A Geometrical Probability Approach to Location-Critical Network Performance Metrics

Yanyan Zhuang Jianping Pan
University of Victoria, Victoria, BC, Canada

***Abstract*—Node locations and distances are of profound importance for the operation of any communication networks. With the fundamental inter-node distance captured in a random network, one can build probabilistic models for characterizing network performance metrics such as k-th nearest neighbor and traveling distances, as well as transmission power and path loss in wireless networks. For the first time in the literature, a unified approach is developed to obtain the closed-form distributions of inter-node distances associated with hexagons. This approach can be degenerated to elementary geometries such as squares and rectangles. By the formulation of a quadratic product, the proposed approach can characterize general statistical distances when node coordinates are interdependent. Hence, our approach applies to both elementary and complex geometric topologies, and the corresponding probabilistic distance models go beyond the approximations and Monte Carlo simulations. Analytical models based on hexagon distributions are applied to the analysis of the nearest neighbor distribution in a sparse network for improving energy efficiency, and the farthest neighbor distribution in a dense network for minimizing routing overhead. Both the models and simulations demonstrate the high accuracy and promising potentials of this approach, whereas the current best approximations are not applicable in many scenarios. This geometrical probability approach thus provides accurate information essential to the successful network protocol and system design.**

***Index Terms*—Probabilistic distance distributions, geometric models, rhombuses, hexagons**

I. Introduction

In a spatially random network, e.g., a sensor network or cellular system, wireless devices are typically distributed over the network area according to a certain distribution. The locations and Euclidean distances between randomly deployed nodes are among the critical factors that determine the system performance metrics. For instance, the k-th nearest neighbor distance is crucial for relay and routing protocols [1]; stochastic mobility models are closely related to the trajectory between random points [2]; energy consumption in sensor networks [3], path loss, interference and capacity in wireless communication networks [4], etc, are all dependent on the location-critical random distances. The probabilistic models based on a fundamental investigation of the distance distributions between random points [5], provide accurate statistical information for protocol design and performance evaluation. They have become the powerful, versatile tools which are built upon the elegant theory of geometric probability, with a rich mathematical background [6].

Traditional methods providing statistical moments, particularly mean and variance, have been long existing in the literature. They give insights for empirical approximations and Monte Carlo simulations. Although these methods have maintained the scale of problems at a tractable level, a nonlinear relationship between location-critical performance metrics and inter-node distances has made the performance metrics much more complicated. Consequently, the numerical averages or empirical distributions have become less accurate, which is obvious from Jensen's inequality. If D is a random variable denoting inter-node distances, and $\varphi(\cdot)$ is a convex function (which is the case for path loss in wireless communication networks), then $\varphi(\mathbb{E}[D]) \leq \mathbb{E}[\varphi(D)]$. Therefore, the complex, non-deterministic nature of random networks makes pertinent a rigorous characterization of network performance metrics by means of the closed-form distance distributions, which is one of the contributions in this paper.

Previously, conducting distribution analysis on network performance metrics has been intractable for complex network geometries. Even for simple square and circular networks, Monte Carlo simulations and empirical approximations have been heavily used [8]. In this paper, we fill in the gap in the literature by developing a unified approach which is able to tackle a problem that has never been solved explicitly: the *closed-form* distance distributions associated with hexagons, a complex geometry commonly used in communication networks. The approach is unified in the sense that its degenerated form gives the exact same results for squares and rectangles as those in the classic geometric probability research. Moreover, the explicit distributions gracefully eliminate the errors in empirical methods. The novelty of this approach is twofold: first, different from [4], [13] and [14], there is no fixed reference point in the network; second, the coordinates of a node can be interdependent, in contrast to the strong assumption on coordinate independence in the literature [3], [5], [17], [18]. While it is relatively easy to derive the distance distribution with a fixed point in the network, the resultant geometrical probability approach only applies to the communication between random points and the fixed infrastructure. The results derived in this paper enable the analytical models in a wider spectrum with less location constraint.

The main contributions of this paper are as follows. First, we develop a novel yet simple formulation, achieving a unified approach to the closed-form node distance distribution functions. This formulation is a quadratic product of conditional probability function, and the probability density of node coordinates. Explicit distributions are given for both elementary and complex geometries, such as squares, rhombuses and hexagons. The results not only are suitable for

978-1-4673-0775-8/12/$31.00 ©2012 IEEE

convex topologies, but also apply to the networks with concave geometric shapes. Second, the rigorousness and accuracy of the derived distributions have been verified through both mathematical validation and simulations. We also illustrate the use of our probabilistic distance models in a computation-effective manner with polynomial fitting. Third, we show by analytical and simulation comparison that, in both sparse and dense network scenarios, the state-of-the-art approximations are not accurate when analyzing non-linear, location-critical network performance metrics. In summary, the probabilistic distance models presented in this paper bridge the gap between the distribution of network performance metrics, and the explicit distribution of random distances in complex geometries. The analytical results are critical to the fine tuning of network protocol parameters, and the accurate modeling of location-critical performance metrics.

The rest of this paper is organized as follows. In Section II we briefly review the research work in the field of geometric probability for random distances, and the applications of distance models for communication networks. In Section III, we present the derivation of the probabilistic distance distributions for hexagons, through a geometric integral of the quadratic product. The case studies in both sparse and dense networks are given in Section IV. Section V concludes the paper with future work.

II. Background and Related Work

Random distances associated with different geometric shapes, where node locations follow a certain distribution, have been a research problem with a long mathematical history. In this section we review the classic work with the main focus in the field of mathematics and statistics, and the application of these results in communication networks.

A. Geometrical Distribution of Node Distances

The study of the distribution of node distances dates back to the 1940's [17], [18], and the problem of deriving the *expected* distance between random points was listed as problem number 75-12 of the Society for Industrial and Applied Mathematics Review [19]. While this problem has drawn considerable attention from the literature (please refer to [20]–[22] and the references therein), obtaining the *distribution* of random distances, which leads to all statistical moments of the distance, turns out to be very challenging yet highly useful when dealing with the problems in random networks. Some research focused on the random distances when one of the endpoints is fixed [9], whereas the problem becomes especially difficult when both the endpoints are random.

[23] is among the first of these efforts, where the classical Crofton technique and its extensions were used for obtaining the geometrical distribution of node distances associated with circles and squares. Later, [9], [24], [25] showed a few simple geometrical cases where their distance distributions can be derived analytically. [5] in particular, is a collection of methods for distance distributions in different elementary geometric shapes. These methods, either using local or global perturbations, differential equations or elementary statistical techniques, provide key insights into our understanding of the probabilistic nature of random networks, including our previous work [3]. However, many of these efforts either studied random distances from a fixed reference point, or employed techniques that yield explicit distributions for very specific network topologies [5]. Particularly, the independence of node coordinates must be maintained [3], [17], [18].

