On the Scalability and Capacity of Wireless Networks with Omnidirectional Antennas

Onur Arpacioglu and Zygmunt J. Haas
School of Electrical and Computer Engineering
Cornell University
Ithaca, NY 14853, USA
o.arpacioglu@cornell.edu, haas@ece.cornell.edu
http://wnl.ece.cornell.edu

ABSTRACT

We consider a wireless network of N nodes equipped with omnidirectional antennas, and we extend the capacity results of some previous works by finding bounds on the maximum achievable per-node end-to-end throughput, λ_e, while using a general network model and a bounded propagation model. Specifically, we show that when the network domain has a fixed area, λ_e is $\Theta(1/N)$ even when the mobility pattern of the nodes, the temporal variation of transmission powers, the source-destination pairs, and the possibly multi-path routes between them are optimally chosen. This result continues to hold even when the nodes are capable of maintaining multiple transmissions and/or receptions simultaneously, or when the communication bandwidth is partitioned into sub-channels of smaller bandwidth. We also address how λ_e depends on the other network parameters such as the area of the network domain, the path loss exponent, or the average number of hops between a source and a destination. Finally, we determine some required conditions to achieve a non-vanishing per-node end-to-end throughput as the number of nodes in the network grows large.

This work has been supported in part by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Office of Naval Research under the grant number N00014-00-1-0564, by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Air Force Office of Scientific Research under the grant number F49620-02-1-0233, and by the National Science Foundation grant number ANI-0081357.

IPSN'04, April 26–27, 2004, Berkeley, California, USA.

Categories and Subject Descriptors
C.2.1. [**Computer-Communication Networks**]: Network Architecture and Design – *Wireless Communication*. G.1.6. [**Numerical Analysis**]: Optimization – *Constrained optimization*. H.1.1. [**Models and Principles**]: Systems and Information Theory – *Information theory*.

General Terms
Performance, Design, Theory.

Keywords
Wireless networks, network capacity, network throughput, network scalability, omnidirectional antennas

1. INTRODUCTION

There are two possible communication paradigms that involve wireless communications – a wireless network supported by a wired infrastructure [9], such as the cellular networks with fixed base-stations, and a peer-to-peer wireless network, such as an ad hoc network. In this work, we mainly focus on the latter, although many of our results are also applicable to the infrastructure-based wireless networks.

The objectives of our work are: (i) to find theoretical results which demonstrate the dependencies among the maximum achievable per-node end-to-end throughput, λ_e, the number of nodes in the network, and the other parameters of the wireless network, and (ii) to determine the implications of these dependencies on the scalability of the wireless network, in other words, to determine the required conditions to support a large number of nodes in the wireless network.

The publication of the capacity results by Gupta and Kumar in [6] has stimulated the scientific community to search for a better understanding of what are the capacity limitations of wireless networks. In that paper, two network models were proposed to analyze the capacity of peer-to-peer wireless networks. In the first network model, called the *arbitrary network* model, there are N immobile nodes in the network, and there are no restrictions on their locations in the network domain, which is assumed to be a disk of area 1 m^2. All nodes are equipped with omnidirectional antennas, and each of the nodes can either transmit or receive, but not both, at a given time. There are no restrictions on the selection of transmission powers, on the source-destination associations, on the routing protocol, and on the spatial-temporal transmission scheduling scheme. In the second model, called the *random network* model, there are three additional restrictions: the node locations are random, the traffic pattern is random, and all transmissions use a fixed transmission power, which is adjusted to ensure that the network is connected as N becomes large.

Moreover, two reception models are proposed in [6]. The first reception model is called the *protocol model*. In this model, a transmission from a particular transmitter to its intended receiver is successful if the receiver is not within interfering ranges of other unintended transmitters. The second model is the *physical model*, which is more directly related to the physical layer design in practical wireless networks. In the physical model, a transmission from a particular transmitter is successful if the Signal-to-Interference-and-Noise Ratio, *SINR*, at the intended receiver of the transmitter stays above a threshold value during the transmission. Finally, if a transmitter is transmitting with power P, then the received power at a distance x from the transmitter is P/x^{γ}, where $\gamma>2$ is the path loss exponent.

The authors in [6] concluded that, with the protocol model, λ_e behaves as $\Theta\left(1/\sqrt{N}\right)$ for arbitrary networks, and $\Theta\left(1/\sqrt{N\log(N)}\right)$ for random networks. On the other hand, with the physical model, λ_e is $O\left(1/N^{1/\gamma}\right)$ and $\Omega\left(1/\sqrt{N}\right)$ for arbitrary networks, whereas, λ_e is $O\left(1/\sqrt{N}\right)$ and $\Omega\left(1/\sqrt{N\log(N)}\right)$ for random networks.

In [5], Grossglauser and Tse investigated whether introducing mobility into the immobile random network model of [6] can increase λ_e. They used the physical model, but to reduce interference, they also allowed wideband communication by incorporating processing gain. To incorporate mobility of nodes, they assumed that the locations of the nodes form a stationary ergodic random process, which has a uniform stationary distribution in the network domain. Furthermore, they assumed that the source-destination pairs never change and that very large end-to-end packet delays are tolerable. They concluded that there is a routing and scheduling scheme that can deliver a packet to its destination with at most two hops and that allows λ_e to be $\Theta(1)$ as N becomes large. Finally, both [6] and [5] concluded that the maximum number of simultaneously successful transmissions in a wireless network is $\Theta(N)$.

