  is the minimum density over  $\mathcal{A}$  and as  $f_{min} > 0$ , the same convergence result would be true even for non-homogeneous node placement.

**Remark 4:** *Random node failures:* After the network is set up, nodes may fail. Let  $\gamma$  denote the probability that a node is ‘good’, i.e., does not fail. From the derivation in Equation 2 it can be seen that the negative exponent in the right hand side of Equation 4 just gets multiplied by  $\gamma$ , thus not affecting the ensuing theorem. The intuition is simple, in the limit as  $n \rightarrow \infty$  there are enough nodes in the ‘strips’, so that in spite of failures, the probability of finding at least one path to the anchor still approaches 1.

**4.1.2 High probability ED bound given HD from  $L$  anchors  $b_l, l = 1, \dots, L$ : Uniform i.i.d. deployment.** For  $L$  anchors, the question arises whether the hop distances from the  $L$  anchors are *feasible* or not, e.g., if we denote a disk with center  $a$  and radius  $r$ , by  $C(a, r) = \{z \in \mathcal{A} : \|z - a\| \leq r\}$ , then a necessary condition for a *feasible*  $\mathbf{h}$  vector ( $\mathbf{h} = [h_1, \dots, h_l, \dots, h_L] \in \mathbb{N}^L$  is the hop distance vector) is that  $\cap_{l=1}^L C(b_l, h_l r(n)) \neq \emptyset$  (there will be other feasibility conditions also). We denote the set of all *feasible*  $\mathbf{h}$  vectors by  $\mathcal{H}(n)$  (note that the feasibility of an  $\mathbf{h}$  vector depends on  $n$ ). We see that  $\forall \mathbf{h} \in \mathcal{H}(n), \cap_{l=1}^L E_{h_l}(n) \supseteq \cap_{l=1}^L \cap_{j=1}^{J(n)} \cap_{i=1}^{h_l-1} A_{i,j}^l$ ,

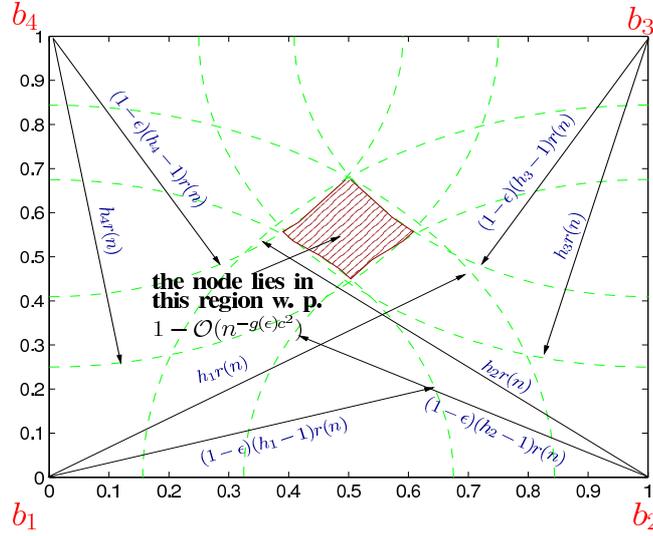


Fig. 7. Graphical illustration of how Theorem 2 yields a location region for a node that is at a HD  $h_l$  from anchor  $b_l$ ,  $1 \leq l \leq 4$ .

which implies that (analyzing similar to Equation 2),

$$\begin{aligned}
 & \mathbb{P}^n \left( \bigcap_{l=1}^L E_{h_l}(n) \right) \\
 & \geq \mathbb{P}^n \left( \bigcap_{l=1}^L \bigcap_{j=1}^{J(n)} \bigcap_{i=1}^{h_l-1} A_{i,j}^l \right) \\
 & \geq 1 - \sum_{l=1}^L (h_l - 1) \left[ \frac{\pi h_l}{2\sqrt{1-p^2}} \right] n^{-q\sqrt{1-p^2}c^2} \\
 & \Rightarrow \mathbb{P}^n \left( \bigcap_{l=1}^L E_{h_l}(n) \right) = 1 - \mathcal{O} \left( n^{-q\sqrt{1-p^2}c^2} \right)
 \end{aligned}$$

Hence we get the following theorem,

**THEOREM 2.** For a given  $1 > \epsilon > 0$ , and  $r(n) = c\sqrt{\frac{\ln n}{n}}$ ,  $c > \frac{1}{\sqrt{\pi}}$ ,  $\forall \mathbf{h} = [h_1, \dots, h_l, \dots, h_L] \in \mathcal{H}(n)$ ,

$$\mathbb{P}^n \left( \bigcap_{l=1}^L E_{h_l}(n) \right) = 1 - \mathcal{O} \left( \frac{1}{n^{g(\epsilon)c^2}} \right)$$

where  $g(\epsilon) = q(\epsilon)\sqrt{1-p^2(\epsilon)}$ ,

$$p(\epsilon) = \frac{1-\epsilon+\sqrt{(1-\epsilon)^2+8}}{4}, q(\epsilon) = \frac{-3(1-\epsilon)+\sqrt{(1-\epsilon)^2+8}}{4}$$

This theorem tells us that for a feasible  $\mathbf{h}$ , the node lies within the intersection of the annuli of inner and outer radii  $(1-\epsilon)(h_l-1)r(n)$  and  $h_l r(n)$  respectively, centered at anchors  $b_l$ ,  $1 \leq l \leq L$ , with a probability that scales as shown in the above theorem. A graphical illustration of this is shown in Figure 7 for  $L = 4$ . The concentric circles denote the ED bounds given the HD from a certain anchor, and

the shaded intersection of these annuli denote the location of the node with high probability. This result motivates us to develop localization schemes that use the hop-distance information from a few fixed anchor nodes which has been described in the Section 6.

#### 4.2 Random Geometric Graphs with Fixed Radius

The scaling of  $r(n)$  with  $n$  as shown in the previous section ensures asymptotic connectivity and increases the precision in localization as  $n \rightarrow \infty$ . But in a wireless sensor network the radius  $r$  of the RGG on which hop-distances are measured often corresponds to the radio range for a given transmit power, and hence does not decrease with  $n$ . So, it is meaningful to use a fixed radius  $r$  for the RGG and it is denoted by  $\mathcal{G}(\mathbf{v}, r)$ . But for connectivity, we need to use number of nodes sufficient to make the network connected (i.e., the radius should scale with  $n$  like  $r(n) = c\sqrt{\frac{\ln n}{n}}$ ,  $c > \frac{1}{\sqrt{\pi}}$ , a constant; see [Gupta and Kumar 1998]), i.e., need at least  $n_0 = \inf\{n : r(n) \leq r\}$  nodes. Using a constant value for radius  $r$ , and redefining  $E_{h_l} = \{\mathbf{v} : (1 - \epsilon)(h_l - 1)r \leq \underline{D}_l(\mathbf{v}, h_l) \leq \overline{D}_l(\mathbf{v}, h_l) \leq h_l r\}$ , where the hop distance is measured on the RGG  $\mathcal{G}(\mathbf{v}, r)$ , we can show (along similar lines as for Equation 2),

$$\begin{aligned} 1 &\geq \mathbb{P}^n(E_{h_l}) \\ &\geq \mathbb{P}^n\left(\bigcap_{j=1}^J \bigcap_{i=1}^{h_l-1} A_{i,j}^l\right) \\ &\geq 1 - (h_l - 1) \left[ \frac{\pi h_l}{2\sqrt{1-p^2}} \right] e^{-nq\sqrt{1-p^2}r^2} \end{aligned} \quad (6)$$

where  $J \leq \left\lceil \frac{\pi h_l}{2\sqrt{1-p^2}} \right\rceil$ . Which implies, as  $n \rightarrow \infty$ ,

$$\begin{aligned} 1 - \mathbb{P}^n(E_{h_l}) &= \mathcal{O}\left(e^{-nq\sqrt{1-p^2}r^2}\right) \end{aligned} \quad (7)$$

Hence,  $\lim_{n \rightarrow \infty} \mathbb{P}^n(E_{h_l}) = 1$ . So, for  $L$  anchors, we will get  $\forall \mathbf{h} \in \mathcal{H}$  (note that the set of feasible  $\mathbf{h}$  vectors,  $\mathcal{H}$ , does not scale with  $n$  in this case),

$$\begin{aligned} 1 - \mathbb{P}^n(\bigcap_{l=1}^L E_{h_l}) &= \mathcal{O}\left(e^{-nq\sqrt{1-p^2}r^2}\right) \end{aligned} \quad (8)$$

Hence we get the following theorem.

**THEOREM 3.** *For a given  $1 > \epsilon > 0$ , and  $r$  fixed,  $\forall n \geq n_0 = \inf\{n : r(n) \leq r\}$ ,  $\forall \mathbf{h} = [h_1, \dots, h_l, \dots, h_L] \in \mathcal{H}$ ,*

$$\mathbb{P}^n(\bigcap_{l=1}^L E_{h_l}) = 1 - \mathcal{O}\left(e^{-ng(\epsilon)r^2}\right)$$

where  $g(\epsilon) = q(\epsilon)\sqrt{1-p^2(\epsilon)}$ ,

$$p(\epsilon) = \frac{1-\epsilon+\sqrt{(1-\epsilon)^2+8}}{4}, q(\epsilon) = \frac{-3(1-\epsilon)+\sqrt{(1-\epsilon)^2+8}}{4}$$

■

**Remark:** We see that,

$$\text{for all } \mathbf{h} \in \mathcal{H}, \lim_{n \rightarrow \infty} \mathbb{P}^n \left( \bigcap_{l=1}^L E_{h_l} \right) = 1$$

but with an *exponential convergence* rate compared to the power law scaling in the previous section. But it also says that the precision of localization remains fixed at  $r$  rather than increasing with  $n$  like in the previous section.

### 4.3 RGGs on Randomized Lattice Node Deployment

In the previous sections we analyzed the performance of ED-HD proportionality approximation for uniform i.i.d. deployment. In this section we will prove a similar result for the randomized lattice deployment. In randomized lattice node deployment, the unit area is split into  $n$  cells each of area  $\frac{1}{n}$ , and in each cell exactly one node is deployed, uniformly over the cell area. The locations of the nodes in two different cells are independent of each other. We denote by  $\mathbb{P}_{RL}^{(n)}(\cdot)$  the probability measure on  $\mathcal{A}^n$  so obtained (this is different from the uniform i.i.d. measure  $\mathbb{P}^n(\cdot)$ ). We will show that, for this deployment also the above theorems hold. Here we consider the case in which the radius  $r(n)$  of the RGG scales with  $n$  as defined before. For fixed  $r$ , the theorem is valid too, which can be proved in a similar way as done in Section 4.2.

We have the following notation,

— $S_k^{i,j}$ : area belonging to the  $i^{\text{th}}$  strip of  $j^{\text{th}}$  blade (refer to Figures 3 and 4) of area  $u(n)t(n)$  that falls in the  $k^{\text{th}}$  cell of the randomized lattice structure.

Thus,  $\sum_{k=1}^n S_k^{i,j} = u(n)t(n)$ ,  $\forall 1 \leq i \leq h_l - 1, 1 \leq j \leq J(n)$ . Since a single node is uniformly distributed over each cell whose area is  $\frac{1}{n}$ ,

$$\mathbb{P}_{RL}^n \left( A_{i,j}^{l,c} \right) = \prod_{k=1}^n \left( 1 - \frac{S_k^{i,j}}{\frac{1}{n}} \right) = \prod_{k=1}^n \left( 1 - nS_k^{i,j} \right)$$

We see that,

$$\sum_{k=1}^n \left( 1 - nS_k^{i,j} \right) = n(1 - u(n)t(n)) \quad \begin{array}{l} \forall 1 \leq i \leq h_l - 1 \\ \forall 1 \leq j \leq J(n) \end{array}$$

Now, we know that the arithmetic mean is no smaller than the geometric mean. It follows that,

$$\begin{aligned} \mathbb{P}_{RL}^n \left( A_{i,j}^{l,c} \right) &= \prod_{k=1}^n \left( 1 - nS_k^{i,j} \right) \leq \left( \frac{1}{n} \sum_{k=1}^n \left( 1 - nS_k^{i,j} \right) \right)^n \\ &= (1 - u(n)t(n))^n \end{aligned} \quad (9)$$

Hence we get (analyzing similar to Equation 2, 3, 4 and 5) the following theorem,

**THEOREM 4.** For a given  $1 > \epsilon > 0$ , and  $r(n) = c\sqrt{\frac{\ln n}{n}}$ ,  $c > \frac{1}{\sqrt{\pi}}$ ,  $\mathbb{P}_{RL}^n(E_{h_l}(n)) = 1 - \mathcal{O}\left(\frac{1}{n^{g(\epsilon)c^2}}\right)$ ,

where  $g(\epsilon) = q(\epsilon)\sqrt{1 - p^2(\epsilon)}$ ,

$$p(\epsilon) = \frac{1 - \epsilon + \sqrt{(1 - \epsilon)^2 + 8}}{4}, \quad q(\epsilon) = \frac{-3(1 - \epsilon) + \sqrt{(1 - \epsilon)^2 + 8}}{4}.$$

Hence,  $\lim_{n \rightarrow \infty} \mathbb{P}_{RL}^n(E_{h_l}(n)) = 1$ . Following a similar analysis as in Section 4.1.2 for  $L$  anchors, we can state the following theorem for randomized lattice node deployment. ■