B. Distance Distributions and Communication Networks

Although the relevant research on random distances has been conducted for a long time, only until recently have people in the field of networking research begun using this tool in the analytical modeling and optimization of network systems. In a spatially random network, nodes are typically distributed over an area or volume, following a certain distribution. The distance between these nodes plays an important role in determining several fundamental performance metrics, as listed briefly in the following.

Position-based routing and hop distance statistics. There are ad hoc routing protocols which make forwarding decisions based on the geographical location of a packet's destination [11]. Nodes are typically stationary, and a sequence of data forwarding results in different covered distances at each hop towards the destination. Hop distance [26] and hop count [1] statistics are critical to the reliability of message delivery, as well as the minimization of multihop energy consumption [3]. These performance statistics, determined by the pairwise distance between intermediate nodes, are crucial to the applications with energy constraints, yet requiring high message delivery ratio.

Stochastic properties of a random mobility model. A device is allowed to move randomly inside a given region, and its trajectory is then formed by a set of polylines between random points [7]. The stochastic process of the distances between random points is characterized by their spatial distribution. By also knowing the speed characteristics of the device, one can obtain its transition process among random points [2]. Similarly, the tour length of the traveling salesman problem (TSP) in a zone, with uniform demand density, can be reduced to the same problem [12]. The resultant travel delay is important to many time-sensitive applications where a minimal service latency is preferred.

Path loss, energy consumption, and interference. In a wireless communication channel, the strength of transmitted signals falls off with the distance between transceivers at rate α, the path loss exponent. Power control schemes are needed to determine the required transmission power to overcome this path loss. On the other hand, the energy required to successfully deliver a packet increases super-linearly with the distance between transceivers. The energy consumption (or received power) thus can be expressed as an α-moment of the (reciprocal) node distance [3], [13]. Meanwhile, the received signal at a receiver is superposed with other unintended signals being transmitted in the vicinity [14]. By knowing the statistical distribution of node distances, the cumulative

interference at the receiver can be modeled as an additive random variable. However, due to the intractability of the problem, approximations are used in most cases [14], [15].

SINR and channel capacity. A successful decoding of the received symbols is a random event with probability dependent on the ratio between the strength of the desired signal from an intended transmitter, and that of the unintended interference plus thermal noise, i.e., the signal-to-interference-and-noise ratio (SINR). Consequently, the channel capacity is determined by the SINR according to Shannon's Theorem [4], and this capacity dictates an acceptable modulation and coding scheme at a given symbol error rate.

All the metrics listed above are related to the node location and distance, which is in close relation to the network deployment (i.e., the distribution of network devices) and geometry (network shape and size). They are particularly important at the network planning and dimensioning stage. However, most existing work has only tackled the moments of the distances (means and variances) [14], characterized the exact distribution for simple network topologies [1], [3], [7], or the distribution with one fixed point [4], [13]. These assumptions have limited the current works to square and rectangular networks [1], [3], or between randomly located users and a fixed reference point [4]. Even though the knowledge of random distances is crucial in the networking research area, relatively little work has been done to give a general, unified formulation, and no explicit results are available in the literature for complex geometries such as hexagons.

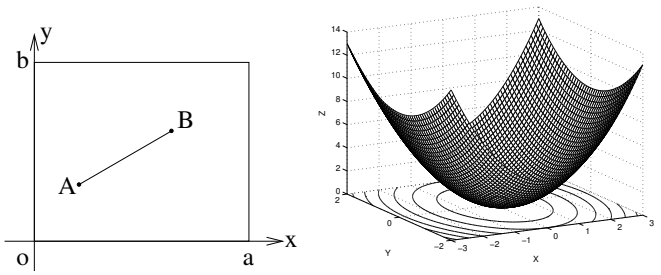
In this paper, the probabilistic distance models that we develop, are exactly filling the gap in the literature. The models are able to deal with both elementary and complex geometries, and convex and concave communication regions, through a simple but elegant formulation.

III. Random Distances Associated with Squares, Rhombuses and Hexagons

A. Formulation and Definitions

The Euclidean distance between random points, (X_1, Y_1) and (X_2, Y_2) on a Cartesian plane, $D = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2}$, is a function of the coordinate differences. Let random variable Z denote the squared Euclidean distance D^2, and $X = X_1 - X_2$ and $Y = Y_1 - Y_2$, then $Z = X^2 + Y^2$ is a function of X and Y. By going back to the definition of distance distribution, we develop a general, unified approach to obtaining the node distance distribution, by a quadratic product of the conditional probability function, and the density of random variables. Using geometric integral, this approach decomposes the geometrical constraints and distributions of random points, i.e.,

1) *The Probability Function* $F_\omega(Z|X, Y)$, given in the form of a conditional probability that is a function of the coordinate differences X and Y with geometrical constraints Ω on (X_1, Y_1) and (X_2, Y_2);



(a) Random Points in a Rectangle. (b) $Z = X^2 + Y^2$ when $a = 3$, $b = 2$.

Fig. 1. Random Distances within a Rectangle: the Geometrical Approach.

2) *The Probability Density* of the coordinate differences, $f_\omega(X, Y)$, where general distributions can be applied to (X_1, Y_1) and (X_2, Y_2).

By the definition of conditional probability, and Z as a function of X and Y as mentioned above, the corresponding formulation is given as follows

$$P_\Omega(Z \le z) = \iint_\omega P\left(Z(X,Y) \le z | X = x, Y = y\right) \cdot f_{X,Y}(x,y)dxdy, \quad (1)$$

where $F_\omega(Z|X, Y) = P\left(Z(X, Y) \le z|X = x, Y = y\right)$ is the conditional probability function, and $f_\omega(X, Y) = f_{X,Y}(x, y)$ is the probability density function. $P_\Omega(Z \le z)$ is obtained by the integration of Z projected onto the $\mathcal{X}$-$\mathcal{Y}$ plane according to ω, which is transformed from Ω. The intuition behind this product is the correspondence between a geometric shape and the characteristics of random coordinates within the shape. This formulation can be expressed in the same form, regardless of the network geometry and the node distribution. Starting from rectangles and squares, we show in this section that our approach can be extended to parallelograms and rhombuses. Using rhombuses as the building block, we finally extend the approach to hexagons.

Henceforth, we use the upper case X for a random variable, and lower case x for a sampled value of X, and $\mathcal{X}$ denotes the set of all possible values of X. We assume that the probability distribution of random points is uniform within the corresponding geometric shape, i.e., the probability measure of any Borel set is proportional to its area, and use notation $U[a, b]$ for a uniform distribution over interval $[a, b]$. Note that in many cases in the literature, the distribution of nodes is assumed to be a Poisson Point Process (PPP). Although the study of such networks is analytically convenient, PPP is only suitable for the scenario where the network size and node population is infinite. In practical networks, nodes are often distributed uniformly at random in a finite region, and the number of nodes in the network is also finite. Therefore we consider a uniform, finite network in the following. Nevertheless, the above formulation also applies to PPP by replacing f_ω in (1) with Skellam distribution [33].