In [13], Toumpis and Goldsmith considered a particular placement of the nodes, and they numerically evaluated the contribution of spatial reuse, multi-hop routing, power control, and successive interference cancellation. In [8], Li et al. concluded that only local traffic patterns can be scalable. In [15], Yi, Pei, and Kalyanaraman explored how much improvement in λ_e can be obtained by using directional antennas instead of omnidirectional antennas in the arbitrary and the random network models of [6]. In [11], Peraki and Servetto studied how much improvement in λ_e can be obtained in random networks by both using directional antennas and by allowing the nodes to maintain multiple simultaneous transmissions or receptions. In [9], Liu, Liu and Towsley studied how much improvement in λ_e can be provided by deploying base stations connected with a wired backbone in the random network of [6]. Additionally, information theoretical approaches such as [7] and [14] concluded that considerable gains in network throughput can be obtained through the usage of more advanced receivers that do not regard interference merely as noise.

The motivations behind our work are (i) to relax some of the limitations of [6] and [5] and (ii) to obtain results under a more general network model than the models used in these studies. Specifically, the results of these works are strongly dependent on the propagation model (P/x^{γ}), which becomes invalid as the transmitter-receiver distance x becomes small. Since the network domain in these studies has a fixed size and since the nodes get closer as the number of nodes grows, the results of these works become unreliable beyond some nodal density. In addition, the nodes are immobile in [6], while the mobility pattern in [5] is a very special one; it has to satisfy certain statistical properties for the results to hold. Furthermore, in [6] and [5] each node can maintain either a single transmission or reception at a given time. Finally, in [5], the source-destination pairs never change, and end-to-end delay can be unbounded.

In this paper, we extend the results of [6] and [5] by (i) using a bounded propagation model, (ii) making no restrictions on the mobility pattern of the nodes, (iii) allowing the nodes to maintain multiple simultaneous transmissions and/or receptions, (iv) making no restrictions on source-destination associations or end-to-end delay, and (v) addressing how λ_e depends on the other parameters of the network such as A, γ, G and β (the effects of these parameters have not been addressed in the works above). Above all, we analyze the implications our results on scalability, and we determine some required conditions to achieve a non-vanishing per-node end-to-end throughput as the number of nodes grows large.

The rest of this paper is organized as follows: Section 2 provides the network model. In Section 3, we put forward the definitions of the quantities that are used in the proofs. In Section 4, we present the upper bounds on the maximum number of simultaneously successful transmissions and the maximum achievable per-node end-to-end throughput. Section 5, provides a detailed analysis of the upper bounds. In Section 6, we show that "λ_e is $\Theta(1/N)$". Section 7 discusses the implications of the results on scalability. Finally, Section 8 concludes the paper.

2. NETWORK MODEL

In this section, we describe the properties of the class of wireless networks where our results hold. We keep some of these properties very general, so as to obtain valid results even for networks in which many of the parameters are optimally chosen. We will discuss these parameters in detail after presenting our results.

2.1 Network Domain and Nodes

We define the *network domain* $\boldsymbol{Q}$ as the space where all of the N nodes are located at all times. We assume that the network domain is a closed disk with area A. [1] We make no restrictions on the mobility pattern of the nodes, so as to obtain results that hold for every mobility pattern, including the immobile pattern of [6] and the specific mobility pattern of [5].

2.2 Transceiver Model

All nodes have the ability to act as a transmitter and/or a receiver at any given time. All transmitters and receivers have omnidirectional antennas. We make no restrictions on how the

[1] All distance measures and area measures have the units "m" and "m^2," respectively.

transmission power is varied during a particular transmission. Also, we make no restrictions on the number of simultaneous transmissions and/or receptions that a node is able to maintain at any given time. Thus, the assumption in [6] and [5] that a node is able to maintain either one transmission or one reception at any given time is only one special case that our model is able to account for. For now, we assume that all transmissions share the same communication bandwidth, so that two transmissions interfere with each other when they occur simultaneously. Another possibility would be splitting the communication bandwidth into sub-channels of smaller bandwidth and allowing the scheduling of the transmissions over separate sub-channels, so as to reduce co-channel interference. We will also generalize our results to this situation. Given a transmitted signal from a particular transmitter, all other signals at the intended receiver of this transmitter are regarded as interference to this transmission. At a given time t, $\zeta_i(t)$ denotes the thermal noise power in the communication bandwidth at receiver i at time t. Also, at time t, information from a particular transmitter can be transmitted to its intended receiver at a rate not exceeding $W_{\max}$ bits/s, provided that the *SINR* at the receiver at time t does not fall below a positive threshold β. Information received when this condition is not met is considered unreliable and, therefore, discarded. In general, β depends on the modulation technique, the desired bit error rate for the reliability of the received information, the desired transmission rate, and the coding employed. Also, we assume that a positive constant G denotes the *processing gain*, which represents the factor by which the total received interference power is reduced at each of the receivers. Typically G exceeds 1 in wideband systems, such as spread spectrum CDMA, and it is taken to be 1 for narrowband systems.

2.3 Propagation Model

Suppose a given transmitter transmits with power $P(t)$ at time t. Also, suppose that $x(t)$ is the distance between the transmitter and a given receiver at time t. We assume that the received power $P_r(t)$ at the receiver is given by the following expression:

$$P_r(t) = P(t)a(x(t)) \,, \qquad \text{(a.1)}$$

where $a(x)$ denotes the *attenuation function*. In most of the previous studies, such as [6] and [5], $a(x)$ is taken to be $x^{-\gamma}$, where γ is the path loss exponent.[2] Nevertheless, this function becomes invalid when the transmitter receiver distance x drops below 1, since the resulting received power $P_r(t)$ exceeds the transmitted power $P(t)$ for $x<1$. Furthermore, $P_r(t)$ becomes arbitrarily large as x becomes sufficiently small. Certainly, this is not realistic and is a consequence of the invalidity of the expression for small transmitter-receiver distances.[3] This problem was also noticed in some previous works such as [3] and [4], and to obtain more meaningful results at small distances, while approximating the conventional model at large distances, the following attenuation function was suggested in those studies:

$$a(x) = (1+x)^{-\gamma}, \qquad x \ge 0 \,. \qquad \text{(a.2)}$$

In this work, due to above reasons, we chose to use the propagation model defined by (a.1) and (a.2) instead of the conventional model, and we call the resulting propagation model as the *power law decaying propagation model*.