**THEOREM 5.** For a given  $1 > \epsilon > 0$ , and  $r(n) = c\sqrt{\frac{\ln n}{n}}$ ,  $c > \frac{1}{\sqrt{\pi}}$ ,  $\forall \mathbf{h} = [h_1, \dots, h_l, \dots, h_L] \in \mathcal{H}(n)$ ,

$$\mathbb{P}_{RL}^n(\cap_{l=1}^L E_{h_l}(n)) = 1 - \mathcal{O}\left(\frac{1}{ng(\epsilon)c^2}\right)$$

where  $g(\epsilon) = q(\epsilon)\sqrt{1-p^2(\epsilon)}$ ,

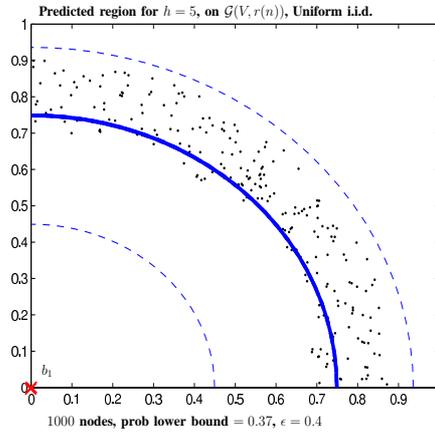
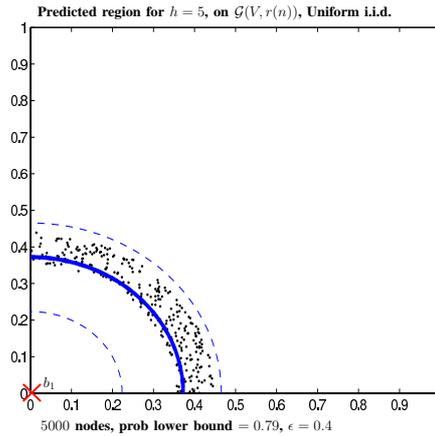
$$p(\epsilon) = \frac{1-\epsilon+\sqrt{(1-\epsilon)^2+8}}{4}, q(\epsilon) = \frac{-3(1-\epsilon)+\sqrt{(1-\epsilon)^2+8}}{4}$$

#### 4.4 RGGs with Poisson Distributed Number of Nodes, Uniform i.i.d. Deployment

Here we consider another kind of deployment, where we pick the number of nodes with the distribution  $\text{Poisson}(n)$  and deploy these nodes uniformly over the area  $\mathcal{A}$ . The number of nodes falling in  $\mathcal{A}$  is a random variable with mean  $n$ , and since we are throwing the picked nodes uniformly over  $\mathcal{A}$ , the nodes falling in disjoint areas are independent and Poisson distributed with rate proportional to the area considered. Hence, for disjoint strips with area  $u(n)t(n)$  each and the number of node selection being  $\text{Poisson}(n)$ , the number of nodes falling in each strip is  $\text{Poisson}(nu(n)t(n))$ , independent and identically distributed. Let the probability law associated with this kind of deployment be denoted by  $\mathbb{P}_{Po}^n(\cdot)$ . Let us focus our attention to a certain blade  $\mathcal{B}_j^l$  as shown in Figure 4 pivoted at the anchor location  $b_l$ . We also denote the maximum and minimum Euclidean distance traveled by a  $h_l$  hop path within this blade by  $\overline{D}_l^{\mathcal{B}_j^l}(\mathbf{v}, h_l)$  and  $\underline{D}_l^{\mathcal{B}_j^l}(\mathbf{v}, h_l)$  respectively. Now, ensuring at least one node in each of the  $h_l - 1$  strips of  $\mathcal{B}_j^l$  will ensure the event  $E_{h_l}^{\mathcal{B}_j^l}(n) = \{\mathbf{v} : (1-\epsilon)(h_l-1)r(n) \leq \underline{D}_l^{\mathcal{B}_j^l}(\mathbf{v}, h_l) \leq \overline{D}_l^{\mathcal{B}_j^l}(\mathbf{v}, h_l) \leq h_l r(n)\}$  also occurs. So, we have for the given  $\epsilon > 0$ ,

$$\begin{aligned} 1 &\geq \mathbb{P}_{Po}^n(E_{h_l}^{\mathcal{B}_j^l}(n)) \\ &\geq \mathbb{P}_{Po}^n\left(\cap_{i=1}^{h_l-1} A_{i,j}^l\right) \\ &= \left(1 - e^{-nu(n)t(n)}\right)^{h_l-1} \\ &= \left(1 - n^{-c^2q\sqrt{1-p^2}}\right)^{h_l-1} \end{aligned} \quad (10)$$

The second inequality comes because  $\{\mathbf{v} \in \cap_{i=1}^{h_l-1} A_{i,j}^l\} \subseteq \{\mathbf{v} \in E_{h_l}^{\mathcal{B}_j^l}(n)\}$  and the first equality comes because of the independence of the number of nodes due to Poisson deployment and disjoint strips. Since in this deployment we are not using the union bound, the expression for probability is exact. Hence the bound on the probability of the event  $E_{h_l}^{\mathcal{B}_j^l}(n)$  is tighter, yet the rate of convergence follows the power law ( $e^{-nu(n)t(n)} = n^{-q\sqrt{1-p^2}c^2}$ ).


 Fig. 8. 1000 nodes,  $\mathbb{P}^n(E_1(n)) \geq 0.37$ .

 Fig. 9. 5000 nodes,  $\mathbb{P}^n(E_1(n)) \geq 0.79$ .

Variation with the number of nodes: locations of nodes that are 5 hops away from an anchor ( $b_1$ ) at the origin (marked as  $\times$ ). The dots denote the true node locations for such a deployment with the hop distance as shown. The thin dashed lines show the ED bounds given by Theorem 1, the thick solid line shows ED  $(h_1 - 1)r(n)$  from  $b_1$ ;  $\epsilon = 0.4$ ,  $r(n) = \frac{4}{\sqrt{\pi}} \sqrt{\frac{\ln n}{n}}$ .

#### 4.5 Simulation Results

In this section, we illustrate Theorem 1 through some simulation examples. We deploy  $n$  nodes in the uniform i.i.d. fashion on the unit square  $\mathcal{A}$ , and form the geometric graph  $\mathcal{G}(\mathbf{v}, r(n))$ , where  $r(n) = \frac{4}{\sqrt{\pi}} \sqrt{\frac{\ln n}{n}}$ . We also have 4 anchors at the 4 corners of  $\mathcal{A}$ .

$n$	$r(n)$	EDLB	$\underline{D}_1$	$\overline{D}_1$	EDUB	PLB	EP
1000	0.1876	0.4494	0.6934	0.9053	0.9362	0.37	1
2000	0.1391	0.3336	0.5196	0.6678	0.6950	0.61	1
3000	0.1166	0.2796	0.4313	0.5590	0.5826	0.70	1
4000	0.1028	0.2465	0.3761	0.4929	0.5136	0.75	1
5000	0.0931	0.2235	0.3428	0.4559	0.4655	0.79	1
6000	0.0859	0.2062	0.3123	0.4191	0.4295	0.81	1

Table I. Euclidean Distance Lower Bound (EDLB) =  $(1 - \epsilon)(h_1 - 1)r(n)$  and Euclidean Distance Upper Bound (EDUB) =  $h_1 r(n)$  are found from Theorem 1.  $\underline{D}_1$  and  $\overline{D}_1$  are the maximum and minimum EDs from anchor 1 given the hop-distance  $h_1 = 5$ . The theoretical Probability Lower Bound (PLB) =  $1 - (h_1 - 1) \left[ \frac{\pi h_1}{2\sqrt{1-p^2(\epsilon)}} \right] e^{-ng(\epsilon)r^2(n)}$ , and the Empirical Probability (EP) is found from this experiment.  $r(n) = \frac{4}{\sqrt{\pi}} \sqrt{\frac{\ln n}{n}}$ ,  $\epsilon = 0.4$ .

#### Illustration of Theorem 1 with increasing $n$ for a fixed $\epsilon$ and HD:

We fix  $\epsilon = 0.4$  and hop-distance  $h_1 = 5$  from anchor  $b_1$  located at the bottom-left corner of the unit square  $\mathcal{A}$ . The results are summarized in Table I and illustrate how the theoretical bounds given in Theorem 1 become tighter as we increase the number of nodes  $n$ , keeping the hop-distance  $h_1$  and  $\epsilon$  fixed.

Figures 8 and 9 show the theoretical bounds given by Theorem 1, and only those nodes are shown that have a hop-distance  $h_1 = 5$  from anchor  $b_1$ , for 1000 and 5000 nodes, respectively. We notice that, for this range of values of  $n$ , while the maximum Euclidean distance,  $\overline{D}_1$ , is quite close to the upper bound (obtained from the triangle inequality), the lower bound is loose when compared to the minimum Euclidean distance,  $\underline{D}_1$ . Indeed, all the node locations lie well within the bounds, and, in fact,  $(h - 1)r(n)$  (the thick solid quarter circle in the figures) could serve as a good approximation to  $\underline{D}_1$ , but this bound is certainly not met with a high probability. The theoretical lower bound on the probability of the upper and lower bounds being respected is seen to be increasing to 1 as the number of nodes,  $n$ , is increased.

**Remark:** The looseness of the lower bound on the Euclidean distance is because of the way the bound is obtained. First, the condition employed in the construction of the lower bound in Figure 4 is only a sufficient one. Moreover, we have used the union bound to get the bound on probability, which further weakens the bound.

#### Illustration of Theorem 1 with decreasing HD for a fixed $n$ and a fixed lower bound on probability:

We have fixed the number of nodes  $n = 5000$  and also fixed the lower bound on probability that the node lies within the bound of  $[(1 - \epsilon)(h_1 - 1)r(n), h_1 r(n)]$  (as given by Theorem 1) at 0.80. Figures 10, 11 and 12 show that as we decrease the hop-distance  $h_1$ , the bound on the ED becomes tighter, which implies that if we keep the lower bound fixed, the  $\epsilon$  that achieves that lower bound will be smaller for smaller hop-distances, as predicted by Theorem 1.

#### Illustration of convergence in probability for the geometric graph with fixed radius:

We fix the radius of the graph  $\mathcal{G}(\mathbf{v}, r)$ ,  $r = 0.1$  and take  $h_1 = 5$ ,  $\epsilon = 0.36$ . The

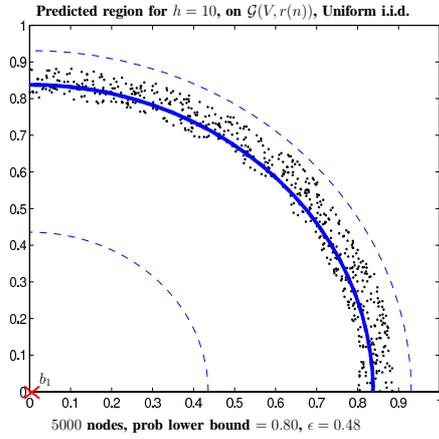
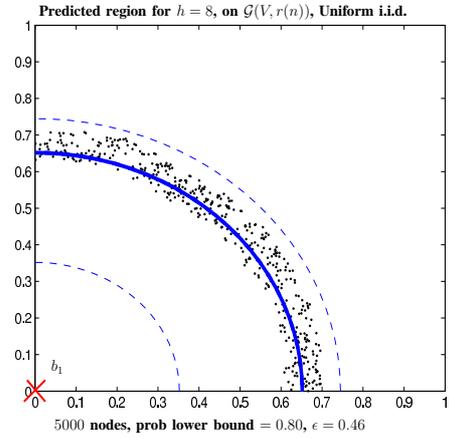
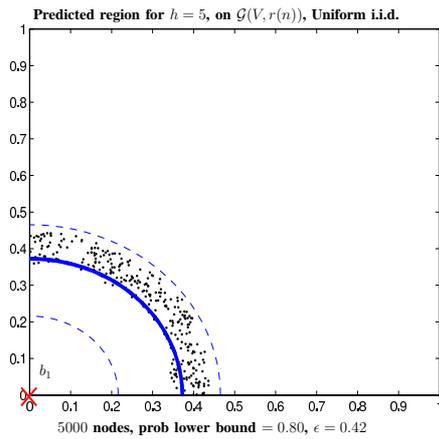

 Fig. 10.  $h_1 = 10, \epsilon = 0.48$ .

 Fig. 11.  $h_1 = 8, \epsilon = 0.46$ .

 Fig. 12.  $h_1 = 5, \epsilon = 0.42$ .

Fig. 13. Variation with hop distance: 5000 nodes were deployed in a uniform i.i.d. fashion. The dots denote the true node locations for such a deployment with the hop distance (from the anchor  $b_1$  marked as  $\times$ ) as shown. The thin dashed lines show the ED bounds given by Theorem 1, the thick solid line shows ED  $(h_1 - 1)r(n)$  from  $b_1$ ;  $\mathbb{P}^n(E_1(n)) \geq 0.80$ .  $r(n) = \frac{4}{\sqrt{\pi}} \sqrt{\frac{\ln n}{n}}$ .

simulation results are summarized in Table II, which shows that for smaller  $n$ , the lower bound of probability (as given by Equation 6) is weak, but the convergence rate, due to its exponential nature, is very rapid with increase in  $n$ .