B. Rectangles and Squares: the Geometrical Approach

In our previous work [3], we used a standard statistical approach to deriving the distribution of random distances between points within and between squares. The standard approach was a four-step process, where the distributions of difference $V = X_1 - X_2$ (or $Y_1 - Y_2$), square $S = V^2$, sum $Z = S_X + S_Y$ and square-root $D = \sqrt{Z}$ of random variables were derived, respectively. The drawback of this approach is, the condition that $Z = S_X + S_Y$ can be derived by convolution is based on the independence of S_X and S_Y, which requires the $\mathcal{X}$ and $\mathcal{Y}$ coordinates of each point on the plane to be stochastically independent. This assumption holds for both squares and rectangles.

[5], [25] used similar approaches and derived the distance distributions for rectangles. The derivation in [17], [18], on the other hand, is based on the joint distribution of $X_1 - X_2$ and $Y_1 - Y_2$, which leads to a closed form only when the corresponding point coordinates are independent. Successful as they were, the approaches with the strong coordinate independence assumption are difficult to extend to other geometrical shapes, such as the random points in parallelograms or rhombuses where the $\mathcal{X}$ and $\mathcal{Y}$ coordinates are interdependent. We show in this section that, using the formulation in (1), the distributions of random distances associated with rectangles, squares, parallelograms and rhombuses, can all be obtained in closed form, no matter whether the $\mathcal{X}$ and $\mathcal{Y}$ coordinates are interdependent or not.

1) Rectangles—an Illustration: Figure 1(a) shows a rectangle of size $a \times b$, with two random points $A(X_1, Y_1)$ and $B(X_2, Y_2)$. Using the same notation as that in [3], we have $Z = X^2 + Y^2$, where $X \in [-a, a]$ and $Y \in [-b, b]$. By the product formulation in (1), the distribution of Z is

$$\begin{aligned} F_Z(z) &= P_\Omega(Z \le z) = P(X^2 + Y^2 \le z) \\ &= \iint P(x^2 + y^2 \le z | X = x, Y = y) \\ &\qquad \cdot f_{X,Y}(x, y) dx dy \end{aligned} \tag{2}$$

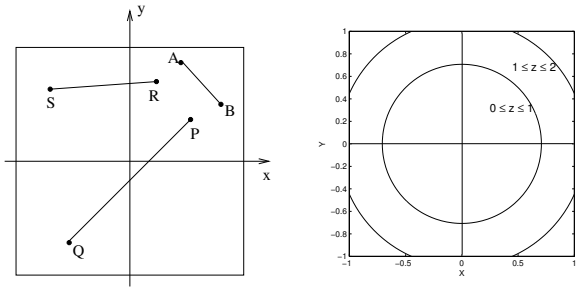
where $F_\omega = P(x^2 + y^2 \le z | X = x, Y = y)$ is the probability function given as a conditional probability, and $f_\omega = f_{X,Y}(x, y)$ is the joint probability density function of random variables X and Y. In rectangles, we have $f_{X,Y}(x, y) = f_X(x) f_Y(y)$, where $f_X(x)$ and $f_Y(y)$ are the (marginal) probability density functions of X and Y,

$$f_X(x) = \frac{1}{a^2} \begin{cases} a + x & -a \le x \le 0 \\ a - x & 0 \le x \le a \end{cases} \tag{3}$$

and

$$f_Y(y) = \frac{1}{b^2} \begin{cases} b + y & -b \le y \le 0 \\ b - y & 0 \le y \le b \end{cases}, \tag{4}$$

both following a symmetric triangular distribution. The geometric interpretation of (2) is shown in Fig. 1(b): a bowl cut off by the boundaries of (3) and (4) at $\mathcal{X} = \pm a$ and $\mathcal{Y} = \pm b$. Figure 1(b) also shows the contours of Z projected onto the $\mathcal{X}$-$\mathcal{Y}$ plane, each of which corresponds to a specific value of Z and together forms a series of concentric circles centered at $(0, 0)$. The above function $X^2 + Y^2 = Z$ can be used for both rectangles and squares, whereas a different function will be used for parallelograms and rhombuses. In the following we first illustrate the derivation of distance distributions for squares where $a = b$.



(a) Random Points in Squares. (b) Two Sub-cases for $Z = |AB|^2$.

Fig. 2. Random Distances Associated with Squares.

2) Distance Distributions Associated with Squares: There are three cases of random distances associated with squares as shown in Fig. 2(a): AB located within the same square; RS inside two parallel adjacent squares sharing a side; and PQ inside two adjacent squares having a common diagonal. Note that PQ is a random node pair communicating across a concave geometric shape.

The projections of Z on the $\mathcal{X}$-$\mathcal{Y}$ plane are shown in Fig. 2(b), assuming $a = b = 1$ for simplicity. Take $|AB|$ in Fig. 2(a) as an example, the $\mathcal{X}$-$\mathcal{Y}$ plane in Fig. 2(b) is divided into four compartments by $\mathcal{X} = 0$ and $\mathcal{Y} = 0$, the transitional values in (3) and (4). The area of the projection $x^2 + y^2 \le z$ in the four compartments is determined by a specific value of Z. Therefore, as the circular projection of Z expanding, we find three transitional values: $z_1 = 0$ where the circle is a point at the origin, $z_2 = 1$ where $\mathcal{X} = \pm 1$ and $\mathcal{Y} = \pm 1$ become the tangent lines of the circle, and $z_3 = 2$ where the radius of the circle is equal to $\sqrt{2}$ (or equal to $\sqrt{a^2 + b^2}$ for a rectangle). Given the three transitional values, we have two subcases when $z \in [z_1, z_2]$ and $[z_2, z_3]$, respectively:

i) $0 \le z \le 1$: The entire circle lies inside the boundary of f_ω, as shown by the small circle in Fig. 2(b). In the first quadrant where $0 \le x \le 1$ and $0 \le y \le 1$, we have $f_{X,Y}(x, y) = (1-x)(1-y)$. Therefore, $F_Z(z)$ in this quadrant is $\int_0^{\sqrt{z}} \int_0^{\sqrt{z-y^2}} (1 - x) dx (1 - y) dy$. Because of symmetry,

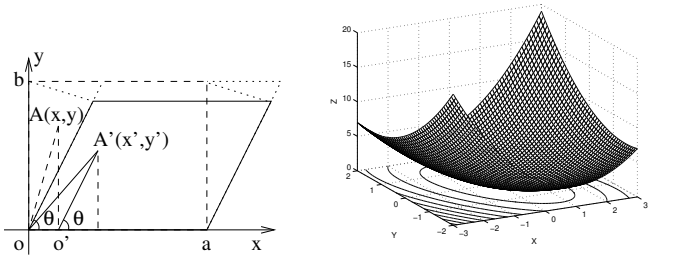
$$F_Z(z) = 4 \int_0^{\sqrt{z}} \int_0^{\sqrt{z-y^2}} (1-x) dx (1-y) dy = \frac{z^2}{2} - \frac{8}{3} z^{3/2} + \pi z.$$