2.4 Traffic Pattern

We make no restrictions on (i) variation of source-destination associations over time, (ii) the sequence or sequences of intermediate nodes involved in routing the information from the sources, and (iii) information segmentation, which allows transmitting different information segments from a given source over multiple paths from the source to its destinations. As in [6] and [5], we assume that intermediate nodes are not involved in a jointly encoding-decoding scheme for transmitting information from different sources. Finally, $\overline{H}$ denotes the *average number of hops between the source and the destination of a bit* and can have any value larger than or equal to 1, since each bit has to be transmitted over at least one hop.

3. DEFINITIONS

A transmission at an arbitrary time t is called a *successful transmission*, if the *SINR* at the intended receiver of the transmission at time t is greater than or equal to β. N_t denotes the number of simultaneously successful transmissions at time t. *Simultaneous transmission capacity of the network*, $N_t^{\max}$, is defined as the maximum value of N_t over all the possible (i) placements of the N nodes, (ii) selection of the transmitters and their intended receivers, and (iii) selection of transmission powers. A closely related definition is the *simultaneous transmission capacity of the network domain*, N_t^{Ω}, which is defined as the maximum value of N_t over (i), (ii), and (iii) from above, given that there are no restrictions on the number of nodes in the network. So, in the computation of N_t^{Ω}, N is a free variable, whereas in the computation of $N_t^{\max}$, N is fixed, and it represents the actual number of nodes in the network. Therefore, $N_t^{\max} \le N_t^{\Omega}$.

$b_i(T)$ denotes the total amount of information (in bits) generated by node i and received by its destinations during a T second time interval $[0,T]$. The *end-to-end throughput of node i*, λ_i, is then defined as follows:

$$\lambda_i := \lim_{T\to\infty} b_i(T)/T \quad \text{for every } 1 \le i \le N \,.$$

The *per-node average end-to-end throughput*, λ, is defined as the arithmetic average of the end-to-end throughputs of all of the nodes; i.e.,

$$\lambda := \frac{1}{N}\sum_{i=1}^{N} \lambda_i \,.$$

Next, an end-to-end throughput λ_0 is said to be *achievable by all nodes*, if there exist (i) a mobility pattern of the nodes, (ii) a traffic pattern, (iii) a spatial-temporal transmission scheduling policy, and (iv) a temporal variation of transmission powers, such that $\lambda_i \ge \lambda_0$ for all $1 \le i \le N$. Similarly, an end-to-end throughput λ_0 is said to be *achievable on average*, if there exist (i), (ii), (iii) and (iv) from above such that $\lambda \ge \lambda_0$. Observe that if λ_0 is achievable by all nodes, then it is also achievable on average. Since the contrapositive of this statement is also true, i.e., if λ_0 is not achievable on average, then it is also not achievable by all nodes, we say that λ_0 is *not achievable* if λ_0 is not achievable on average.

Next, the *per-node end-to-end throughput capacity*, λ_e, is defined as the supremum of all end-to-end throughputs that are

[2] The parameter γ equals to 2 in free space, but in realistic radio channels, values of γ between 1.6 and 6 have been observed [1],[12]. In this paper, we let γ to have any non-negative value.

[3] See [2] for a more detailed explanation on this issue.

achievable by all nodes. The *per-node average end-to-end throughput capacity*, λ_m, is defined as the supremum of all end-to-end throughputs that are achievable on average. It follows immediately from these definitions that $\lambda_m \geq \lambda_e$.

Finally, we use the standard asymptotic notations; given non-negative functions f and g of a variable x, f is said to be $O(g)$ with respect to x, if there are positive real numbers x_0 and y_0 for which $0 \leq f \leq y_0 g$ for all $x \geq x_0$. Likewise, f is said to be $\Omega(g)$ with respect to x, if there are positive real numbers x_1 and y_1 for which $0 \leq y_1 g \leq f$ for all $x \geq x_1$. Finally, f is said to be $\Theta(g)$ with respect to x, if f is both $O(g)$ and $\Omega(g)$ with respect to x. We will omit the phrase "with respect to x" when it is understandable from the context. We will also make use of the fact that f is $\Theta(g)$ with respect to x if $0 < \lim_{x\to\infty} \frac{f}{g} < \infty$.

4. UPPER BOUNDS ON $N_t^{\max}$, N_t^{Q}, λ_e, AND λ_m

In this section, we derive two theorems. The first theorem provides two upper bounds: an upper bound on N_t^{Q} (hence, $N_t^{\max}$) that is independent of N, and an upper bound on $N_t^{\max}$ that is independent of A and γ. The theorem is as follows:

Theorem 1: *For every time instant t, N_t^{Q} and $N_t^{\max}$ have the following upper bounds:*

$$N_t^{\max} \leq N_t^{Q} \leq U_\gamma , \qquad \text{(T1.1)}$$

$$N_t^{\max} \leq N(1+G/\beta) , \qquad \text{(T1.2)}$$

where

$$U_\gamma := \begin{cases} \dfrac{2(\gamma-1)(\gamma-2)\left(1+\frac{G}{\beta}\right)A}{\pi c_2\left(1+\frac{\gamma-2}{\left(1+2\sqrt{\frac{A}{\pi c_2}}\right)^{\gamma-1}}-\frac{\gamma-1}{\left(1+2\sqrt{\frac{A}{\pi c_2}}\right)^{\gamma-2}}\right)} & \gamma \notin \{1,2\}, \quad \text{(T1.3)} \\ \dfrac{\left(1+\frac{G}{\beta}\right)\sqrt{A}}{\sqrt{\pi c_2}\left(1-\frac{\log\left(1+2\sqrt{\frac{A}{\pi c_2}}\right)}{2\sqrt{\frac{A}{\pi c_2}}}\right)} & \gamma = 1, \quad \text{(T1.4)} \\ \dfrac{2\left(1+\frac{G}{\beta}\right)A}{\pi c_2\left(\log\left(1+2\sqrt{\frac{A}{\pi c_2}}\right)-\frac{2\sqrt{\frac{A}{\pi c_2}}}{1+2\sqrt{\frac{A}{\pi c_2}}}\right)} & \gamma = 2, \quad \text{(T1.5)} \end{cases}$$

$$c_2 := \tfrac{2}{3} - \tfrac{\sqrt{3}}{2\pi} . \qquad \text{(T1.6)}$$

Proof:

Due to space limitations, we only provide a brief outline of the proof to demonstrate the technique that we use. The complete proof can be found in [2].

Firstly, we consider an arbitrary time instant t, and we index each transmitter-receiver pair that belongs to the same transmission with a unique number between 1 and N_t. So, receiver i is the intended receiver of transmitter i for every $1 \leq i \leq N_t$. Let $P_r^{ji}(t)$ be the power received by receiver i from transmitter j at time t. Also, let $SINR_i(t)$ be the $SINR$ at receiver i at time t. Next, we write the $SINR$ expression for each of the N_t transmissions at time t, as follows:

$$SINR_i(t) = \frac{P_r^{ii}(t)}{\zeta_i(t) + \frac{1}{G}\sum_{\substack{j=1,\\(j\neq i)}}^{N_t} P_r^{ji}(t)} , \quad 1 \leq i \leq N_t . \qquad (1)$$

From the definition of a successful transmission, N_t simultaneously successful transmissions can take place at time t, if and only if $SINR_i(t) \geq \beta$ for every $1 \leq i \leq N_t$. By using this condition, together with some other inequalities, such as the triangular inequality, following inequality can be derived as a necessary condition for the existence of N_t simultaneously successful transmissions at time t:

$$\sum_{m=1}^{N_t-1}\sum_{i=1}^{N_t}\frac{1}{\left(1+u_{im}(t)\right)^{\gamma}} \leq \frac{GN_t}{\beta} , \qquad (2)$$

where $u_{im}(t)$ is the Euclidean distance between receiver i and the m^{th} nearest receiver to receiver i at time t. We proceed from here by using the following geometric result derived in [2], which provides an upper bound on the sum of the m^{th} nearest receiver distances when receivers are arbitrarily located in the network domain:

$$\sum_{i=1}^{N_t}\left[u_{im}(t)\right]^2 \leq md^2 , \quad 1 \leq m \leq N_t - 1 , \qquad (3)$$

where $d := 2\sqrt{A/(\pi c_2)}$.

Next, using Kuhn-Tucker Theory [10], frequently used in constrained optimization problems, we find the minimum value of the left hand side of inequality (2) subject to the constraint in (3), and obtain the following inequality as a necessary condition for N_t simultaneously successful transmissions at time t:

$$\sum_{m=1}^{N_t-1}\frac{1}{\left(1+d\sqrt{\frac{m}{N_t}}\right)^{\gamma}} \leq \frac{G}{\beta} . \qquad (4)$$

We proceed from here by using some other inequalities, such as the integral bounds on summations, and we finally obtain the following upper bound on N_t

$$N_t \leq U_\gamma , \qquad (5)$$

where U_γ is defined as in (T1.3), (T1.4), and (T1.5).

During the derivation of (5), we have made no restrictions on the choices of transmitters, on their intended receivers, on the transmission powers, or on the number of nodes in the network. Therefore, the right hand side of (5) is also an upper bound on N_t^{Q}, which is greater than or equal to $N_t^{\max}$. This proves (T1.1).

Finally, (T1.2) follows directly from (2): Suppose we only have a single receiver node which is receiving N_t successful transmissions at time t. Then, in (2), $u_{im}(t)$ is equal to zero for every i, j, and every t. Thus, N_t cannot exceed $1+G/\beta$. This shows that no node can receive more than $1+G/\beta$ successful transmissions simultaneously. Since there are N nodes, this implies that $N_t^{\max}$ cannot exceed $N(1+G/\beta)$, which proves (T1.2).

■

The main reason for the existence of the upper bound in (T1.1) is the co-channel interference. It turns out that, when all transmissions occur at the same channel and thus interfere with each other, it becomes impossible to satisfy the $SINR$ requirements of all of the transmissions at the same time beyond some number of transmissions. On the other hand, the primary reason for the existence of the upper bound in (T1.2) is the limitation on the number of simultaneously successful transmissions that each node can receive; as the above proof demonstrates, none of the nodes can receive more than $1+G/\beta$ simultaneously successful transmissions.

The second theorem follows from Theorem 1, and it provides two upper bounds on λ_e and λ_m. Again, one of these upper bounds is due to co-channel interference, and the other one is due to the limitation on the number of simultaneously successful transmissions that each node can receive. The theorem is as follows:

Theorem 2: *λ_e and λ_m have the following upper bounds:*

$$\lambda_e \le \lambda_m \le \frac{W_{\max} U_\gamma}{\overline{H} N}, \tag{T2.1}$$

$$\lambda_e \le \lambda_m \le \frac{W_{\max}}{\overline{H}} \left(1 + \frac{G}{\beta}\right). \tag{T2.2}$$

Proof:

Form Theorem 1, $N_t^{\max}$ cannot exceed U, where U is equal to either U_γ or $N(1+G/\beta)$. Since the information transmission rate of each of the transmissions cannot exceed $W_{\max}$, the total information transmission rate of the network cannot exceed $W_{\max}U$ at all times. Thus the time average of the total information transmission rate of the network over the time interval $[0,\infty)$, say $\overline{C}$, cannot exceed $W_{\max}U$, either. On the other hand, each bit of information that reaches its destination is transmitted in the network $\overline{H}$ times on average. Thus, $\overline{C}$ is not less than $\overline{H}\Sigma_{i=1}^{N} \lambda_i = \overline{H} N \lambda$. As a result,

$$\overline{H} N \lambda \le \overline{C} \le W_{\max} U, \tag{7}$$

and we find the following upper bound on λ:

$$\lambda \le \frac{W_{\max} U}{\overline{H} N}. \tag{8}$$

In the derivation of (8), we have made no restrictions on the mobility pattern of the nodes, on the traffic pattern, on the spatial-temporal transmission scheduling policy, or on the temporal variation of transmission powers. Since λ_m is defined as $\sup\{\lambda_0 : \lambda \ge \lambda_0\}$ over these parameters, the right hand side of (8) is also an upper bound on λ_m, which is greater than or equal to λ_e. This proves (T2.1) and (T2.2).

■

Until now, we have made no restrictions on the number of transmissions and/or receptions that a node is able to maintain simultaneously. If, as in [6] and [5], there is the extra constraint that a node is able to maintain either one transmission or one reception at any time, a case which we refer to as the *half-duplex restricted* case, the following corollary to Theorem 2 holds:

Corollary 1: *In the half-duplex restricted case, λ_e and λ_m have the following upper bound:*

$$\lambda_e \le \lambda_m \le \frac{W_{\max}}{\overline{H}} \min\left\{\frac{U_\gamma}{N}, \frac{1}{2}\right\}.$$

Proof:

In the half-duplex restricted case, each transmission makes two of the N nodes unavailable for other transmissions. Therefore, $N_t^{\max} \le N/2$. Proceeding as in the proof of Theorem 2, we find that $\lambda_e \le \lambda_m \le W_{\max} / (2\overline{H})$. Combining this with (T2.1) and (T2.2) completes the proof.

■

Until now, we have been assuming that all transmissions occur within the same communication bandwidth. As, we noted in Section 2.2, another possibility is to split the communication bandwidth into smaller sub-channels. This has the potential advantage of reducing co-channel interference, but it also has the disadvantage of reducing the maximum transmission rate, since as the bandwidth is reduced, the maximum transmission rate is reduced as well. The following corollary states that splitting the communication bandwidth into sub-channels of smaller bandwidth does not change the terms other than $W_{\max}$ in (T2.1) and in (T2.2):

Corollary 2: *Suppose each of the transmissions occurs through one of the M non-overlapping sub-channels, whose maximum transmission rates are $W_1^{\max}, W_2^{\max}, \dots W_M^{\max}$. Then, the upper bounds on λ_e and λ_m in Theorem 2 are still valid if W_{max} is replaced with $\Sigma_{m=1}^{M} W_m^{\max}$.*

Proof:

Let $\overline{C}_m$ be the time average of the total information transmission rate of the network over the time interval $[0,\infty)$ in sub-channel m. Proceeding as in the proof of Theorem 2, we find that $\overline{C}_m \le W_m^{\max} U$, where U is defined as in the proof of Theorem 2. Since $\overline{C} = \Sigma_{m=1}^{M} \overline{C}_m$ and $\overline{C}$ is not less than $\overline{H} N \lambda$ from (7), we find that $\overline{H} N \lambda \le \overline{C} \le U \Sigma_{m=1}^{M} W_m^{\max}$.

The remaining part of the proof follows along the same lines as the proof of Theorem 2 after (7).

■

5. ANALYSIS OF THE RESULTS

In this section, we analyze the upper bounds on $N_t^{\mathbf{Q}}$, $N_t^{\max}$, λ_e, and λ_m, provided by Theorems 1 and 2.

Firstly, the analysis of the upper bound U_γ in Theorem 1 leads to the following conclusions on $N_t^{\mathbf{Q}}$:

- $\lim_{\gamma\to\infty} \frac{U_\gamma}{\gamma^2} = \frac{2}{\pi c_2}\left(1+\frac{G}{\beta}\right) A$ $\Rightarrow$ $N_t^{\mathbf{Q}}$ is $O(\gamma^2)$
- $\lim_{A\to\infty} \frac{U_\gamma|_{\gamma=1}}{A^{1/2}} = \frac{1}{\sqrt{\pi c_2}}\left(1+\frac{G}{\beta}\right)$ $\Rightarrow$ $N_t^{\mathbf{Q}}$ is $O(A^{1/2})$ if $\gamma = 1$
- $\lim_{\substack{A\to\infty \\ (\gamma<2)}} \frac{U_\gamma}{A^{\gamma/2}} = \frac{2^\gamma}{(\pi c_2)^{\gamma/2}}\left(1-\frac{\gamma}{2}\right)\left(1+\frac{G}{\beta}\right)$ $\Rightarrow$ $N_t^{\mathbf{Q}}$ is $O(A^{\gamma/2})$ if $\gamma < 2$
- $\lim_{A\to\infty} \frac{U_\gamma|_{\gamma=2}}{A/\log(A)} = \frac{4}{\pi c_2}\left(1+\frac{G}{\beta}\right)$ $\Rightarrow$ $N_t^{\mathbf{Q}}$ is $O(A/\log(A))$ if $\gamma = 2$
- $\lim_{\substack{A\to\infty \\ (\gamma>2)}} \frac{U_\gamma}{A} = \frac{2(\gamma-1)(\gamma-2)}{\pi c_2}\left(1+\frac{G}{\beta}\right)$ $\Rightarrow$ $N_t^{\mathbf{Q}}$ is $O(A)$ if $\gamma > 2$
- $\lim_{G/\beta\to\infty} \frac{U_\gamma}{G/\beta} = f(\gamma,d) := \frac{U_\gamma}{1+G/\beta}$ $\Rightarrow$ $N_t^{\mathbf{Q}}$ is $O(G/\beta)$
- $\lim_{\gamma\downarrow 0} U_\gamma = \lim_{A\downarrow 0} U_\gamma = 1+\frac{G}{\beta}$ $\Rightarrow$ Lack of attenuation and lack of space are equivalent.