$n$	EDLB	$\underline{D}_1$	$\overline{D}_1$	EDUB	PLB	EP
1000	0.2560	0.3483	0.4574	0.5000	0.0000	1
2000	0.2560	0.3479	0.4677	0.5000	0.0000	1
3000	0.2560	0.3775	0.4835	0.5000	0.0000	1
4000	0.2560	0.3741	0.4843	0.5000	0.2906	1
5000	0.2560	0.3831	0.4897	0.5000	0.7733	1
6000	0.2560	0.3705	0.4826	0.5000	0.9275	1

Table II. Radius  $r = 0.1$ ,  $h_1 = 5$ ,  $\epsilon = 0.36$ . The theoretical PLB =  $1 - (h_1 - 1) \left[ \frac{\pi h_1}{2\sqrt{1-p^2(\epsilon)}} \right] e^{-ng(\epsilon)r^2}$ . Abbreviations are as defined in Table I.

## 5. DISTRIBUTION OF PAIRWISE ED GIVEN THE HD

In Section 4 we provided a characterization of the support of the *joint* distribution of the locations of all nodes that are at a given HD from an anchor. In this section we turn our attention to a (more refined) characterization of the *marginal* distribution of the ED of a node from an anchor, when we are given the HD of the node from the anchor.

### 5.1 Results in the literature

There are several results in the literature that deal with the probability density functions relating the ED and the HD under various settings and assumptions. A brief overview of the different results is discussed in this section.

[Ta et al. 2007b], [Ta et al. 2007a], [Vural and Ekici 2005] address the problem of obtaining the probability density function (pdf) of the pairwise ED given HD. Ta et al. [Ta et al. 2007b], [Ta et al. 2007a], obtain a recursive expression for the pdf of the pairwise ED given the HD under certain independence assumptions. The pdf they derive becomes equal to the exact distribution for hop counts 1 and 2. For higher hop-counts, simulation results match well with the analytical results and they also discuss an empirical correction that further improves the distribution. However, the expressions that they derive are recursive in nature and hence would be difficult to compute as the hop count increases.

Vural *et al.* [Vural and Ekici 2005] obtain the distribution of the maximum distance traveled in a given number of hops. One dimensional sensor networks are considered and the approximate mean, variance etc are calculated. Gaussianity of the distribution is commented upon but the central limit theorem cannot be applied as successive hops are dependent in their construction. They also present a discussion on obtaining the kurtosis of the distribution relating it to the Gaussian distribution. Results relating the distribution of the maximum distance traveled in a given number of hops and the distribution of ED given HD are obtained. Using these results the density function of the HD given ED is derived. The recursive equation is difficult to compute for hop-counts more than 5. A direction propagation model is proposed for 2D networks, but no analysis is provided.

Dulman *et al.* [Dulman et al. 2006] also consider the problem of obtaining the distribution of the pairwise HD given ED for the cases of 1D and 2D networks. For the 1D case, exact recursive expressions are obtained for the mean of the distribu-

tion. For 2D case, an approximate algorithm is discussed and a recursive formula for the distribution is obtained. Since the distribution is difficult to compute, two approaches for computing the distribution are discussed and simulation results are presented, comparing the simulated and the approximated distribution. A localization algorithm using Least Squares Method, which makes use of the variances of the distribution is discussed.

[Miller 2001] and [Bettstetter and Eberspaecher 2003] consider the problems of obtaining the absolute probability density functions of the ED and the HD. Miller [Miller 2001] obtains the absolute pdf of the *link distance*. Link distance is the distance between any two random nodes whose locations are i.i.d in a rectangular field. Results are obtained for the uniform case, where the sensor location coordinates are uniform i.i.d and extended to the Gaussian case where the sensor location co-ordinates are Gaussian. Bettstetter et al. [Bettstetter and Eberspaecher 2003] consider the problem of obtaining the absolute pdf of the hop counts ( $P(H = h)$ ). The exact distribution is obtained for hop counts 1 and 2. An upper bound is obtained for the cumulative density function ( $P(H \leq h)$ ), which is evaluated asymptotically and compared with simulation results. Bounds on the expected HD are also obtained.

[De et al. 2006], [Zorzi and Rao 2003] obtain bounds for HD given ED and average HD given ED. These are for the forwarding algorithm that they use. De et al. [De et al. 2006] propose a greedy algorithm where one attempts to minimize the remaining distance to the destination in each hop. Lower and upper bounds on the HD as relative to the ED are obtained. Zorzi et al. [Zorzi and Rao 2003] propose a geographic packet forwarding algorithm (GeRaF) using hop-counts, and the average number of hops to reach the destination as a function of the distance is derived. A random topology where nodes could be following a sleep-wake cycle is considered. The node closest to the destination is chosen as the relay node in each hop. Bounds on the expected number of hops are derived using Wald's Lemma and the lower bound is shown to be tight through simulations. A practical protocol for the proposed GeRaF scheme is also discussed.

Ding et al [Ding et al. 2008] view the problem of localization as a standard ED matrix completion problem. Given the upper bounds, lower bounds and noisy versions of the different distances between the sensors and anchors, and amongst sensors, and the exact distances amongst the anchors nodes, one can formulate a weighted least squares problem with suitably defined weights. This problem has been studied in literature and the authors apply semi-definite programming to solve the problem.

Kuo et al [Kuo and Liao 2007] considers the problem of obtaining the HD distribution for a mobile node network. Note that since mobiles are roaming around there is only a notion of initial distance between the sender and the receiver. A flooding algorithm is considered and its analysis is provided.

In the present paper, we consider the problem of obtaining the distribution of ED given the HD. Our approach differs from prior approaches in the literature in the following respects. For the distribution of the ED given HD, the greedy algorithm that we propose is similar to the direction-propagation model discussed in [Vural and Ekici 2005]; the authors do not however, carry out the analysis for this model

and leave it as a future research topic. The region in which we choose our next hop sensor node is different from that proposed by Vural et al. By splitting the analysis along  $x$  and  $y$  co-ordinates we are able to carry forward the analysis. The greedy algorithm proposed in [De et al. 2006], aims to minimize the remaining distance at each stage, whereas we maximize the distance progressed in each step. [De et al. 2006] obtain bounds on the HD and not the distribution per se whereas we are interested in obtaining an approximation to the distribution. The assumptions made in our derivation enable us to use the central limit theorem to provide an explanation as to why the observed distribution is Gaussian-like. While, as a result of the assumptions, there is some deviation in the derived distribution from the true distribution, the derivation presented here does allow one to quickly obtain the distribution without need of recursive calculations. This could be useful in practice, when the distributions are used in practice for self-localization.

Following this, we propose two algorithms for localization based on our theoretical results, and discuss how *belief propagation* (BP) could be used to obtain improvements. [Ihler et al. 2005] discuss the use of BP for self-localization in a general setting. However to the best of our knowledge, we have not seen any published work that discusses the use of BP for hop-count-based localization algorithms.

## 5.2 The setting

Our interest here is in determining the pdf of the Euclidean distance  $D_{l,i}(\mathbf{v})$  of the  $i$ th node from the  $l$ th anchor node, given that the hop-count  $H_{l,i}(\mathbf{v}) = h_l$ . Given the intractable nature of this problem, we will determine instead, the pdf of  $D_{l,i}(\mathbf{v})$  given the hop-count as determined by a greedy algorithm presented below. Thus we regard the hop-count as determined by the greedy algorithm to be a good approximation to the true hop count and we provide below, a heuristic argument as well as simulations to back up this claim.

For the purposes of this analysis, we will assume the center of the unit area to have coordinates  $(0,0)$  and the presence of just a single anchor node located at the center, thereby staying away from edge effects in the analysis. Without loss of generality, we assume the node  $v_i$  to be located at coordinates  $(d, 0)$ . Since the distribution is independent of the value of  $i$  we will drop subscripts and write  $D(\mathbf{v})$  and  $H(\mathbf{v})$  in place of  $D_{l,i}(\mathbf{v})$ ,  $H_{l,i}(\mathbf{v})$  and  $h$  in place of  $h_l$ . However, we shall retain the subscript in  $v_i$ .

Let  $\mathcal{C}(u)$  denote a circle of radius  $r \triangleq r(n)$  centered at a point  $u$  in the unit area,  $\mathcal{C}_H(u)$  denote the half circle centered at  $u$  along the positive  $x$ -axis (i.e., the right half of a circle centered at  $u$ ) and let  $\mathcal{S}(u)$  denote the segment of the circle  $\mathcal{C}(u)$  cut out by a chord whose midpoint has coordinates given by  $u + (\frac{r}{2}, 0)$ . (see Fig. 14 below.). When a deployment  $\mathbf{v}$  is such that the region  $\mathcal{S}(u)$ , for  $u$  in the unit area, contains at least one node, we define the “furthest node” in the segment  $\mathcal{S}(u)$  as the node in  $\mathcal{S}(u)$  having largest  $x$ -coordinate and we use the notation  $\phi(\mathcal{S}(u), \mathbf{v})$  to denote the coordinates of this node<sup>4</sup>. Starting with the anchor node located at the origin, we next attempt to define a succession of nodes  $v_{i_j}$ ,  $j = 0, 1, \dots, h - 1$

<sup>4</sup>Ties can be broken for instance, by selecting the node whose  $y$ -coordinate has smallest magnitude.

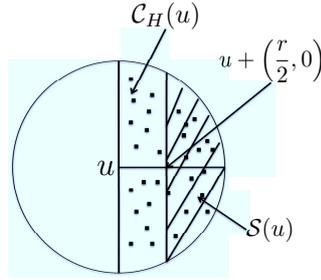


Fig. 14. Circle and associated segment.

as follows:

$$\begin{aligned} v_{i_0} &= (0, 0) \\ v_{i_j} &= \phi(\mathcal{S}(v_{i_{j-1}}), \mathbf{v}), \quad j \geq 1. \end{aligned}$$

In terms of this notation, let us define the event  $\mathcal{GE}_h$  as simply the event that the node  $v_i \in \mathcal{C}_H(v_{i_{h-1}})$ , i.e.,

$$(d, 0) \in \mathcal{C}_H(v_{i_{h-1}}),$$

(see Fig. 15). When this event occurs, we will say that the hop-count as determined by the greedy algorithm equals  $h$ . Note that this is well-defined since  $v_i \in \mathcal{C}_H(v_{i_{h-1}})$  is possible only for one value of  $h$ , by nature of the greedy algorithm.

The reasoning behind our belief that the change in conditioning from the event  $H(\mathbf{v}) = h$  to event  $\mathcal{GE}_h$  will leave the density function of  $D(\mathbf{v})$  essentially unchanged is simply that it seems reasonable to assume that one will hit upon a path from anchor node to node  $v_i$  with least hop count if when proceeding node-by-node from anchor node to node, at every stage, we select as the next node, the node that takes us farthest along in the direction of the node  $v_i$ . The reason for choosing segments in which the chord defining the segment is at distance  $\frac{r}{2}$  from the origin is to ensure that any two distinct segments  $\mathcal{S}(v_{i_k}), \mathcal{S}(v_{i_l})$  are disjoint. As  $n$  becomes large, with high probability, there will be a node in  $\mathcal{S}(v_{i_j})$  at each stage.

We now proceed to compute the density function of the distance  $D(\mathbf{v})$  conditioned upon the event  $\mathcal{GE}_h$ . For this purpose, we will first evaluate the probability that the hop count to reach the point  $(d, 0)$  using the greedy algorithm is  $h$ . We will then use the Bayes rule to obtain the density function of  $D(\mathbf{v})$ .

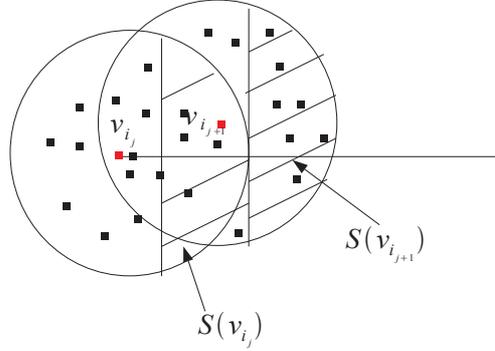


Fig. 15. Evolution of the greedy algorithm.