The probability density function (PDF) is the derivative of $F_Z(z)$, i.e.,

$$f_Z(z) = F_Z'(z) = z - 4\sqrt{z} + \pi.$$

ii) $1 \le z \le 2$: Part of the circle is cut off by the boundary of f_ω. The corresponding $f_Z(z)$ can be derived similarly:

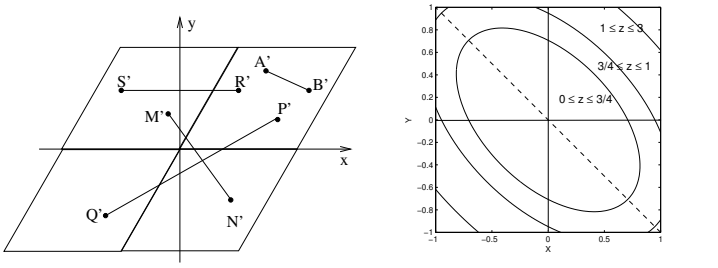
$$f_Z(z) = 2 \sin^{-1} \left(\frac{2}{z} - 1 \right) - z + 4\sqrt{z - 1} - 2.$$



(a) A Rectangle Becomes a Parallelogram.

(b) $Z = X^2 + 2\cos\theta XY + Y^2$ when $a = 3$, $b = 2$ and $\theta = \frac{\pi}{3}$.

Fig. 3. A Rectangle Squeezed into a Parallelogram.



(a) Random Points in Rhombuses.

(b) Three Sub-cases for $Z = |A'B'|^2$.

Fig. 4. Random Distances Associated with Rhombuses.

With $D = \sqrt{Z}$, the distance distribution $f_D(d)$ is

$$f_D(d) = F'_Z(d^2) = 2df_Z(d^2). \tag{5}$$

$f_D(d)$ for $|SR|$ and $|PQ|$ in Fig. 2(a) are derived in a similar way, by shifting the transitional values in (3) and (4), and plugging (5) into the corresponding $F_Z(z)$. The results obtained here agree with the previous work [3], [5], [17], [18], [25]. However, the new derivation does not rely on the assumption of coordinate independence, and does not require complicated convolution.

Although unit squares are assumed throughout above, the distance distribution can be easily scaled by an arbitrary nonzero scalar. Let the scalar be s, then

$$F_{sD}(d) = P(sD \le d) = P(D \le \frac{d}{s}) = F_D(\frac{d}{s}).$$

Therefore,

$$f_{sD}(d) = F'_D(\frac{d}{s}) = \frac{1}{s}f_D(\frac{d}{s}). \tag{6}$$

C. Parallelograms & Rhombuses: An Intermediate Geometry

1) Parallelograms—the Squeezed Rectangles: From a geometric perspective, parallelograms and rhombuses are the *squeezed* version of rectangles and squares, respectively. Suppose that initially, a point $\mathrm{A}(x, y)$ lies in the rectangle shown in Fig. 3(a). This point forms a right triangle OAO' with the $\mathcal{X}$-axis. Squeezing the rectangle by $\frac{\pi}{2} - \theta$ (assuming $0 \le \theta \le \frac{\pi}{2}$), $\mathrm{A}(x, y)$ becomes $\mathrm{A}'(x', y')$ in the parallelogram, which forms an obtuse triangle $\mathrm{OA'O'}$ with the $\mathcal{X}$-axis. From A to A′, there is the following affine transformation [34]:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & \cos\theta \\ 0 & \sin\theta \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}, \quad \text{or} \quad \begin{cases} x' = x + y\cos\theta \\ y' = y\sin\theta \end{cases}. \tag{7}$$

Therefore, a rectangle is a degenerated case of a parallelogram when $\theta = \frac{\pi}{2}$. In the following, we use (X'_1, Y'_1) and (X'_2, Y'_2) (and (X_1, Y_1) and (X_2, Y_2)) as the random variables denoting the coordinates after (and before) the squeeze transformation in (7). For the points within the same parallelogram,

$$\begin{aligned} Z &= D^2 = (X')^2 + (Y')^2 = (X'_1 - X'_2)^2 + (Y'_1 - Y'_2)^2 \\ &= [(X_1 - X_2) + \cos\theta(Y_1 - Y_2)]^2 + [\sin\theta(Y_1 - Y_2)]^2 \\ &= (X_1 - X_2)^2 + 2\cos\theta(X_1 - X_2)(Y_1 - Y_2) + (Y_1 - Y_2)^2 \end{aligned} \tag{8}$$

where (X_1, Y_1) and (X_2, Y_2) are the corresponding coordinates in the original rectangle. As the value of θ changes, only the coefficient $2\cos\theta$ will change in (8). Still let $X = X_1 - X_2$ and $Y = Y_1 - Y_2$ be the difference of $\mathcal{X}$ and $\mathcal{Y}$-coordinates *before* the transformation, then $Z = X^2 + 2\cos\theta XY + Y^2$. In analytic geometry, this equality satisfies the implicit equation of a non-degenerated real ellipse [30] when $\theta \ne \frac{\pi}{2}$. The probability function therefore becomes $P(x^2 + 2\cos\theta xy + y^2 \le z | X = x, Y = y)$. When $\theta = \frac{\pi}{2}$, (8) degenerates to the formulation for a rectangle as shown in (2).

Without loss of generality, for a parallelogram of side length of a and b, and with uniformly distributed random points, then $X_1, X_2 \sim U[0, a]$ and $Y_1, Y_2 \sim U[0, b]$ in the original rectangle. Therefore, the distributions of X and Y are the same as in (3) and (4). The geometric interpretation of $Z = X^2 + 2\cos\theta XY + Y^2$ in the three-dimensional space, assuming $a = 3$, $b = 2$ and $\theta = \frac{\pi}{3}$, is shown as the squeezed bowl in Fig. 3(b), where the projections of Z on the $\mathcal{X}$-$\mathcal{Y}$ plane are concentric ellipses centered at $(0, 0)$, with cutoffs at $\mathcal{X} = \pm a$ and $\mathcal{Y} = \pm b$. It is interesting to observe that after a rectangle has been squeezed to a parallelogram, the shape of $Z = D^2$ has also been squeezed, and the projections on the $\mathcal{X}$-$\mathcal{Y}$ plane are squeezed from circles to ellipses. The level of the squeeze from circles to ellipses in the projected domain is determined by the same squeeze from rectangles to parallelograms.