These results show that $N_t^{\mathbf{Q}}$ is $O(A^{\min\{\gamma/2,1\}})$ if $\gamma \ne 2$ and $O(A/\log(A))$ if $\gamma = 2$. For any value of γ, these imply that $N_t^{\mathbf{Q}}$ does not grow faster than linearly with the area of the network domain. Also, if $\gamma \le 2$, then linear growth is not possible, and linear growth may only be possible if $\gamma > 2$.

In Figure 1, the common upper bound U_γ on $N_t^{\max}$ and N_t^{Q} is plotted as a function of the area of the network domain and the path loss exponent for $G = \beta = 10$. For a given value of the path

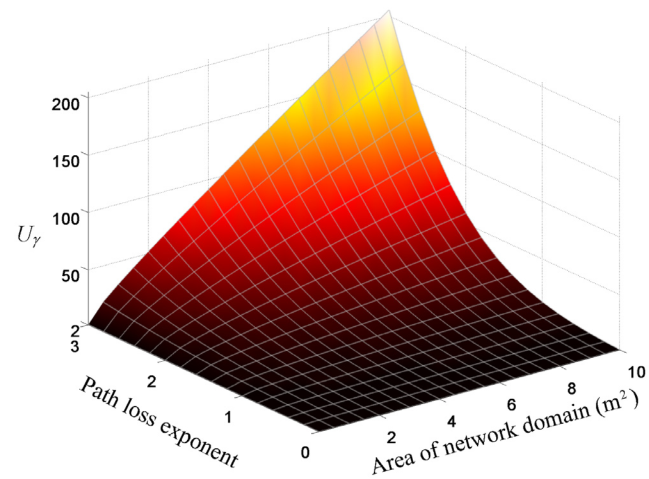

Fig. 1: Upper bound on N_t^Q as a function of the area of the network domain and the path loss exponent

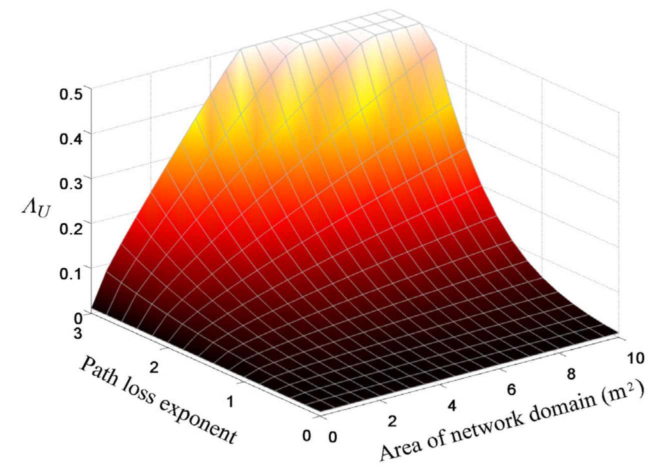

Fig. 2: Upper bound on the normalized λ_e and λ_m as a function of the area of the network domain and the path loss exponent

loss exponent between 0 and 2, the upper bound grows sub-linearly with the area of the network domain, and for values of the path loss exponent exceeding 2, the upper bound grows linearly with the area of the network domain. Figure 1 also illustrates the quadratic growth of the upper bound with the path loss exponent for a given value of the area of the network domain. Additionally, the figure demonstrates the equivalence of lack of attenuation and lack of space; we observe that when either the path loss exponent or the area approaches 0, the upper bound approaches 2 transmissions, which is equal to $1+G/\beta$. Consequently, in this example, at most two simultaneous transmissions are possible, if the network domain lacks attenuation or lacks space.

Next, we draw conclusions on the asymptotic behavior of $N_t^{\max}$. The upper bound in (T1.1) is independent of N, which implies that $N_t^{\max}$ is $O(1)$ with respect to N. Since $N_t^{\max}$ cannot exceed N_t^Q, each of the above $O(\cdot)$ results associated with N_t^Q holds for $N_t^{\max}$, too. However, (T1.2) shows that $N_t^{\max}$ has an upper bound that does not depend on A and γ, which shows that $N_t^{\max}$ is also $O(1)$ with respect to A and γ.

Next, we draw conclusions on the asymptotic behavior of λ_e and λ_m. From (T2.1) we conclude that λ_e and λ_m are $O(1/N)$, $O(1/\overline{H})$ and $O(G/\beta)$.[4] The upper bound in (T1.2) is independent of A and γ, which implies that λ_e and λ_m are $O(1)$ with respect to A and γ.

The above results show that (unlike N_t^Q) $N_t^{\max}$, λ_e, and λ_m are $O(1)$ with respect to A and γ. This discrepancy is due to the upper bounds that result from the limitation on the number of simultaneous receptions that each node can maintain. It turns out that beyond some finite values of A or γ, the limitation on the number of simultaneously successful transmissions that each node can receive does not allow further increases in $N_t^{\max}$, λ_e and λ_m. This is so, in spite of the fact that the network domain provides sufficient space and sufficient attenuation to put more simultaneously successful transmissions inside the network domain. The above phenomenon is easy to observe in the half-duplex restricted case of [6] and [5]; although A or γ keeps growing, there will never be more than $N/2$ simultaneously successful transmissions. Therefore, neither $N_t^{\max}$ exceeds $N/2$ nor λ_e and λ_m exceed $W_{\max}/(2\overline{H})$, regardless of how large A or γ are. On the other hand, for small values of A, γ, and a sufficiently large value of N, the dominant factor that limits $N_t^{\max}$, λ_e, and λ_m is the shortage of space and attenuation, which can be observed from (T1.1) and (T2.1). This behavior is consistent with the claim that for a given N, there exists a region of (A,γ) pairs, where additional space or additional attenuation leads to significant growths in $N_t^{\max}$, λ_e, and λ_m, and beyond this region the contribution of additional space or additional attenuation to $N_t^{\max}$, λ_e, and λ_m vanishes. We intend to further justify this claim in our future work.