### 5.3 Determining the Density Function of $D(\mathbf{v})$

Let  $N$  nodes with locations given by  $\{v_1, v_2, \dots, v_N\}$  be deployed in a unit area where  $N$  is Poisson with mean  $n$ , i.e.,  $N \sim Poi(n)$ . Let  $(x_i, y_i)$  denote the  $x$  and  $y$  coordinates of the  $i$ th node  $v_i$ , i.e.,  $v_i = (x_i, y_i)$  (note that  $i$  is fixed here from our assumption in the beginning of this derivation). Given a particular deployment  $\mathbf{v}$ , let the nodes  $v_{i_j}$  be as defined above and let there be  $N_j$  nodes in segment  $\mathcal{S}(v_{i_j})$ . We shall neglect the probability that there are no nodes in a particular segment which is very small for the large density of nodes that we consider. Set

$$\begin{aligned} \mathcal{S}(v_{i_j}) &= \{(x_{i_j}, y_{i_j}) + (X_{j,k}, Y_{j,k}) \mid 1 \leq k \leq N_j\}, \\ \phi(\mathcal{S}(v_{i_j}), \mathbf{v}) &= (x_{i_{j+1}}, y_{i_{j+1}}) = (x_{i_j}, y_{i_j}) + (X_j, Y_j), \end{aligned}$$

where

$$X_j = \max\{X_{j,k} \mid 1 \leq k \leq N_j\}.$$

so that

$$\begin{aligned} x_{i_{j+1}} &= x_{i_j} + X_j, \\ y_{i_{j+1}} &= y_{i_j} + Y_j. \end{aligned}$$

The area of the segment  $\mathcal{S}(0)$  can be computed as the difference between the areas of the corresponding sector and triangle and is hence given by

$$\frac{1}{2}r^2 \frac{2\pi}{3} - \frac{\sqrt{3}r}{2} \times \frac{r}{2} = r^2 \left( \frac{\pi}{3} - \frac{\sqrt{3}}{4} \right) \triangleq A_s.$$

The probability that there is no node in a segment is given by  $(1 - A_s)^n$ .  $r$  needs to be at least  $c\sqrt{\frac{\ln(n)}{n}}$  for a graph to be connected where  $c$  is a constant. For large  $n$  this probability can be shown to be  $O(\frac{1}{n})$  which is small, thus justifying our assumption of at least one node in the segment.

It can be easily shown that, the density function of any of the random variables

$X_{j,k}$  is given by

$$f_{X_{j,k}}(x) = \begin{cases} \frac{2}{A_s} \sqrt{r^2 - x^2} & \frac{r}{2} \leq x \leq r \\ 0 & \text{else.} \end{cases}$$

The distribution of  $X_j$  is now obtained as follows. From our assumptions we have that all segments  $S(v_{i_j})$ ,  $1 \leq j \leq h-1$  are nonempty. We thus have

$$F_{X_{j,k}}(x) = \int_{\frac{r}{2}}^x \frac{2\sqrt{r^2 - t^2}}{A_s} dt.$$

Since

$$\int_{a_1}^{a_2} \sqrt{r^2 - t^2} dt = \frac{r^2}{2} \left[ \sin^{-1}\left(\frac{a_2}{r}\right) + \frac{a_2}{r} \sqrt{1 - \frac{a_2^2}{r^2}} - \sin^{-1}\left(\frac{a_1}{r}\right) - \frac{a_1}{r} \sqrt{1 - \frac{a_1^2}{r^2}} \right],$$

we get

$$F_{X_{j,k}}(x) = \frac{1}{\left(\frac{\pi}{3}\right) - \frac{\sqrt{3}}{4}} \left[ \sin^{-1}\left(\frac{x}{r}\right) + \frac{x}{r} \sqrt{1 - \frac{x^2}{r^2}} - \frac{\pi}{6} - \frac{\sqrt{3}}{4} \right].$$

Consider the joint pdf  $F_{X_j, N_j}(x, l) = P(X_j \leq x, N_j = l)$ . We have

$$F_{X_j, N_j}(x, l) = F_{X_j|N_j}(x|l)P(N_j = l).$$

We have

$$\begin{aligned} F_{X_j|N_j}(x|l) &= P(\max(X_{j,1}, \dots, X_{j,N_j}) \leq x | N_j = l), \\ &= P(\max(X_{j,1}, \dots, X_{j,l}) \leq x), \\ &= (F_{X_{j,k}}(x))^l. \end{aligned}$$

$$\begin{aligned} F_{X_j}(x) &= \mathbb{E}_{N_j} F_{X_j, N_j}(x, l), \\ &= \mathbb{E} [(F_{X_{j,k}}(x))^{N_j}] \\ &= \sum_{l=1}^{\infty} \frac{(nA_s)^l \exp(-nA_s)}{l!} (F_{X_{j,k}}(x))^l. \end{aligned}$$

Upon differentiating, we get

$$\begin{aligned} f_{X_j}(x) &= \sum_{l=1}^{\infty} \frac{(nA_s)^l \exp(-nA_s)}{(l-1)!} \left[ \frac{1}{\left(\frac{\pi}{3}\right) - \frac{\sqrt{3}}{4}} \right]^l \\ &\quad \left[ \sin^{-1}\left(\frac{x}{r}\right) + \frac{x}{r} \sqrt{1 - \frac{x^2}{r^2}} - \frac{\pi}{6} - \frac{\sqrt{3}}{4} \right]^{l-1} \frac{2\sqrt{r^2 - x^2}}{r^2}, \quad \frac{r}{2} \leq x \leq r \end{aligned}$$

and  $f_{X_j}(x) = 0$  else. By marginalizing, we find that

$$\begin{aligned} f_{Y_j}(y) &= \int f_{X_j, Y_j}(x, y) dx, \\ &= \int f_{X_j}(x) f_{Y_j|X_j}(y|x) dx, \\ &= \int_{-\sqrt{r^2-y^2}}^{\sqrt{r^2-y^2}} f_{X_j}(x) \frac{1}{2\sqrt{r^2-x^2}} dx. \end{aligned}$$

Since the deployment is Poisson and the nodes are chosen from disjoint areas, it follows that the random-variable pairs,  $(X_0, Y_0), \dots, (X_{h-2}, Y_{h-2})$  are i.i.d.

Let  $\mu_x, \mu_y$  and  $\sigma_x^2, \sigma_y^2$  be the means and variances of  $X_j$  and  $Y_j$  respectively. Then, from the central limit theorem we get that the sum  $x_{i_{h-1}} = X \sim N(h\mu_x, h\sigma_x^2)$  and  $y_{i_{h-1}} = Y \sim N(h\mu_y, h\sigma_y^2)$ . Note that the only dependence relation amongst the random variables in this collection of  $2(h-1)$  random variables is that between the two random variables constituting a pair  $(X_j, Y_j)$ . Let  $X_j = r_j \cos(\alpha_j)$  and  $Y_j = r_j \sin(\alpha_j)$ . From symmetry about the x-axis, we have

$$\mathbb{E}(Y_j) = 0.$$

Thus we obtain

$$\begin{aligned} \mathbb{E}(XY) &= \sum_i \sum_j \mathbb{E}(X_i Y_j) \\ &= \sum_j \mathbb{E}(X_j Y_j) \\ &= \sum_j \mathbb{E}\left(r_j^2 \frac{\sin(2\alpha_j)}{2}\right) \\ &= \sum_j \int_{r_j} r_j^2 f_{R_j}(r_j) \mathbb{E}\left(\frac{\sin(2\alpha_j)}{2} | R_j = r_j\right) dr_j \\ &= 0, \end{aligned}$$

since

$$\mathbb{E}\left(\frac{\sin(2\alpha_j)}{2} | R_j = r_j\right) = 0.$$

It follows from this and the fact that  $\mathbb{E}(Y) = 0$  that

$$\mathbb{E}([X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]) = 0.$$

Thus  $X$  and  $Y$  are uncorrelated. Since they are Gaussian, they are also independent. The probability that the node  $v_i$  located at location  $(d, 0) \in \mathcal{C}_H(v_{i_{h-1}})$  is reachable in  $h$  hops is obtained by integrating over all the points  $(x, y)$  such that  $(d, 0) \in \mathcal{C}_H((x, y))$ . Thus we have

$$P(H(\mathbf{v}) = h | d, \mathcal{G}\mathcal{E}_h) = \int_x \int_y f_{X,Y}(x, y) \mathcal{I}_{\{(d,0) \in \mathcal{C}_H((x,y))\}} dx dy$$

Using Bayes rule, the density function of the distance is obtained as follows

$$f_{D(\mathbf{v})}(d|\mathcal{G}\mathcal{E}_h) = \frac{P(H(\mathbf{v}) = h|d, \mathcal{G}\mathcal{E}_h)f_D(d)}{P(H(\mathbf{v}) = h|\mathcal{G}\mathcal{E}_h)}.$$

Assuming a circularly symmetric unit area, since the points are uniformly distributed we get

$$f_D(d) = 2\pi d$$

$$P(H(\mathbf{v}) = h|\mathcal{G}\mathcal{E}_h) = \int_d P(H(\mathbf{v}) = h|d, \mathcal{G}\mathcal{E}_h)f_D(d)dd.$$

The simulation results comparing the derived density function and the simulated density functions are shown in Figure 16. The simulated density function is obtained by averaging over a large number of Poisson deployments with mean node density of  $n = 8000$ . It can be seen that the density functions match pretty well at when the number of hops is large. However, when the number of hops is small, the density functions do not match quite that well. This is because the Gaussian approximation following from an application of the central limit theorem does not hold quite as well in the case of lower hop counts. This also means that we cannot then ignore the correlation between the variables  $X_j$  and  $Y_j$  random variables for small hop counts. However, it is theoretically hard to characterize the discrepancy in the distributions for lower hop counts, since the correlation is hard to characterize analytically. In a large network with sparsely spread out anchor nodes, one expects that the hop counts would be large enough for the approximation to be valid.

#### 5.4 Distribution of the location given the hop-count

In the previous section we obtained the distribution of the distance,  $D_{l,i}(\mathbf{v}) = d$ , of the  $i$ th node, with co-ordinates  $v_i = (x_i, y_i)$  given that the hop-count  $H_{l,i}(\mathbf{v}) = h$  from the  $l$ th anchor node with co-ordinates  $b_l = (u_l, v_l)$  as determined by the greedy algorithm, or else, equivalently, the distribution of the distance,  $D_{l,i}(\mathbf{v}) = d$  conditioned on the event  $\mathcal{G}\mathcal{E}_h$ . It is clear that the knowledge of the hop-count does not provide any information on the direction of the  $i$ th node. Thus based on the fact that the node location is uniform i.i.d, we have the following:

$$x_i = u_l + D_{l,i}(\mathbf{v}) \cos(\alpha)$$

$$y_i = v_l + D_{l,i}(\mathbf{v}) \sin(\alpha).$$

where  $\alpha \sim U(0, \phi)$  and  $\phi$  depends on the location of the anchor node. If the anchor node is at the center of the field then  $\phi = 2\pi$ . If the anchor node is at the corner of a square field then,  $\phi = \frac{\pi}{2}$ . We have:

$$f_{D_{l,i}(\mathbf{v}),\alpha}(d, \alpha | \mathcal{G}\mathcal{E}_h) = f_{D_{l,i}(\mathbf{v})}(d | \mathcal{G}\mathcal{E}_h)f_\alpha(\alpha).$$

Using transformation of variables we get the following:

$$f_{V_i|H_{l,i}(\mathbf{v})}(x_i, y_i | \mathcal{G}\mathcal{E}_h) = f_{D_{l,i}(\mathbf{v}),\alpha} \left( \|v_i - b_l\|, \tan^{-1} \left( \frac{y_i - v_l}{x_i - u_l} \right) | \mathcal{G}\mathcal{E}_h \right) \frac{1}{|J|}.$$

## 6. ALGORITHMS FOR SELF-LOCALIZATION

Based on the theory discussed in the previous sections, we propose two algorithms for self-localization, namely, Hop Count derived ED Bounds based Localization

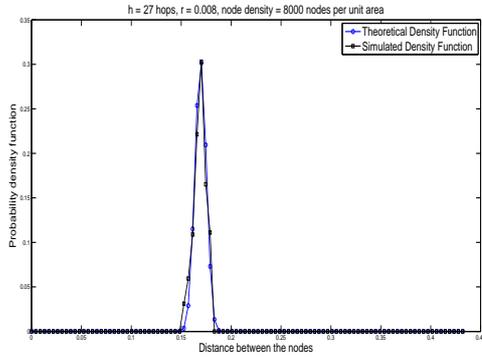
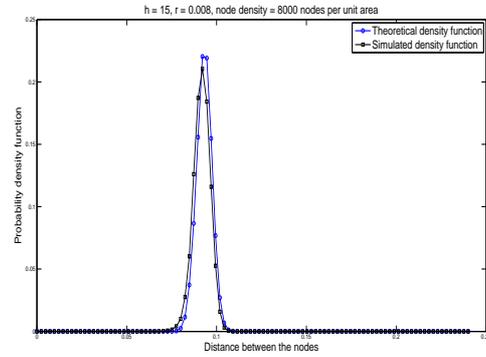
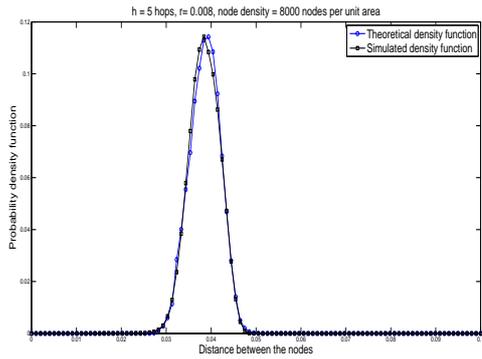
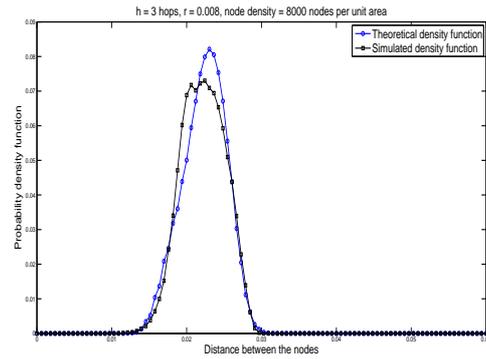
(a)  $h=27, r=0.008$ (b)  $h=15, r=0.008$ (c)  $h=5, r=0.008$ (d)  $h=3, r=0.008$ 

Fig. 16. Simulation results comparing the derived and simulated density functions. The simulations are for Poisson deployment with mean node density of 8000 nodes per unit area.