2) Distance Distributions Associated with Rhombuses: Define a *unit rhombus* as the rhombus with an acute angle of $\theta = \frac{\pi}{3}$ and a side length of 1. The coordinates of a point in the rhombus are $x' = x + \frac{y}{2}$ and $y' = \frac{\sqrt{3}}{2}y$, according to (7), where x and y are the coordinates in the original unit square and follow $U[0, 1]$. Thus,

$$Z = D^2 = X^2 + XY + Y^2. \tag{9}$$

Note that this method can be applied to more general parallelograms and rhombuses where $\theta \in [0, \frac{\pi}{2}]$. By Z given in (9), the distribution becomes

$$\begin{aligned} F_Z(z) &= P_\Omega(Z \le z) = P(X^2 + XY + Y^2 \le z) \\ &= \iint \left[P(x^2 + xy + y^2 \le z | X = x, Y = y) \right. \\ &\qquad \left. \cdot f_{X,Y}(x, y)\right] dxdy. \end{aligned} \tag{10}$$

Consequently, the elliptical projections of Z on the $\mathcal{X}$-$\mathcal{Y}$ plane, are cut off by $\mathcal{X} = \pm 1$ and $\mathcal{Y} = \pm 1$, the boundaries of f_ω. There are four cases of the geometric locations of two random points, when rhombuses are adjacent and similarly

oriented, as shown in Fig. 4(a): i.e., A′B′ that are within the same rhombus; R′S′ that are inside two parallel rhombuses sharing a side; P′Q′ and M′N′ that are inside two rhombuses sharing a common diagonal. Here P′Q′ and M′N′ are two different cases between two points in a concave geometry.

Take $|A'B'|$ in Fig. 4(a) for example. Because of symmetry, there are only four transitional values of Z as the ellipse expanding: z_1 where the ellipse is a point at the origin, z_2 where $\mathcal{X} = \pm 1$ or $\mathcal{Y} = \pm 1$ become the tangent lines of the ellipse, z_3 where the semimajor axis of the ellipse is equal to $\sqrt{a^2 + b^2} = \sqrt{2}$, and z_4 where the semiminor axis of the ellipse is equal to $\sqrt{2}$. The semimajor axis and semiminor axis of ellipse $X^2 + XY + Y^2 = Z$ are $\sqrt{2Z}$ and $\sqrt{\frac{2}{3}Z}$, and thus $z_1 = 0$, $z_2 = \frac{3}{4}$, $z_3 = 1$ and $z_4 = 3$. When $0 \le z \le \frac{3}{4}$, we have

$$\begin{aligned} F_Z(z) &= 2\left[\int_0^{\sqrt{z}} \int_0^{-\frac{y}{2}+\sqrt{z-\frac{3}{4}y^2}} (1-x)dx(1-y)dy \right. \\ &\quad \left. +2\int_{-\sqrt{z}}^{0} \int_{-y}^{-\frac{y}{2}+\sqrt{z-\frac{3}{4}y^2}} (1-x)dx(1+y)dy\right] \\ &= \left(\frac{2}{3} + \frac{\pi}{9\sqrt{3}}\right) z^2 - \frac{32}{9} z^{3/2} + \frac{2\pi}{\sqrt{3}} z, \end{aligned}$$

and

$$f_Z(z) = \left(\frac{4}{3} + \frac{2\pi}{9\sqrt{3}}\right) z - \frac{16}{3}\sqrt{z} + \frac{2\pi}{\sqrt{3}}.$$

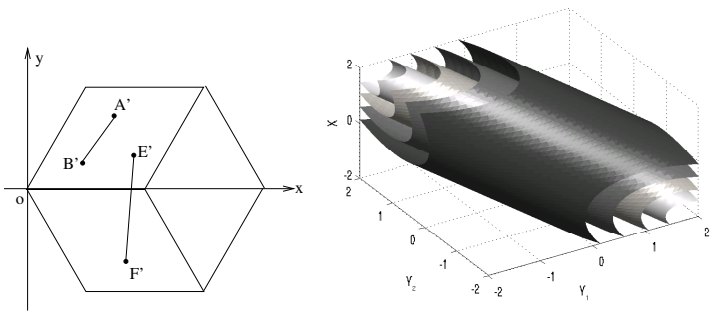
When $\frac{3}{4} \le z \le 1$, the ellipse intersects with boundaries $\mathcal{X} = \pm 1$ and $\mathcal{Y} = \pm 1$. $f_Z(z)$ can be derived by careful integration along the common boundary determined by both the ellipse and f_ω, and the result for this subcase is

$$\begin{aligned} f_Z(z) &= \frac{8}{\sqrt{3}}\left(1 + \frac{z}{3}\right) \sin^{-1}\frac{\sqrt{3}}{2\sqrt{z}} + \left(\frac{4}{3} - \frac{10\pi}{9\sqrt{3}}\right) z \\ &\quad + \frac{10}{3}\sqrt{4z-3} - \frac{16}{3}\sqrt{z} - \frac{2\pi}{\sqrt{3}}. \end{aligned}$$

The rest of the derivation can be done in the same way. The results for $|S'R'|$, $|P'Q'|$ and $|M'N'|$ shown in Fig. 4(a) are summarized in our technical report [28]. This is, for the first time in the literature, that the distribution functions of random distances are derived for rhombuses. Also note that in a cellular system, when hexagonal cells are split into different sectors using directional antennas, the covered area of each sector resembles a rhombus. Being the building block of a hexagon, rhombuses bridge the gap between the elementary geometry of squares, and the complex geometry of hexagons.

D. Distance Distributions Associated with Hexagons

Define a *unit hexagon* as the regular hexagon with a side length of 1. Two random endpoints of a given link inside a unit hexagon, as shown in Fig. 5(a), will fall into either one of the two cases: both endpoints are inside the same unit rhombus, e.g., A′B′, with probability $\frac{1}{3}$; each endpoint falls into one of the two adjacent rhombuses sharing a side, e.g., E′F′, with probability $\frac{2}{3}$. Denoting the distribution for $|A'B'|$ as $f_{D_{\mathrm{R}_1}}(d)$, and $|E'F'|$ as $f_{D_{\mathrm{R}_2}}(d)$, the probability density function of



(a) Relationship between Rhombuses and a Hexagon. (b) Geometric Interpretation of (11).

Fig. 5. Random Distances Associated with a Hexagon.

the random Euclidean distances between two endpoints in a hexagon is $\frac{1}{3} f_{D_{\mathrm{R}_1}}(d) + \frac{2}{3} f_{D_{\mathrm{R}_2}}(d)$, i.e., a probabilistic sum.