As an example of the half-duplex restricted case, in Figure 2 we plot as a function of A and γ the upper bound on λ_e and λ_m normalized to $W_{\max}$, which we will henceforth denote by Λ_U. The remaining parameters are: $G=\beta=10$, $N=220$, and $\overline{H}=1$.[5] As we observe from the figure, similar to U_γ, Λ_U has different growth trends with A for different values of γ. The figure also illustrates the existence of a region of (A,γ) pairs where shortage of space and attenuation limits λ_e and λ_m. Finally, the figure demonstrates that if A or γ becomes sufficiently large, shortage of inactive pairs of nodes limits λ_e and λ_m, and, thus, they cannot exceed $W_{\max}/(2\overline{H})$.

In Figure 3, the network parameters are the same, with the exception that now the path loss exponent has a fixed value of 3 and the number of nodes is an independent variable. This figure illustrates that if the network domain has a fixed size and the number of nodes grows, then λ_e and λ_m decay down to zero, since Λ_U is $\Theta(1/N)$. However, the figure also demonstrates that if we

[4] The conclusion that λ_e and λ_m are $O(G/\beta)$ is based on the assumption that $W_{\max}$ does not depend on G/β. For some practical systems, this may not be true. We will elaborate on this in Section 7.

[5] This is the smallest possible value of $\overline{H}$, and it is achieved when each generated bit is destined for a node one hop away.

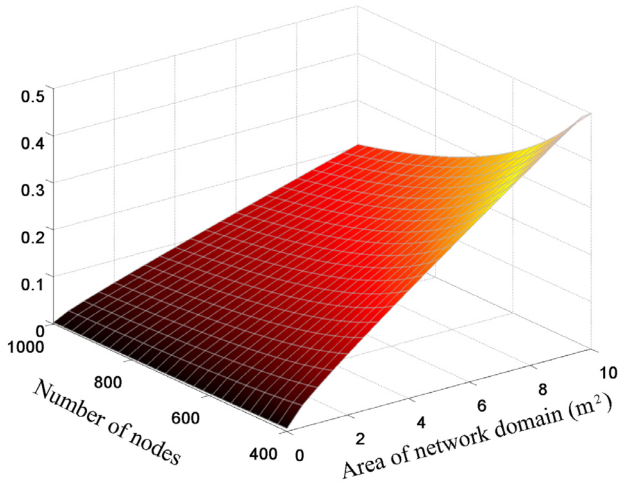

Fig. 3: Upper bound on the normalized λ_e and λ_m as a function of the area of the network domain and the number of nodes

increase the size of the network domain while N is increased, then it is possible to keep the upper bound away from zero, which is required to achieve a non-vanishing per-node end-to-end throughput (i.e., scalability with N). In Section 7, we will find the exact form of the asymptotic relation that A and N must satisfy to keep this upper bound at a constant level.

6. THE $O(1/N)$ RESULT IS TIGHT

In this section, we prove the tightness of the result which states that λ_e and λ_m are $O(1/N)$. To show this we make two additional assumptions: each successful transmission can occur at the rate of W bits/s and the noise power in the communication bandwidth is bounded; i.e., it cannot exceed a constant ζ. With these additional assumptions, we find the following theorem:

Theorem 3: A per-node end-to-end throughput of W/N is achievable by all nodes.

Proof: We divide the time into consecutive slots of equal duration, and consecutive cycles of N slots. Now, we assign each node a unique slot in each cycle and let it transmit at the rate W directly to its destination during the slots assigned to itself with a power that exceeds $\beta\zeta\left(1+2\sqrt{A/\pi}\right)^{\gamma}$. This power level ensures that the transmission is successful. Therefore, each node transmits information successfully to its destination at rate W during $1/N$ fraction of the time. This completes the proof.

■

Theorem 3 shows that λ_e and λ_m are $\Omega(1/N)$. Since they are also $O(1/N)$, they are $\Theta(1/N)$, too. It is worth noting that this throughput scaling is achievable even without employing simultaneous transmissions, as the above proof demonstrates.

7. IMPLICATIONS ON SCALABILITY

In this section, we discuss several required conditions to achieve a non-vanishing per-node end-to-end throughput as the number of nodes in the network tends to infinity. Suppose $\lambda_0 > 0$ is a throughput value that we desire to achieve. Theorem (T1.1) shows that λ_e and λ_m cannot exceed $W_{\max}U_\gamma/(\overline{H}N)$. This implies that unless one or more of the parameters γ, $W_{\max}$, G/β, or A grow with N, λ_0 is not achievable. Note that decreasing $\overline{H}$ is insufficient, since $\overline{H}$ cannot be decreased beyond the value of 1.

In *practical systems*, $W_{\max}$ cannot grow arbitrarily large with N, due to the existence of noise and constraints on the maximum transmission power. The path loss exponent γ depends on the radio channel and it cannot grow arbitrarily large with N. G/β depends on the properties of the communication system and increasing it typically makes it necessary to decrease $W_{\max}$. For instance, in spread spectrum CDMA [12], given the communication bandwidth is fixed, the processing gain and the symbol transmission rate are inversely proportional to each other. Similarly, decreasing the *SINR* threshold β typically requires reducing the symbol transmission rate proportionally to achieve a desired bit error rate.