(HCBL) and Hop Count derived ED Distribution based Localization (HCDL). Simulation results show that these algorithms outperform some of the well known algorithms present in literature. We then use Belief Propagation (BP) to improve the performance of these algorithms. Based on this we get two algorithms HCBL-BP and HCDL-BP.

We emphasize that the algorithms mentioned here take as input the geometric graph with radius  $r(n)$ . Getting such a graph from the radio graph, i.e., the graph formed by neighbor relation induced by inter-node radio communication, is a different question altogether and is not addressed here. In [Acharya et al. 2010], a technique is provided for obtaining an approximation to the *critical geometric graph*, whose radius  $r(n)$  is the smallest such that the resulting graph is connected. In [Narayanaswamy et al. 2002], a distributed protocol named COMPOW has been

introduced, which constructs a critical geometric graph with nodes operating at a common power level. On obtaining such a graph, we find hop distances between the nodes and the anchors using the distributed Bellman-Ford algorithm; see [Kurose and Ross 2010]. Once the algorithm converges, we get the hop-distances of the nodes from the anchors and also the inter-anchor hop-distances.

Furthermore, for the following algorithms, we need an estimate of the  $r(n)$  at the center (note that the nodes do not estimate this, since we have moved all the computation to a central hub which knows the hop distances of all nodes from the anchors). An estimate can be made by obtaining the hop distances between the anchors, for which the locations are known and hence the Euclidean distances between them. The known Euclidean distances along with hop distances between the anchors, provides an estimate of the radius of the geometric graph.

### 6.1 Hop Count derived ED Bounds based Localization (HCBL)

The setting and notation are the same as in Section 2. Theorem 1 and Theorem 2, explore the relationship between the HD and the ED. From Theorem 2, a node having HDs  $h_l$  from the anchors  $b_l$ ,  $l = 1, \dots, L$ , lies with high probability in the region of intersection of the bands  $[(1 - \epsilon)(h_l - 1)r(n), h_l r(n)]$ ; see Figure 7. This is the basic idea behind HCBL. However, the simulation results in Figures 11 to 12 suggest a smaller  $\epsilon$  than that suggested by theory. The following is the algorithm that we propose.

#### **The HCBL Algorithm:**

*STEP 1: (Initialization)* Given the geometric graph  $\mathcal{G}(\mathbf{v}, r(n))$ , each node finds the minimum hop-distances from the  $L$  anchors and sets up its own  $\mathbf{h} = [h_1, \dots, h_L]$  vector, where  $h_l$  is the hop distance of the node from the  $l^{\text{th}}$  anchor  $b_l$  (This can be carried out by the Bellman-Ford algorithm).

*STEP 2: (Region of Intersection)* For a given node, given its  $\mathbf{h}$  vector as found from *STEP 1*, set an  $\epsilon$ , small enough, and find the region of intersection formed by the annuli of radii  $\{(1 - \epsilon)(h_l - 1)r(n), h_l r(n)\}$  centered at the  $l^{\text{th}}$  anchor location,  $l = 1, \dots, L$ .

*STEP 3: (Terminating Condition)* Check if the region of intersection is non-empty, otherwise increase the value of  $\epsilon^5$ . For a finite number of nodes  $n$  and a small enough  $\epsilon$ , it is possible that the annuli do not have a common region. A graphical illustration is given in Fig. 7 for 4 anchors. The value of  $\epsilon$  for which an intersection is found, can be different for different nodes. Hence, this step can be stated as follows.

*IF* there is an intersection, declare the centroid of the region of intersection as the estimate of the node. *GO TO STEP 4.*

*ELSE* increase  $\epsilon$  by an amount  $\delta$ ,  $0 < \delta < 1$ . *GO TO STEP 2.*

*STEP 4: (Repetition)* Repeat *STEP 2* to *STEP 3* for all  $n$  nodes.

*STEP 5: STOP*

<sup>5</sup>We could use the  $\epsilon$  as proposed by our theory. However, since the bounds are not tight, the performance of the algorithm would be poor.

**Time complexity of HCBL:** Assuming that given a region of intersection, the computation of the centroid takes constant time, the time complexity to find the hop distances of  $n$  nodes from the  $L$  anchors is  $\mathcal{O}(n^2)$ . Hence, *STEP 1* completes in  $\mathcal{O}(n^2)$  time. After getting the hop distances, we are increasing the  $\epsilon$  in each iteration and if there is an intersection in the annuli, we compute the centroid, which is an  $\mathcal{O}(1/\delta)$  ( $\delta$  is the step size of the increment of  $\epsilon$ ) computation. So, for each node the time complexity of the *STEPS 2 to 4* is  $\mathcal{O}(1/\delta)$ . For  $n$  nodes, it is  $\mathcal{O}(n/\delta)$ . Hence, the time complexity of HCBL algorithm is  $\mathcal{O}(n^2) + \mathcal{O}(n/\delta)$ . There is a tradeoff for the choice of  $\delta$  since a small  $\delta$  will increase the accuracy of the localization but the time complexity of HCBL will be higher whereas a larger value will converge HCBL faster at the cost of reduced accuracy.

The performance of this algorithm is shown in Figure 21. The x-axis of the plot is the localization error. The y-axis is the cumulative density function of the error. The performance is compared with two other algorithms, Proximity Distance Map (PDM) by [Lim and Hou 2005] and Hop Count Ratio based Localization (HCRL) by [Yang et al. 2007]. Distance Vector hop (DV-hop) by [Niculescu and Nath 2003], is shown to be better than HCRL and hence we do not plot DV-hop in our simulations. It can be seen that HCBL performs better than the other algorithms.

## 6.2 Hop Count derived ED Distribution based Localization (HCDL)

The setting and notation are the same as in Section 2. In the previous section, we obtained an approximation to the density function of the node location given the hop-count. For the case when there are  $L$  anchor nodes, every node  $i$ , will have a hop-tuple  $\{H_{l,i}(\mathbf{v})\}_{l=1}^L$ . If one could obtain access to the density function of the node location, one could then proceed to derive the Minimum Mean Square Error (MMSE) estimate of the node location and this is the basic principle behind the Hop-Count-derived, ED-Distribution-based Localization (HCDL) technique. There are two issues however. The first is that we do not have access to the true density function of node location, but only to an approximate version of it. We handle this by simply working with the approximate density function as if it were the true density function. The second issue is that if there are  $L$  anchor nodes, and each node is aware of its hop count with respect to each of the  $L$  anchor nodes, then it would have  $L$  estimates of the density function of its node location, leading to the question of how one would proceed to fuse these  $L$  pieces of information. We answer this below.

Let  $f_{V_i|\{H_{l,i}(\mathbf{v})\}_{l=1}^L}(v_i | \{h_{l,i}\}_{l=1}^L)$  denote the distribution of the  $i$ th node location given its hop-count tuple. We then have the following:

$$\begin{aligned} f_{V_i|\{H_{l,i}(\mathbf{v})\}_{l=1}^L}(v_i | \{h_{l,i}\}_{l=1}^L) &= \frac{f(\{h_{l,i}\}_{l=1}^L | v_i)f(v_i)}{f(\{h_{l,i}\}_{l=1}^L)} \quad \text{using Bayes rule} \\ &\propto f(\{h_{l,i}\}_{l=1}^L | v_i) \\ &\approx \prod_{l=1}^L f(h_{l,i} | v_i) \\ &\propto \prod_{l=1}^L f_{V_i|H_{l,i}(\mathbf{v})}(v_i | h_{l,i}). \end{aligned}$$

The third approximation can be justified as follows. Suppose that the anchor nodes are spaced sufficiently far apart (i.e., have large angular separation), given

the location  $v_i$  and the location of the anchor nodes,  $H_{l,i}(\mathbf{v})$  depends only on the distribution of the nodes linking the anchor node<sup>6</sup>  $l$  and location  $v_i$ . Suppose that the subsets of nodes linking the node  $v_i$  to the  $L$  anchor nodes are pairwise disjoint, then since the node locations are assumed to be i.i.d the above approximation is justified. Note that the pair-wise disjoint assumption holds well when the anchor nodes are well separated in terms of angular separation. For example the anchors could be aligned along the periphery of the field maximally spaced apart. They could also be placed inside the field, in a deterministic hexagonal pattern, etc. However we do not quantify as to how far apart the anchors need to be placed for this assumption to hold and is more of an approximate derivation. In practice for a military setting, anchors are usually base stations that are placed at the corners of the area that is accessible and sensors are thrown in the inaccessible areas (as in our simulations) for which the algorithm seems to work well in simulations.

Given the hop count tuple  $\{H_{l,i}(\mathbf{v})\}_{l=1}^L$ , we know the distribution of  $f_{V_i|\{H_{l,i}(\mathbf{v})\}_{l=1}^L}(v_i | \{h_{l,i}\}_{l=1}^L)$  and thus we know the marginals  $f(x_i | \{h_{l,i}\}_{l=1}^L)$  and  $f(y_i | \{h_{l,i}\}_{l=1}^L)$ . Hence the MMSE estimate of the location is given by

$$\begin{aligned}\hat{x}_{i,MMSE} &= \mathbb{E}(X_i|\{H_{l,i}(\mathbf{v})\}_{l=1}^L). \\ \hat{y}_{i,MMSE} &= \mathbb{E}(Y_i|\{H_{l,i}(\mathbf{v})\}_{l=1}^L).\end{aligned}$$

Thus the HCDL proceeds as follows. Each node initially gets the hop-count tuple from the anchor nodes. Based on the probability distribution, the node calculates its MMSE estimate. Simulation results show that this algorithm has the best performance amongst all the better-known algorithms to be found in the literature. The advantage of this approach is that if the distribution is a good approximation of the true distribution, then this algorithm gives the best possible mean-square performance with respect to any hop-count based approaches.

### 6.3 Simulation results

The simulation results for all the algorithms are shown in Fig. 17. The x-axis is the localization error and the y-axis is the cumulative density function (cdf) of the errors. The simulation setting is as follows. A set of nodes are uniformly deployed over a unit area. The number of nodes is Poisson distributed with mean  $n = 1000$ . The localization algorithms are run over the critical graph. The location estimates of the nodes are obtained and the location errors are stored for the different algorithms. The location errors are the standard mean square errors in localization calculated as  $\|v_i - \hat{v}_i\|_2$  where  $v_i$  is the location of the  $i$ th node and  $\hat{v}_i$  is its estimate and  $\|\cdot\|_2$  is the standard L2 norm. The experiment is repeated many times and the error cdf is obtained by averaging over the error cdf's of the individual deployments. Note that the localization errors are normalized, since the node deployment is over a unit area. For these simulation results, the experiment has been carried out four times. It can be seen that the error performance of both the algorithms is better than PDM and HCRL. To the best of our knowledge we are not aware of any theoretical work that characterizes the best possible achievable error cdf or lower bounds on the error in localization for localization for hop-count

<sup>6</sup>Note that the anchor node locations are assumed to be known.

based approaches. Usually the Cramer Rao Lower bound is used as a benchmark to compare the performance of any algorithm. However this requires the knowledge of the exact probability distribution of the hop-distances as a function of the true distances which as we saw is hard to obtain. Hence we restrict ourselves to comparing our results with existing algorithms in the literature.

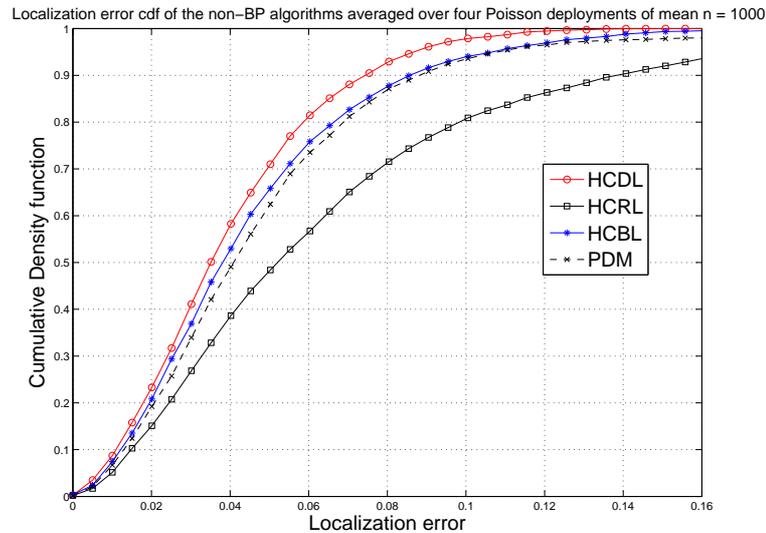


Fig. 17. Location error cdf for different approaches.

## 7. BELIEF PROPAGATION BASED ALGORITHMS

In a scenario in which every node  $v_i$  knows its hop counts  $\{H_{l,i}(\mathbf{v})\}_{l=1}^L$ , we have shown how the node can use this set of hop counts to estimate its location. This however raises the question: could a node use the hop counts of its neighbors to improve the estimate of its own location? We answer this question in the affirmative in this section by showing how one can use belief propagation (BP) to improve estimates of node location. We will abbreviate and write  $H_{l,i}$  in place of  $H_{l,i}(\mathbf{v})$ .