Following the same squeeze transformation as that in (7), and denoting the coordinates as $\mathrm{E}'(X_1', Y_1')$ and $\mathrm{F}'(X_2', Y_2')$, then $X_1' = X_1 + \frac{Y_1}{2}$, $Y_1' = \frac{\sqrt{3}}{2} Y_1$, $X_2' = X_2 + \frac{Y_2}{2}$, and $Y_2' = -\frac{\sqrt{3}}{2} Y_2$, where X_1, Y_1, X_2 and $Y_2 \sim U[0,1]$. We then have the squared Euclidean distance Z as

$$\begin{aligned} Z &= (X_1' - X_2')^2 + (Y_1' - Y_2')^2 \\ &= \left[(X_1 - X_2) + \frac{1}{2}(Y_1 - Y_2)\right]^2 + \left[\frac{\sqrt{3}}{2}(Y_1 + Y_2)\right]^2 \\ &= (X_1 - X_2)^2 + (X_1 - X_2)(Y_1 - Y_2) + Y_1^2 + Y_1 Y_2 + Y_2^2 \end{aligned}$$

Let $X = X_1 - X_2$, and thus

$$Z = X^2 + X(Y_1 - Y_2) + Y_1^2 + Y_1 Y_2 + Y_2^2 \qquad (11)$$

Note that the two rhombuses where E′ and F′ are located, are flipped with respect to the $\mathcal{X}$-axis, with different orientations, which differentiates them from all the four cases shown in Fig. 4. Therefore, for $|E'F'|$, the rhombus technique needs to be further extended. The geometric interpretation of (11) with different values, or the isosurfaces of Z, is concentric cylinders as shown in Fig. 5(b) when the value of Z increases. Z is hence a four-dimensional geometric shape. The projection ω is then drawn on a three-dimensional effective region, defined by $f_\omega = f_{X,Y_1,Y_2}(x, y_1, y_2)$. The final results of the distance distributions inside a hexagon, as well as between two adjacent hexagons sharing a side, are given in our technical report [29]. The distance distributions for random distances in hexagons are also obtained for the first time in the literature.

E. Verification of Distance Distributions

Looking at Fig. 2(a), the four adjacent unit squares together resemble a large square, with a side length of 2. Assuming that the distance distribution for $|AB|$ in Fig. 2(a) is $f_{D_{\mathrm{S}_1}}(d)$, and those for $|SR|$ and $|PQ|$ are $f_{D_{\mathrm{S}_2}}(d)$ and $f_{D_{\mathrm{S}_3}}(d)$. Then according to (6), the distance distribution in the large square is

$$f_{2D}(d) = \frac{1}{2} f_{D_{\mathrm{S}_1}}\left(\frac{d}{2}\right). \qquad (12)$$

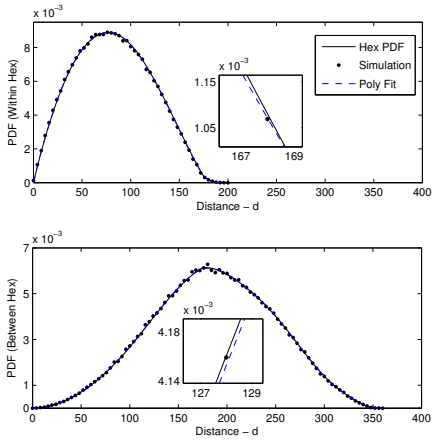


Fig. 6. Distance Distributions of Hexagons, Simulation and Polynomial Fit.

On the other hand, the two random endpoints of a given link inside the large square will fall into one of the three cases: both endpoints are inside the same unit square, with probability $\frac{1}{4}$; each of the two endpoints falls into one of the two parallel squares, with probability $\frac{1}{2}$; both endpoints fall into two diagonal squares, with probability $\frac{1}{4}$. The distance density function for the large square can also be given by a probabilistic sum,

$$f_{2D}(d) = \frac{1}{4} f_{D_{S_1}}(d) + \frac{1}{2} f_{D_{S_2}}(d) + \frac{1}{4} f_{D_{S_3}}(d). \qquad (13)$$

It can be verified that the RHS of (12) and (13) are equivalent. The same verification can be applied to both rhombuses and hexagons, with similar scaling functions and probabilistic sums, the details of which are given in [28] and [29]. The corresponding results are a strong validation of the correctness of the distance distribution functions that we have derived.

[28] and [29] also give the statistical moments, variances, and polynomial approximations of the distance distribution functions, which facilitate the practical utilization of the distribution functions. Polynomial fit in particular, makes it especially convenient to utilize these results, meanwhile maintaining the level of accuracy. Our previous work [3] shows examples of generating and using some of the polynomials, in modeling the energy consumption in wireless sensor networks.

Figure 6 plots the PDFs of the distance distributions within a hexagon and between two adjacent hexagons, of which the side lengths are both $s = 100\,m$ using (6). The validity of the PDFs is also cross-checked using Monte Carlo simulation results by $100,000$ randomly generated node locations in one hexagon, and between two hexagons. Figure 6 also plots the polynomial fits given in [29] in the zoom-in subfigure. In both cases, the polynomials fit very well with the analytical distributions, and the differences are negligible until zoomed in.

IV. Performance Study Using Distance Distributions

Using the distance distribution functions derived in Section III, we are able to further analyze the performance metrics

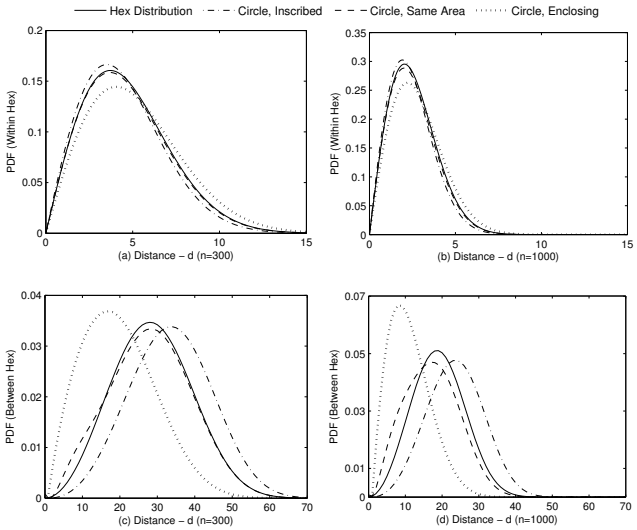


Fig. 7. Nearest Neighbor Distribution.

discussed in Section II. Starting from a challenging scenario where nodes are sparsely deployed, we analyze the nearest neighbor of a node in a hexagonal network. In a dense network, we analyze the farthest neighbor that a node can reach through a single hop. By choosing appropriate neighbors for packet relay, the underlying routing protocols benefit from the minimization of energy consumption and routing overhead in these two scenarios, respectively.

Previously, the distance distributions of circles have been widely used to approximate a hexagon: the inscribed or enclosing circles, and the circles with the same area as the hexagon [16]. When the communication involves adjacent hexagons, however, circular geometries either overlap or leave gaps between each other. Therefore, we also show that when compared with the explicit hexagon distribution models, traditional circular approximations are less accurate or even not applicable, in analyzing both the statistical distribution and expected average of the system performance metrics.

A. Sparse Network Scenario: the Nearest Neighbor

In a sparse network where the network size is larger than the communication range of a node, an energy-efficient routing protocol will choose the nearest neighbor as the packet forwarder in order to conserve energy at each node. This is particularly important when the network device has very limited energy supply. Suppose there are n nodes located in the same area, for a random node i, the minimum distance from $n-1$ nodes to i is

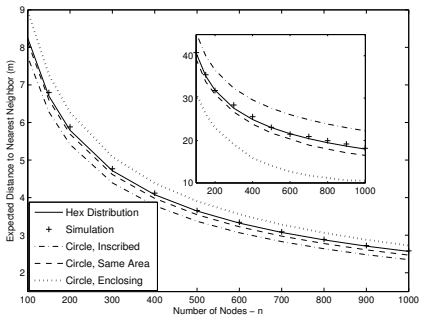
$$\delta = \min\{d_1, d_2, \cdots, d_{n-1}\}. \qquad (14)$$

The distribution function of δ is

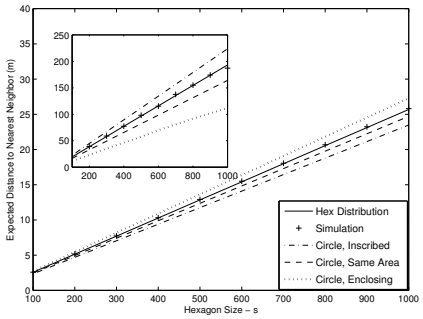
$$F_\Delta(\delta) = 1 - P(d_i \geq \delta)^{n-1} = 1 - [1 - F_D(\delta)]^{n-1},$$

where $F_D(\cdot)$ is any of the distance distribution CDFs derived in the last section. The PDF of δ is thus

$$f_\Delta(\delta) = (n-1) f_D(\delta) [1 - F_D(\delta)]^{n-2}. \qquad (15)$$



(a) Distance vs. No. of Nodes.



(b) Distance vs. Network Size.

Fig. 8. Expected Distance to Nearest Neighbor.

Similarly, the k-th nearest neighbor to node i can be derived by order statistics. In the following we use hexagonal topology since it is commonly used in communication networks. Figure 7(a)–(d) show the corresponding PDFs in (15) using the hexagon distribution functions we derived. It also shows those three circular approximations, of which the distance distribution functions are given in [10] for non-overlapping circles. When circles overlap, results from Monte Carlo simulations are used, because no closed-form PDFs are available in the literature except for empirical approximations on CDFs [31].

Figure 7(a) and (b) plot the PDFs in (15) using distance distribution *within* a hexagon, when the side length of the hexagon is $100\ m$, and the total number of nodes is $n = 300$ and $1,000$, respectively. When the network size is fixed, the increased number of nodes makes the nearest neighbor closer to node i. Also note that, the tails of the PDFs are truncated at $15\ m$ in the figure. Although the longest distance between nodes can be $200\ m$, it is very likely that the nearest neighbor is only a few meters away. Figure 7(c) and (d) use the distribution *between* two adjacent hexagons, with the same parameters as Fig. 7(a) and (b), and are truncated at $70\ m$. In these two figures, none of the circular approximations give a good fit to the hexagon distributions. Intuitively, inscribed circles skew nodes apart from the hexagon boundary, whereas same-area circles and enclosing circles both include areas not belonging to the hexagon.

The expectation of the distance to the nearest neighbor can be obtained by integrating (15) over the hexagonal area. The results are given in Fig. 8(a) and (b). In both figures, the zoom-in subfigures are between adjacent hexagons. In Fig. 8(a), the

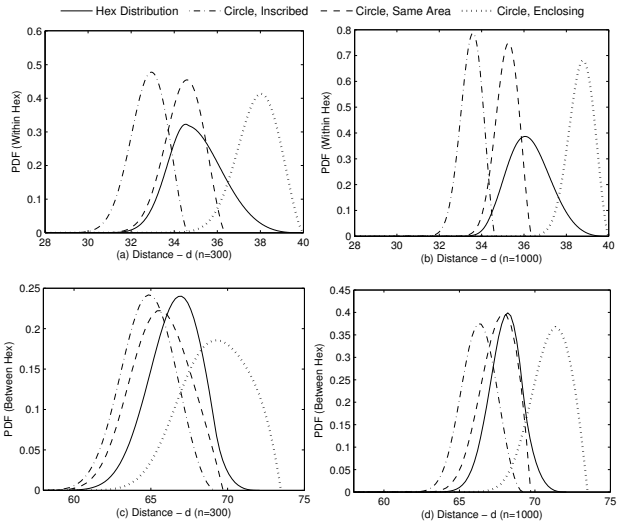


Fig. 9. Farthest Neighbor Distribution.

hexagon size is fixed at $100\ m$, and Fig. 8(b) fixes the number of nodes at $1,000$. From Fig. 8(a), if the underlying routing protocol chooses the nearest neighbor within a hexagon of size $100\ m$, the expected distance is less than $8\ m$ with more than 100 nodes, whereas an average distance of $82.6\ m$ is needed to reach an arbitrary node [29]. We may also observe that in a single hexagon, circles can approximate relatively well; but when nodes communicating between hexagons, the error of circular approximations grows super-linearly. It is also interesting to see in Fig. 8(b) that the distance to the nearest neighbor increases linearly with the network size because of the uniform distribution.

B. Dense Network Scenario: the Farthest Neighbor

In a sparse network, it is highly desirable to choose the nearest neighbor for improving energy efficiency. In a small, densely deployed network, on the other hand, it is common for a node to have several neighbors simultaneously. As a result, the number of transmissions and routing overhead can be minimized by choosing the farthest node as the packet forwarder. Suppose the total number of nodes is n, for node i define

$$\gamma = \max\{d_1, d_2, \cdots, d_{n-1}\}. \tag{16}$$

The distribution of γ has CDF

$$F_\Gamma(\gamma) = P(d_i \le \gamma)^{n-1} = F_D^{n-1}(\gamma),$$

and the corresponding PDF

$$f_\Gamma(\gamma) = (n-1) f_D(\gamma) F_D^{n-2}(\gamma). \tag{17}$$

Figure 9(a) and (b) plot the PDFs in (17) using distance distribution within a hexagon, of which the side length is $20\ m$, and the total number of nodes is $n = 300$ and $1,000$, respectively. The increased number of nodes brings the farthest neighbor closer to the network boundary at $40\ m$. Thus the left tails of the PDFs are truncated at $28\ m$. Figure 9(c) and (d)

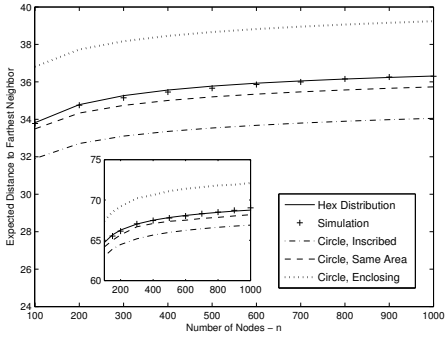


Fig. 10. Expected Distance to Farthest Neighbor.

show the distribution between two adjacent hexagons with the same parameters, and are truncated at $58\ m$. In Figure 9(a)–(d), it is obvious that none of the circular topologies give a good approximation for hexagons.