BP algorithms, also often referred to as message-passing algorithms, have been extensively used in the literature to obtain significant performance improvements, as can be seen from the impact of turbo and low-density parity-check codes upon the literature in coding theory [Aji and McEliece 2002], [Kschischang et al. 2001]. In a scenario where the available information is distributed, BP provides a framework that permits efficient data fusion. Pearl's [Pearl 1988] book is a classic reference which deals with BP in detail. There has also been some interesting work on using BP for self-localization in sensor networks [Ihler et al. 2005]. Here, the authors discuss localization for range based localization techniques but one does not have a parametric model for the observations. There has been some work in trying to

understand BP in loopy networks [Crick and Pfeffer 2003], [Yedidia et al. 2003], [Yedidia et al. 2001]. In the present work, we structure the computation in such a way that the resultant graph is acyclic.

### 7.1 BP for hop-count based localization

Our aim in this subsection, is to use belief propagation to improve localization accuracy of a node in the network by incorporating knowledge of certain select hop counts in the network. We make two assumptions with heuristic justification, that make the problem tractable and which lead to an acyclic graphical model to which belief propagation can then be applied. Simulations, presented in the subsection following, show that this approach can significantly increase localization accuracy.

Let  $v_a$  be a node<sup>7</sup> whose hop-distance set to the  $L$  anchors, is given by  $\{H_{l,a}\}_{l=1}^L$ . This means that with respect to a fixed anchor node  $b_l$ , there are  $(H_{l,a} - 1)$  nodes on the minimum hop-count path  $b_l$  to node  $a$ . Let  $\{v_{l_i}\}_{i=1}^{\mu_l}$  be a subset of these nodes selected in such a way that the sequence of nodes from  $b_l$  to node  $v_a$  reads as

$$(v_{l_1}, v_{l_2}, \dots, v_{l_{\mu_l}}),$$

in that order; thus  $v_{l_1}$  is closest to anchor node  $b_l$ . Within this chain of nodes, it is clear that, for  $j > i$ , node  $v_{l_j}$  knows the minimum hop count to it from node  $v_{l_i}$  and we will extend the hop-count notation and write  $H_{l_i,l_j}$  to denote this hop count. Also, we will use  $H_{l_{\mu_l},a}$  to denote the hop count from the last node in this chain to node  $v_a$ . The selection of the nodes in the chain will be carried out keeping in mind the need to select those hop counts that result in a stronger link between hop count and Euclidean distance. Fig. 18 illustrates the setting for the case of 4 anchor nodes, i.e.,  $L = 4$ . We collect together the set of relevant hop counts into three disjoint sets:

—  $\mathcal{H}_{\text{anchor}}$ , which denotes the set of minimum hop counts of the nodes in Fig. 18 to the anchors, i.e.,

$$\mathcal{H}_{\text{anchor}} = \{H_{l,l_j} \mid 1 \leq l \leq L, 1 \leq j \leq \mu_l\} \cup \{H_{l,a} \mid 1 \leq l \leq L\},$$

—  $\mathcal{H}_{\text{inter}}$ , which denotes the set of minimum inter-node hop counts between adjacent nodes in Fig. 18, in the direction anchor node to node  $v_a$ , i.e.,

$$\mathcal{H}_{\text{inter}} = \{H_{l_i,l_j} \mid 1 \leq l \leq L, 1 \leq i < j \leq \mu_l\},$$

—  $\mathcal{H}_{\text{target}}$ , which denotes the set of minimum hop counts between nodes  $v_{l_{\mu_l}}$  and node  $v_a$  in Fig. 18, i.e.,

$$\mathcal{H}_{\text{target}} = \{H_{l_{\mu_l},a} \mid 1 \leq l \leq L\}.$$

We set

$$\mathcal{H} = \mathcal{H}_{\text{anchor}} \cup \mathcal{H}_{\text{inter}} \cup \mathcal{H}_{\text{target}}.$$

In the belief-propagation-based computation to follow, we will make the following assumptions:

<sup>7</sup>We will often refer to a node by its location.

- Dependency-only-through-minimum-Hop-Count (HCD) assumption*: Under this assumption, given  $\mathcal{H}$ , the location of two nodes are dependent on each other only if the minimum hop count between the two nodes is known. The justification here being that given knowledge of hop counts to the anchor nodes, knowledge of node  $v_i$ 's location strongly affects a second node  $v_j$ 's location only if the minimum hop count between  $v_i$  and  $v_j$  is known.
- Closest-Neighbor Dependency (CND)*: Here we address the situation when hop counts of a node  $v_{l_i}$  to both anchor node  $b_l$  and node  $v_{l_j}$  are known and node  $v_{l_i}$  lies on a minimum-hop-count path leading from node  $v_{l_j}$ ,  $j > i$ , to the anchor node  $b_l$ . In this situation, we assume that  $v_{l_j}$  is independent of  $b_l$  given knowledge of  $v_{l_i}$ . A similar situation arises when hop counts of a node  $v_{l_k}$  to both node  $v_{l_i}$  and node  $v_{l_j}$ ,  $i < j < k$ , are known and node  $v_{l_j}$  lies on a minimum-hop-count path leading from node  $v_{l_i}$  to the node  $v_{l_k}$ . Again, in this situation, we assume that  $v_{l_k}$  is independent of  $v_{l_i}$  given knowledge of  $v_{l_j}$ . The justification in both instances being that knowledge of the hop count to a geographically closer node is likely to render obsolete, information conveyed by the hop distance to the more distant node.

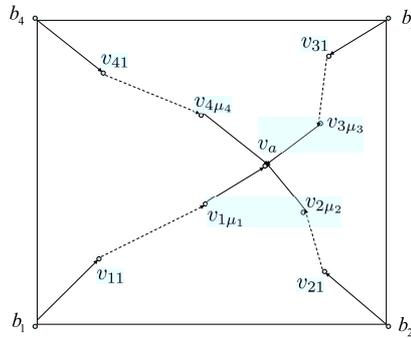


Fig. 18. Chain of nodes along which beliefs are passed.

Our interest is in determining the density function

$$f(v_a | \mathcal{H}),$$

under the assumption that the location  $\{b_l\}_{l=1}^L$  of the anchor nodes is fixed and known. We shall now illustrate the computation under these two assumptions, of  $f(v_a | \mathcal{H})$  for a sample network as shown in Figure. 19. We will compute  $f(v_a | \mathcal{H})$  by first computing  $f(v_a, v_{11}, v_{12}, v_{21}, v_{22} | \mathcal{H})$  and then marginalizing to obtain the desired density function by integrating over  $v_{11}, v_{12}, v_{21}, v_{22}$ . For the example

network of Figure. 19, we have

$$\begin{aligned}\mathcal{H}_{\text{anchor}} &= \{H_{l,11}, H_{l,12}, H_{l,21}, H_{l,22}, H_{l,a} \mid l = 1, 2\} \\ \mathcal{H}_{\text{inter}} &= \{H_{11,12}, H_{21,22}\} \\ \mathcal{H}_{\text{target}} &= \{H_{12,a}, H_{22,a}\}.\end{aligned}$$

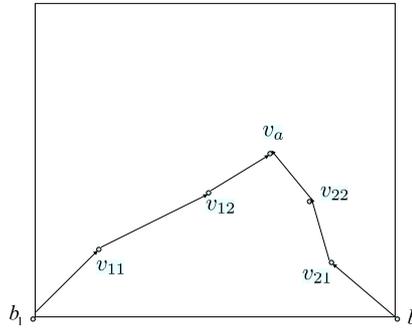


Fig. 19. Sample belief propagation network with two anchor nodes.

By making use of our HCD and CND assumptions above, the problem can be presented as one of marginalizing a product function [Aji and McEliece 2002]:

$$f(v_a|\mathcal{H}) = \int_{v_{11}, v_{12}, v_{21}, v_{22}} f(v_{11}|\mathcal{H})f(v_{12}|v_{11}, \mathcal{H})f(v_{21}|\mathcal{H})f(v_{22}|v_{21}, \mathcal{H})f(v_a|v_{12}, v_{22}, \mathcal{H})dv_{11}dv_{12}dv_{21}dv_{22}.$$

The distributive law can be applied to yield an efficient, message-passing solution, which in the present instance means reorganizing the sequence of marginalizations in the form:

$$\begin{aligned}f(v_a|\mathcal{H}) &= \int_{v_{12}, v_{22}} f(v_a|v_{12}, v_{22}, \mathcal{H})dv_{12}dv_{22} \\ &\quad \int_{v_{11}} f(v_{11}|\mathcal{H})f(v_{12}|v_{11}, \mathcal{H})dv_{11} \\ &\quad \int_{v_{21}} f(v_{21}|\mathcal{H})f(v_{22}|v_{21}, \mathcal{H})dv_{21}.\end{aligned}$$

The sequence of marginalizations is most easily derived from the corresponding junction tree, shown in Fig. 20, in which the dotted lines indicate the stages in which marginalization of the different variables takes place. Such a junction tree can always be constructed as it is derived from the directed acyclic graph (DAG) representing factorization of the joint probability density function according to:

$$f(v_a, v_{11}, v_{12}, v_{21}, v_{22} | \mathcal{H}) = f(v_{11} | \mathcal{H})f(v_{12} | v_{11}, \mathcal{H})f(v_{21} | \mathcal{H})f(v_{22} | v_{21}, \mathcal{H}), f(v_a | v_{12}, v_{22}, \mathcal{H}),$$

see [Aji and McEliece 2002] for details. While we have presented this by example, the technique clearly carries over to the general situation.

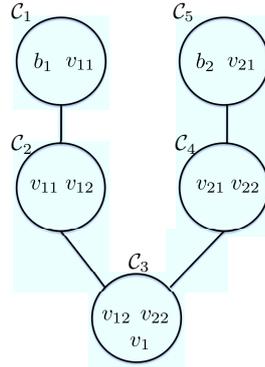


Fig. 20. Junction Tree associated to the marginalize-a-product function problem.

Let us now specialize the density functions for the cases of HCDL-BP and HCBL-BP.

*Messages for HCDL-BP:* For the case of HCDL-BP, whenever we need to evaluate a density function of the form  $f(v_a | v_{l_{\mu_l}}, H_{l_{\mu_l}, a})$ , i.e., the density function of a node  $v_a$  given the location of a second node  $v_{l_{\mu_l}}$  and the minimum hop count between the two nodes, we will make use of the the density function expressions derived in Section 5. We will also have to compute density functions of the form

$$f(v_a | \{v_{l_{\mu_l}}\}_{l=1}^L, \{H_{l_{\mu_l}, a}\}_{l=1}^L).$$

Such densities can be evaluated (we make use of our CND assumption above here) according to:

$$\begin{aligned} f(v_a | \{v_{l_{\mu_l}}\}_{l=1}^L, \{H_{l_{\mu_l}, a}\}_{l=1}^L) &\propto f(v_a, \{v_{l_{\mu_l}}\}_{l=1}^L | \{H_{l_{\mu_l}, a}\}_{l=1}^L) \\ &= f(v_a | \{H_{l_{\mu_l}, a}\}_{l=1}^L) \prod_{l=1}^L f(v_{l_{\mu_l}} | v_a, H_{l_{\mu_l}, a}) \\ &\propto \prod_{l=1}^L f(v_a | v_{l_{\mu_l}}, H_{l_{\mu_l}, a}) \end{aligned}$$

since

$$f(v_a | \{H_{l_{\mu_l}, a}\}_{l=1}^L) = f(v_a)$$

is assumed to be uniform and hence not a function of  $v_a$ . In the above derivation, the  $\propto$  sign at each step indicates that the constant of proportionality is independent of  $v_a$ .

*Messages for HCBL-BP:* For the case of HCBL-BP, the density functions are defined as follows. The density function  $f(v_a | v_{l_{\mu_l}}, H_{l_{\mu_l}, a})$ , of a node  $v_a$  given the location of a second node  $v_{l_{\mu_l}}$  and the minimum hop count between the two nodes, would be the uniform distribution over the support

$$\{v_a | \|v_a - v_{l_{\mu_l}}\| \in [(1 - \epsilon)(H_{l_{\mu_l}, a} - 1), H_{l_{\mu_l}, a}]\}.$$

To evaluate a density function of the form

$$f(v_a | \{v_{l_{\mu_l}}\}_{l=1}^L, \{H_{l_{\mu_l}, a}\}_{l=1}^L),$$

we will use a uniform distribution over the support

$$\{v_a \in \bigcap_l \{v_a | \|v_a - v_{l_{\mu_l}}\| \in [(1 - \epsilon)(H_{l_{\mu_l}, a} - 1), H_{l_{\mu_l}, a}]\}\}.$$

In other words, we take the node location  $v_a$  to be uniform over the intersection of the bands dictated by the different hop-counts from given locations.

**Computational cost of BP:** It is easy to observe that the message passed by a node is independent of the node to be localized. The message passed by a node to its neighbor is the posteriori probability of the location of the neighbor. Thus the communication cost of a node is  $kO(\log n)$  where  $k$  is a constant which depends on the resolution of the region over which the posterior probability is given. For example if the entire region is split into  $p$  squares and the node has to be localized to a square, then  $k = p$ .