Figure 10 shows the expected distance to the farthest neighbor, with increasing number of nodes. The zoom-in subfigure shows that between two adjacent hexagons. Compared with the nearest neighbor, the increase in the number of nodes does not improve the expected distance much, since the farthest node is already around the network boundary. We can anticipate that in a dense network, single hop is very likely to reach the farthest neighbor. The underlying routing protocol thus can achieve the minimal routing overhead.

V. Conclusions

In this paper, we develop a unified approach to deriving the distribution of random distances, by a quadratic product of the probability function and the probability density. The closed-form probability density functions of the random distances associated with hexagons, are derived for the first time in the literature. We further use the results to investigate the nearest and farthest neighbor statistics in both sparse and dense network scenarios. The analytical and simulation results show the high accuracy and promising potentials of this approach. Our future work includes deriving the conditional probability of the distance distributions associated with hexagons, which can be utilized to model location-dependent, additive interference, and cooperative communications. We believe the probabilistic models presented in this paper and their future extensions will provide important guidelines for more accurate network dimensioning and better protocol design.

References

[1] C. Bettstetter and J. Eberspacher, "Hop distances in homogeneous ad hoc networks", in *Proc. 57th Vehicular Technology Conference (VTC'03)*, pp. 2286–2290, 2003.

[2] C. Bettstetter, H. Hartenstein and X. Pérez-Costa. "Stochastic properties of the random waypoint mobility model", *Wireless Networks*, 10(5):555–567, 2004.

[3] Y. Zhuang, J. Pan, and L. Cai, "Minimizing energy consumption with probabilistic distance models in wireless sensor networks", in *Proc. 29th IEEE International Conference on Computer Communications (INFOCOM'10)*, pp. 2453–2461, 2010..

[4] Y. Zhuang, Y. Luo, J. Pan, and L. Cai, "A geometric probability model for capacity analysis and interference estimation in wireless mobile cellular systems", in *Proc. 54th Global Telecommunications Conference (GLOBECOM'11)*, 2011.

[5] A. M. Mathai, *An introduction to geometrical probability: distributional aspects with applications*, Gordon and Breach Science, 1999.

[6] D. Stoyan, W. S. Kendall and J. Mecke, "Stochastic geometry and its applications", Wiley & Sons, 1978.

[7] L. He, Y. Zhuang, J. Pan, and J. Xu, "Evaluating on-demand data collection with mobile elements in wireless sensor networks", in *Proc. 72th Vehicular Technology Conference (VTC'10)*, 2010..

[8] K. B. Baltzis, "Empirical description of node-to-node distance density in non-overlapping wireless networks", *Journal of Microwaves, Optoelectronics and Electromagnetic Applications*, 9(1): 57–68, 2010.

[9] A. M. Mathai, "Random distances associated with triangles", *Int'l J of Mathematical and Statistical Sciences*, 7(1):77–96, 1998.

[10] A. M. Mathai and G. Pederzoli, "Random points with reference to a circle, revisited", *Rendiconti del Circolo Matematico di Palermo, Serie II, Suppl.*, 50: 235–258, 1997.

[11] M. Mauve, A. Widmer and H. Hartenstein, "A survey on position-based routing in mobile ad hoc networks", *IEEE Network*, 15(6): 30–39, 2001.

[12] L. Quadrifoglio and R. W. Hall, etc, "Performance and design of mobility allowance shuttle transit services: bounds on the maximum longitudinal velocity", *Transportation Science*, 40(3): 351–363, 2006.

[13] K. B. Baltzis, "Analytical and closed-form expressions for the distribution of path loss in hexagonal cellular networks", *Wireless Personal Communications*, 60(4):599–610, 2010.

[14] S. Srinivasa and M. Haenggi, "Modeling interference in finite uniformly random networks", in *Proc. International Workshop on Information Theory for Sensor Networks (WITS'07)*, 2007.

[15] X. Wang and L. Cai, "Interference analysis of co-existing wireless body area networks", in *Proc. 54th Global Telecommunications Conference (GLOBECOM'11)*, 2011.

[16] J. S. Evans and D. Everitt, "On the teletraffic capacity of CDMA cellular networks", *Vehicular Technology, IEEE Transactions on*, 48(1): 153–165, 1999.

[17] B. Ghosh, "On the distribution of random distances in a rectangle", *Science and Culture*, 8: 388, 1943.

[18] B. Ghosh, "Random distances within a rectangle and between two rectangles", *Bull. of the Calcutta Mathematical Society*, 43: 17–24, 1951.

[19] J. D. Murchland and D. J. Daley, "Problem 75-12", *SIAM Review*, 18(3): 497–501, 1976.

[20] T. S. Hale, "A taxonomy of expected distance functions with applications to facility location problems", Ph.D. dissertation, Texas A&M University, College Station, Texas, 1997.

[21] D. J. Gates, "Asymptotics of two integrals from optimization theory and geometric probability", *Advances in Applied Probability*, 17(4): 908–910, 1985.

[22] R. E. Stone, "Technical note—some average distance results", *Transportation Science*, 25(1): 83–90, 1991.

[23] V. S. Alagar, "The distribution of the distance between random points", *J of Applied Probability*, 13(3): 558–566, 1976.

[24] T. K. Sheng, "The distance between two random points in plane regions", *Advances in Applied Probability*, 17(4): 748–773, 1985.

[25] A. M. Mathai, P. G. Moschopoulos and G. Pederzoli, "Random points associated with rectangles", *Rendiconti del Circolo Matematico di Palermo*, XLVIII: 163–190, 1999.

[26] L. E. Miller, "Distribution of link distances in a wireless network", *Journal of Research of the National Institute of Standards and Technology*, 106(2): 401–412, 2001.

[27] S. Srinivasa and M. Haenggi, "Distance distributions in finite uniformly random networks: theory and applications", *Vehicular Technology, IEEE Transactions on*, 59(2): 940–949, 2010.

[28] Y. Zhuang and J. Pan, "Random distances associated with rhombuses", ArXiv Technical Report, arXiv: 1106.1257, 2011.

[29] Y. Zhuang and J. Pan, "Random distances associated with hexagons", ArXiv Technical Report, arXiv: 1106.2200, 2011.

[30] J. D. Lawrence, "A catalog of special plane curves", Dover Publ, 1972.

[31] K. B. Baltzis, "A geometric method for computing the nodal distance distribution in mobile networks", *Progress In Electromagnetics Research*, 114: 159–175, 2011.

[32] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *Wireless Communications, IEEE Trans on*, 1(4): 660–670, 2002.

[33] Skellam distribution, http://en.wikipedia.org/wiki/Skellam_distribution.

[34] E.W. Weisstein, "Affine transformation", http://mathworld.wolfram.com/AffineTransformation.html.