## 7.2 Simulation results

The simulation results for all the algorithms are shown in Fig. 21. The x-axis is the localization error and the y-axis is the cumulative density function of the errors. The sample error-vector plots or ‘Mikado’<sup>8</sup> diagrams over a particular deployment for the different approaches are shown in Fig. 22. The simulation set-up is as follows. The number of nodes deployed in an unit square is taken to be Poisson with mean  $n = 1000$ . The algorithms are run over the critical graph. The performance is obtained by averaging the errors over different node locations and multiple deployments.  $H_{BP} = 5$  in these simulations. The average error cdf is obtained over four node deployments. We can see that BP gives significant improvements over the previous algorithms. Since the experiment is repeated for a lesser number of trials, the performance of PDM and HCBL is not clearly distinguishable. But this is clearer in Figure 17.

<sup>8</sup>Mikado is the ancient name for the game commonly called pick-up sticks. Here sticks of different colors are dropped and the players need to pick up each stick without disturbing the others. The error patterns generated here look like the fallen sticks in a game of Mikado, hence our use of this name.

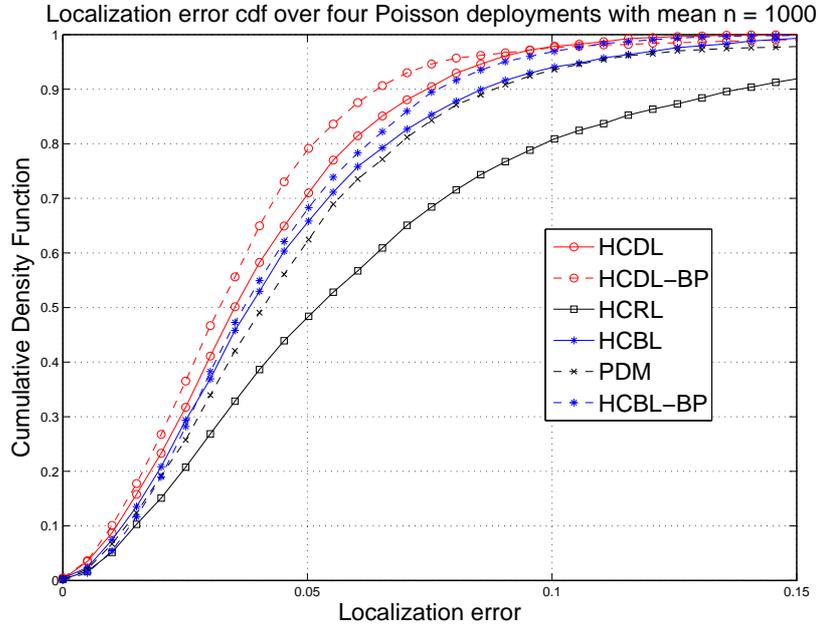


Fig. 21. cdf for location error for each of the localization algorithms HCRL, PDM, HCDL, HCBL, HCDL-BP and HCBL-BP.

There are a few points noted here. It is possible that in individual cases of node locations BP might worsen the location error. However on average BP provides improvements. Also, here we have just a single iteration of BP. In practice, one could have multiple iterations of BP with the posterior probabilities of the node locations obtained in one iteration can be used as prior probabilities in the next iteration.

## 8. CONCLUSIONS

In this paper, we have formally studied the often used heuristic that HD on a geometric graph on a plane is proportional to the ED. For arbitrary 2-D node placements, we saw in Lemma 1, that for HD  $h \geq 2$ ,  $r < ED \leq hr$  and this is not even roughly proportional to HD. For homogeneous random deployments, we have found that for given HD  $h$ ,  $(1 - \epsilon)(h - 1)r < ED \leq hr$  w.h.p. (this is true even for a non-homogeneous node placement with a positive density over all points of the area). Our proof techniques rely on a certain geometric construction and the union bound. The parameter  $\epsilon$  provides a trade-off between ED-HD proportionality and the rate of convergence of the desired probability to 1. This result holds for both uniform i.i.d. (Theorem 1) and randomized lattice (Theorem 4) node placements. We also provide Theorems 2, 3 and 5, that can be useful for GPS-free localization. Simulation results show that the actual probability of ED-bounds being respected is much closer to 1 than that provided by our bounding arguments. Given an HD,

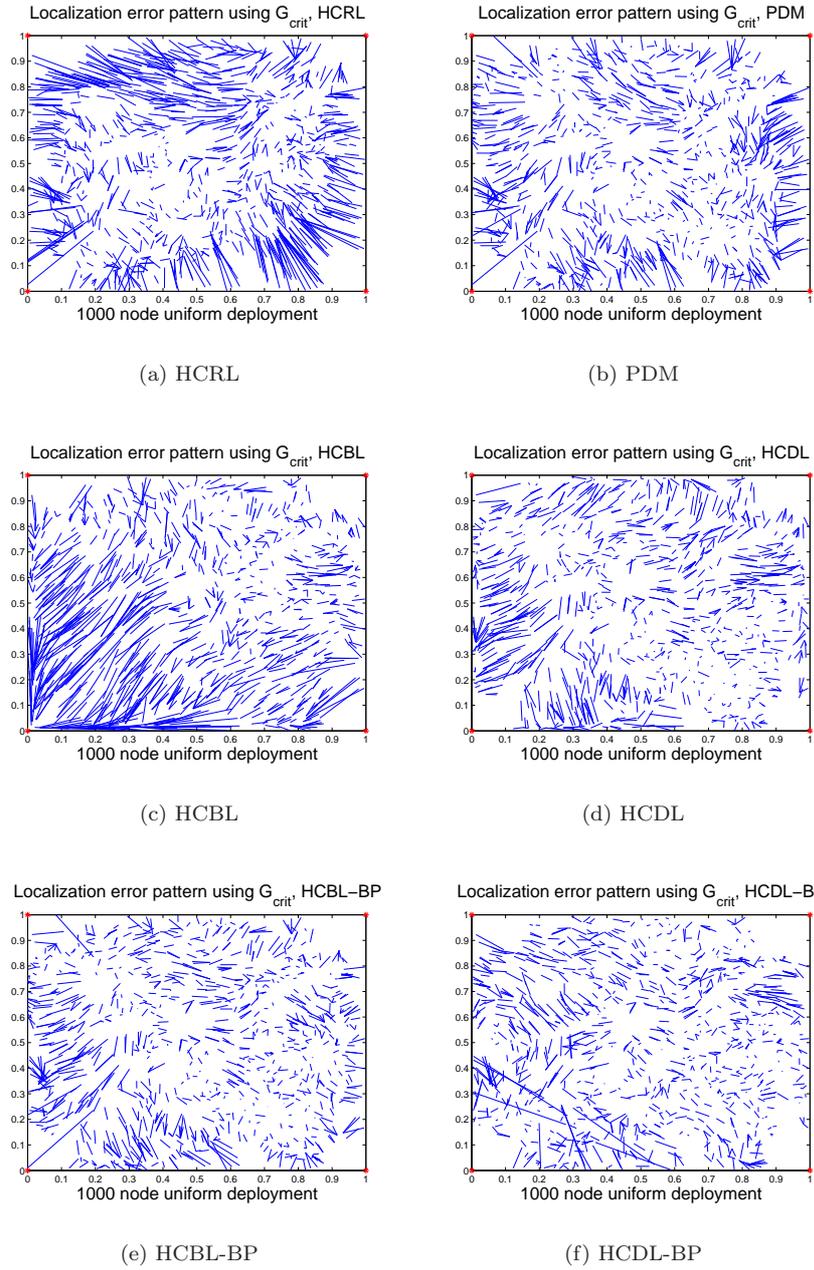


Fig. 22. Error vector plots (‘Mikado diagrams’) for the different approaches for a sample node deployment.

$h$ , the simulations suggest that  $[(h-1)r, hr]$  could serve as a good approximation to the interval in which the EDs of the nodes lie; however, this might not be a high probability bound.

In the second part of the paper we discuss the problem of obtaining the pdf of the pairwise Euclidean distance given the hop distance. A survey of the existing results in the literature is provided. Using a greedy algorithm, we heuristically obtain the distance distribution given the hop-count and show a close match with the simulation results. Our assumptions help us to invoke the central limit theorem, and thus we are easily able to obtain the approximate probability density functions for different hop-counts. The simulation results match well for larger hop count values, which is expected due to the nature of the central limit theorem.

We propose two algorithms HCBL and HCDL whose motivation arise from the developed theory and show that the performance is better than some of the proposed algorithms in literature. We then discuss the utility of belief propagation in obtaining performance improvements over the proposed algorithms.

In our future work, we seek more precise bounds on the ED with a provably higher probability of being respected. We would also like to develop theories and algorithms for anisotropic networks.

#### ACKNOWLEDGMENTS

#### REFERENCES

- ACHARYA, S., KUMAR, A., DEWANGAN, V., SANKARA, N., HEGDE, M., AND ANAND, S. V. R. 2010. Distributed Construction of the Critical Geometric Graph in Dense Wireless Sensor Networks. *ArXiv e-prints* 2010arXiv1011.3482A (Nov.).
- AJI, S. AND McELIECE, R. 2002. The generalized distributive law. *Information Theory, IEEE Transactions on* 46, 2, 325–343.
- BETTSTETTER, C. AND EBERSPAECHER, J. 2003. Hop distances in homogeneous ad hoc network. In *IEEE Vehicular Technology Conference 2003*.
- COSTA, J., PATWARI, N., AND HERO III, A. 2006. Distributed weighted-multidimensional scaling for node localization in sensor networks. *ACM Transactions on Sensor Networks (TOSN)* 2, 1, 39–64.
- CRICK, C. AND PFEFFER, A. 2003. Loopy belief propagation as a basis for communication in sensor networks. In *Uncertainty in Artificial Intelligence*. Vol. 18.
- DE, S., CARUSO, A., CHAIRA, T., AND CHESSA, S. 2006. Bounds on Hop Distance in Greedy Routing Approach in Wireless Ad Hoc Networks. *International Journal of Wireless and Mobile Computing* 1, 2, 131–140.
- DING, Y., KRISLOCK, N., QIAN, J., AND WOLKOWICZ, H. 2008. Sensor network localization, euclidian distance matrix completions, and graph realization. In *Workshop on mobile entity localization and tracking in GPS-less environments 2008*.
- DULMAN, S., ROSSI, M., HAVINGA, P., AND ZORZI, M. 2006. On the hop count statistics for randomly deployed wireless sensor networks. *Int. J. Sensor Networks* 1, 1/2.
- GUPTA, P. AND KUMAR, P. R. 1998. Critical Power for Asymptotic Connectivity in Wireless Networks. *Stochastic Analysis, Control, Optimization and Applications*.
- GUVENÇ, I. AND CHONG, C. 2009. A survey on TOA based wireless localization and NLOS mitigation techniques. *Communications Surveys & Tutorials, IEEE* 11, 3, 107–124.
- HE, T., HUANG, C., BLUM, B. M., STANKOVIC, J. A., AND ABDELZAHER, T. 2003. Range-Free Localization Schemes for Large Scale Sensor Networks. In *ACM MobiCom*. 81–95.
- HU, L. AND EVANS, D. 2004. Localization for Mobile Sensor Networks. In *ACM MobiCom*. 45–57. *ACM Transactions on Sensor Networks*, Vol. V, No. N, Month 20YY.

- IHLER, A., FISHER III, J., MOSES, R., AND WILLSKY, A. 2005. Nonparametric belief propagation for self-localization of sensor networks. *IEEE Journal on Selected Areas in Communications* 23, 4, 809–819.
- KHAN, U., KAR, S., AND MOURA, J. 2009. Distributed sensor localization in random environments using minimal number of anchor nodes. *Signal Processing, IEEE Transactions on* 57, 5, 2000–2016.
- KSCHISCHANG, F., FREY, B., AND LOELIGER, H. 2001. Factor graphs and the sum-product algorithm. *IEEE Transactions on information theory* 47, 2, 498–519.
- KUO, J. C. AND LIAO, W. 2007. Hop Count Distance in Flooding-Based Mobile Ad Hoc Networks With High Node Density. *IEEE Transactions on Vehicular Technology* 56, 1357–1365.
- KUROSE, J. AND ROSS, K. 2010. *Computer Networking: A Top Down Approach Featuring the Internet*. Pearson Education, Inc.
- LANGENDOEN, K. AND REIJERS, N. 2003. Distributed localization in wireless sensor networks: a quantitative comparison. *Computer Networks* 43, 4, 499–518.
- LI, M. AND LIU, Y. 2007. Rendered Path: Range-free localization in anisotropic sensor networks with holes. In *Proceedings of Mobicom*. ACM.
- LIM, H. AND HOU, J. 2005. Localization for anisotropic sensor networks. *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE* 1.
- MILLER, L. E. 2001. Distribution of Link Distances in a Wireless Network. *Journal of Research of the National Institute of Standards and Technology* 106, 2, 401–412.
- NAGPAL, R., SHROBE, H., AND BACHRACH, J. 2003. Organizing a global coordinate system from local information on an ad hoc sensor network. In *IPSN*. IEEE.
- NARAYANASWAMY, S., KAWADIA, V., SREENIVAS, R., AND KUMAR, P. R. 2002. Power control in ad hoc networks: Theory, architecture, algorithm and implementation of the compow protocol. In *European Wireless, Next Generation Wireless Networks: Technologies, Protocols, Services and Applications*.
- NATH, S. AND KUMAR, A. 2008. Performance Evaluation of Distance-Hop Proportionality on Geometric Graph Models of Dense Sensor Networks. In *Proceedings of Performance Evaluation Methodologies and Tools (Valuetools)*. ICST. Athens, Greece.
- NICULESCU, D. AND NATH, B. 2001. Ad hoc Positioning System (APS). In *IEEE Globecom*. IEEE.
- NICULESCU, D. AND NATH, B. 2003. DV Based Positioning in Ad Hoc Networks. *Telecommunication Systems* 22, 1, 267–280.
- OZGUR, A., LEVEQUE, O., AND TSE, D. 2007. Hierarchical Cooperation Achieves Optimal Capacity Scaling in Ad Hoc Networks. *IEEE Transactions on Information Theory* 53, 10 (October), 3549–3572.
- PEARL, J. 1988. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann.
- PENROSE, M. D. 2003. *Random Geometric Graphs*. Oxford University Press.
- TA, X., MAO, G., AND ANDERSON, B. D. O. 2007a. Evaluation of the probability of k-hop connection in homogeneous wireless sensor networks. In *IEEE Globecom 2007*.
- TA, X., MAO, G., AND ANDERSON, B. D. O. 2007b. On the probability of k-hop connection in wireless sensor networks. *IEEE Communications Letters* 11, 8, 662–664.
- VURAL, S. AND EKICI, E. 2005. Analysis of Hop-Distance Relationship in Spatially Random Sensor Networks. In *Mobihoc*. ACM.
- YANG, S., YI, J., AND CHA, H. 2007. HCRL: A Hop-Count-Ratio based Localization in wireless sensor networks. In *IEEE SECON*. IEEE.
- YEDIDIA, J., FREEMAN, W., AND WEISS, Y. 2003. Understanding belief propagation and its generalizations. *Exploring artificial intelligence in the new millennium*, 239–236.
- YEDIDIA, J., FREEMAN, W. T., AND WEISS, Y. 2001. Generalized belief propagation. In *In NIPS 13*. MIT Press, 689–695.
- ZORZI, M. AND RAO, R. R. 2003. Geographic random forwarding (GeRaF) for ad hoc and sensor networks: multihop performance. *IEEE Transactions on Mobile Computing* 2, 337–348.

### A. A LOWER BOUND WITH IMPROVED CONVERGENCE RATE

Previously, in Section 4.1, a lower bound to the Euclidean distance between any two nodes separated by a HD  $h$  was derived by considering the case when there was at least one node in each rectangular strip within each blade shown in Figure 4. It turns out that the rate at which the probability of the lower bound holding converges to 1 can be improved by replacing the rectangular strips by lens-shaped areas and this is the derivation presented here.

For simplicity, we confine ourselves to a setting where there is just a single anchor, located at the center of the unit area, thereby allowing us to stay clear of any edge effects. We adopt a coordinate system in which the coordinates of this center are  $(0, 0)$ . The interest here is in a lower bound to the Euclidean distance  $D_{l,i}(\mathbf{v})$ , given hop distance  $H_{l,i}(\mathbf{v})$ . Without loss of generality, we assume  $i = 1$ , so that we can abbreviate and write  $D(\mathbf{v})$  and  $H(\mathbf{v})$  in place of  $D_{l,i}(\mathbf{v})$  and  $H_{l,i}(\mathbf{v})$  respectively. We will also write  $r$  in place of  $r(n)$ .

Given that  $H(\mathbf{v}) = h$ , the triangle inequality gives us the upper bound  $D(\mathbf{v}) \leq hr$ . We will now provide a corresponding lower bound to  $D(\mathbf{v})$  that holds with high probability with this probability converging to 1 at a rate faster than that in the case of the blades-and-strips argument.

We have the construction as shown in Figure 23 shows a collection of intersecting circles drawn in the unit plane. All circles have centers that are located along the  $x$ -axis. The circles at the two ends have radius  $r$  (we term these big circles) while the circles in the middle have half this radius (and are termed small circles). The circles are drawn as follows. First, a big circle is drawn centered at the origin (i.e., from the location of the anchor). We next draw a small circle such that the intersection of the big circle, the small circle and the  $x$ -axis, is a line segment of width  $\delta r$ . We then draw a collection  $(h - 4)$  additional small circles such that the distance between the centers of any two adjoining small circles equals  $\alpha r(n)$ ,  $0 \leq \alpha \leq 1$ . Finally, we draw a big circle such that the intersection between the rightmost small circle, the big circle on the right and the  $x$ -axis is symmetrically, a line segment of width  $\delta r$ .

We will refer to the intersection of two adjoining circles as a lens. Thus there are a total of  $(h - 2)$  lenses in all, which we will label as lenses  $L_1$  through  $L_{h-2}$ , running from left to right. Unlike the  $(h - 4)$  lenses in the middle, the two lenses  $L_1, L_{h-2}$  at either end are asymmetric. Let  $a_s$  and  $a_{as}$  denote the areas of the symmetric and asymmetric lenses respectively.

Next let us define  $A_i$ ,  $1 \leq i \leq (h - 2)$  as the event that there is at least one node in lens  $L_i$ . Let  $\Gamma$  be the intersection event under which the deployment  $\mathbf{v}$  is such that  $\mathbf{v} \in A_i$ ,  $\forall 1 \leq i \leq h - 2$ . Then whenever  $\mathbf{v} \in \Gamma$ , there exists at least one node in each of the  $h - 2$  lenses (see Figure 23). It follows then from the figure that for  $\mathbf{v} \in \Gamma$ , all nodes at a distance  $< ((h - 4)\alpha + 3 - 2\delta)r(n)$  from the fixed anchor are reachable in at most  $h - 1$  hops, hence will have a hop distance  $\leq h - 1 < h$ . So, we have  $D(\mathbf{v}) \geq ((h - 4)\alpha + 3 - 2\delta)r$ , for such a deployment  $\mathbf{v}$ ; see Figure 23. It follows then that  $D(\mathbf{v}) \geq ((h - 4)\alpha + 3 - 2\delta)r$  given  $H(\mathbf{v}) = h$  with probability  $\geq Pr(\Gamma)$ .

We will now proceed to lower bound the probability of the event  $\Gamma$  and show that

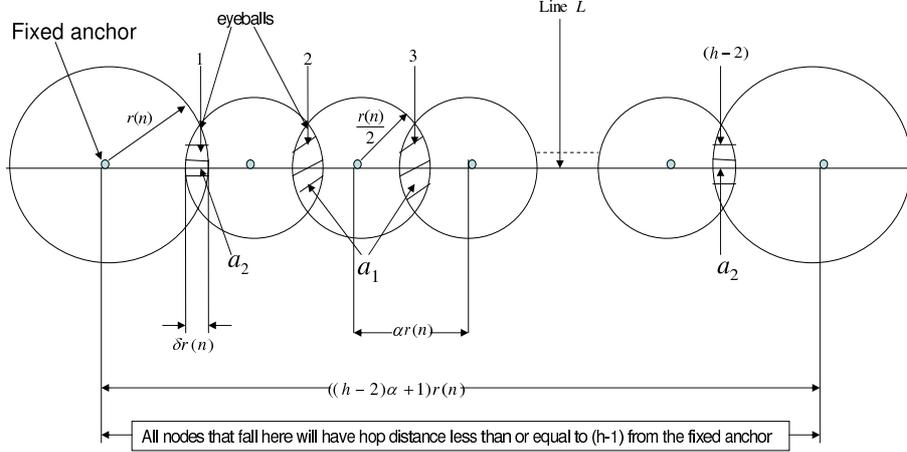


Fig. 23. Construction using lenses.

it converges to unity at a rate that is more rapid than for the case when one used strips in place of lenses. While one could attempt to optimize the choice of  $\delta$ , we have for the ease of calculations, taken  $\delta = 1 - \alpha$  (i.e., the width of each the  $(h - 2)$  lenses is now the same), which causes the lower bound on the Euclidean distance to equal  $r(n)((h - 2)\alpha + 1)$ . For this choice of  $\delta$ , we have the lower bound:

$$\begin{aligned}
 \mathbb{P}(\Gamma) &= \mathbb{P}^n (\cap_{i=1}^{h-2} A_i) \\
 &= 1 - \mathbb{P}^n (\cup_{i=1}^{h-2} A_i^c) \\
 &\geq 1 - \sum_{i=1}^{h-2} \mathbb{P}^n (A_i^c) \\
 &= 1 - \left[ \sum_{i=2}^{h-3} \mathbb{P}^n (A_i^c) + \mathbb{P}^n (A_1^c) + \mathbb{P}^n (A_{h-2}^c) \right] \\
 &= 1 - [(h - 4)(1 - a_s)^n + 2(1 - a_{as})^n] \\
 &\geq 1 - [(h - 4)e^{-na_s} + 2e^{-na_{as}}] \\
 &\geq 1 - (h - 2)e^{-na_s} \quad \text{since } a_s < a_{as} \\
 &= 1 - (h - 2)n^{-g'(\epsilon, h)c^2} \\
 &\xrightarrow{n \rightarrow \infty} 1
 \end{aligned} \tag{11}$$

We can carry out simple calculations to find out the expressions for  $a_s$  and  $a_{as}$ .

$$a_s = \frac{r^2(n)}{2}(\theta - \alpha\beta) \quad \theta = \cos^{-1} \alpha, \beta = \sqrt{1 - \alpha^2}$$

$$a_{as} = r^2(n) \left( \phi_1 - \frac{\sin 2\phi_1}{2} \right) + \frac{r^2(n)}{4} \left( \phi_2 - \frac{\sin 2\phi_2}{2} \right)$$

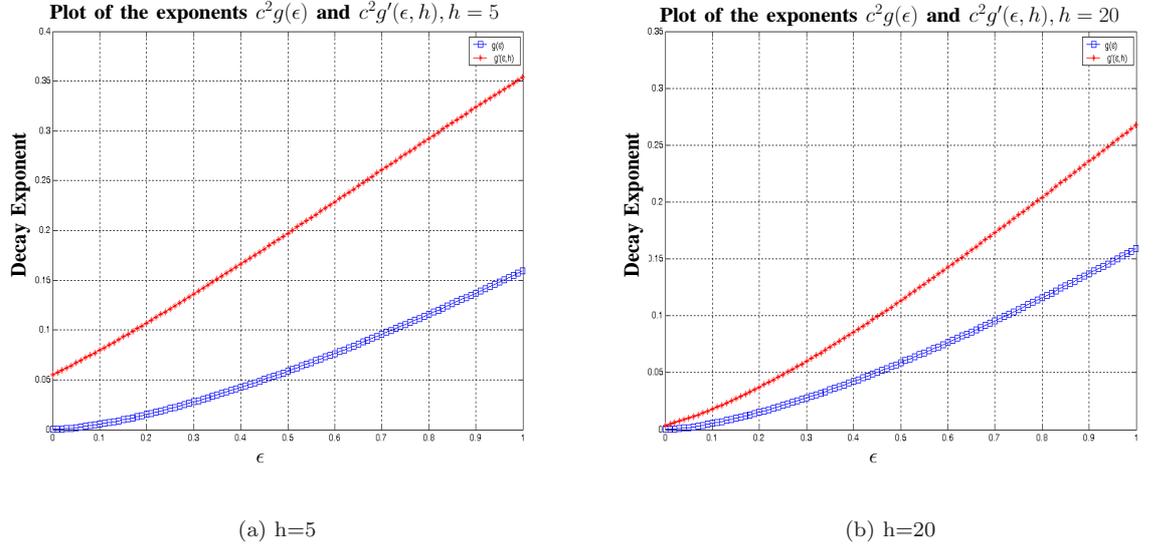


Fig. 24. Comparison of the exponents for hop counts 5 and 20.

Where,

$$\begin{aligned}\phi_1 &= \cos^{-1} \left( \frac{3}{4} - \frac{\delta}{2} - \frac{3}{2(6-4\delta)} \right) \\ \phi_2 &= \cos^{-1} \left( \frac{3}{2} - \delta - \frac{3}{6-4\delta} \right) \\ \delta &= 1 - \alpha\end{aligned}$$

For comparing this result with that of the previous, we put the lower bound on the Euclidean distance to be the following.

$$\begin{aligned}(h-1)(1-\epsilon)r(n) &= r(n)((h-2)\alpha + 1) \\ \alpha &= 1 - \frac{h-1}{h-2}\epsilon\end{aligned}$$

hence,  $g'(\epsilon, h) = \frac{1}{2}(\theta - \alpha\beta)$ . It depends on  $h$  as  $\alpha$  depends on  $h$ . In Figure 24, we show the convergence rate exponents  $g(\epsilon)$  and  $g'(\epsilon, h)$ . For same  $\epsilon$ , the exponent  $g'(\epsilon, h) > g(\epsilon)$ , which shows that the rate of convergence is more rapid than in the case of the blades-and-strips argument. As a reference point, we note that since the probability of a circular area of radius  $r(n)$  containing a node at its center, but not having a node anywhere else is on the order of  $n^{-1}$ , no lower bound can be guaranteed to hold with probability that converges to 1 at a rate faster than  $\frac{1}{n}$ .