The only remaining parameter whose growth seems feasible is the area A of the network domain. As we noted at the end of Section 5, it is possible to keep the upper bound on λ_0 at a constant level by increasing the area of the network domain while N is growing. In the beginning of Section 5, we concluded that the term U_γ in the upper bound (T2.1) is $\Theta(A^{\min\{\gamma/2,1\}})$ if $\gamma \neq 2$ and $\Theta(A/\log(A))$ if $\gamma = 2$. So, assuming that A is the only parameter that is growing with N, N must be $O(A^{\min\{\gamma/2,1\}})$ if $\gamma \neq 2$ and $O(A/\log(A))$ if $\gamma = 2$, to keep the upper bounded away from zero. Otherwise, λ_0 is not achievable.

On the other hand, $\overline{H}$ should not grow arbitrarily large with the number of nodes because of the following reason. From (T2.2), it follows that $\lambda_0 \leq W_{\max}(1+G/\beta)/\overline{H}$. As we noted above, increasing G/β typically makes it necessary to decrease $W_{\max}$ proportionally. In such a case, the numerator of the upper bound cannot grow arbitrarily large to compensate for the arbitrarily large $\overline{H}$. This shows that $\overline{H}$ must be $O(1)$ with respect to N. Since we also know that $\overline{H} \geq 1$, we conclude that $\overline{H}$ must be $\Theta(1)$ with respect to N.

Our observations in the previous two paragraphs lead to the corollary regarding practical systems, for which we assume that increasing A is the only way to compensate for increasing N:

Corollary 3: (*Necessary condition for scalability of practical systems*) *A desired per-node end-to-end throughput is not achievable as N tends to infinity, unless* $\overline{H}$ *is* $\Theta(1)$ *with respect to N, A grows with N, and the following condition is satisfied:*

$$N \text{ is } O(A^{\min\{\gamma/2,1\}}) \text{ if } \gamma \neq 2 \text{ and } O(A/\log(A)) \text{ if } \gamma = 2. \qquad (*)$$

■

An equivalent formulation of the above corollary in terms of the *node density*, $\rho := N/A$, is also possible. The only changes are that, in the condition (*), N is replaced with ρ and the expressions inside the $O(\cdot)$ results are divided by A. The resulting condition is that ρ must be $O(A^{\min\{\gamma/2-1,0\}})$ if $\gamma \neq 2$ and $O(1/\log(A))$ if $\gamma = 2$. These expressions show that a desired throughput λ_0 is not achievable if the node density grows indefinitely. Additionally, it also shows that if $\gamma \leq 2$, λ_0 is not achievable unless the node density converges to zero as N becomes large. Finally, these expressions show that only when $\gamma > 2$ it may be possible to achieve λ_0 (by all nodes or on average), while keeping the node density constant.

Figure 4 demonstrates Corollary 3. In Figure 4, $G=\beta=10$ and $\overline{H}=1$. The curves consist of the (A,N) pairs for which $\Lambda_U=0.1$ and $\gamma\in\{0,1,2,3\}$. We know that normalized λ_e and λ_m cannot exceed Λ_U, which decreases with N and increases with A provided that $\Lambda_U\,\overline{H}<0.5$. Therefore, in the regions above each of the curves, the normalized per-node end-to-end throughput of 0.1 is not achievable, whereas it may be achievable (by all nodes or on average) below the curves. We use the phrase "may be" achievable, because the curves are contour plots of an "upper" bound on λ_e and λ_m. This is the case for the throughput of 0.1 if $\gamma=2$ and $(A,N)=(3,100)$, whereas the throughput of 0.1 is not achievable if $\gamma=2$ and $(A,N)=(3,400)$. Corollary 3 gives us the asymptotic relation between the (A,N) pairs on each curve in Figure 4. For example, if γ is equal to 0, 1, 2, or 3, then N is $\Theta(1)$, $\Theta(A^{1/2})$, $\Theta(A/\log(A))$, and $\Theta(A)$, respectively. Equivalently, for the (A,ρ) pairs associated with the curves, if γ is equal to 0, 1, 2, or 3, then ρ is $\Theta(1/A)$, $\Theta(1/A^{1/2})$, $\Theta(1/\log(A))$, and $\Theta(1)$, respectively.

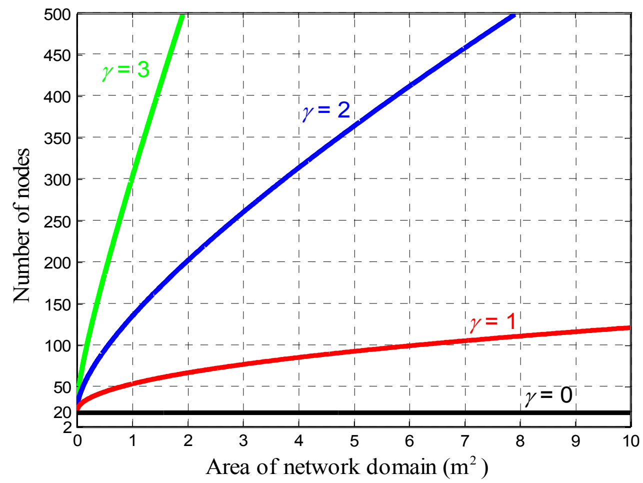

Fig. 4: Curves formed by the (A,N) pairs for which $\Lambda_U=0.1$. For the (A,N) pairs above the curves, any normalized throughput greater than or equal to 0.1 is not achievable

8. CONCLUDING REMARKS

In this paper, we considered a wireless network of N nodes that have omnidirectional antennas, and we extended the capacity results of [6] and [5] while using a more general network model and a bounded propagation model.

Due to the different propagation model, we reached a different conclusion, as compared with the conclusion in [6] and [5] stating that $N_t^{\max}$ is $\Theta(N)$. Rather, we showed that $N_t^{\max}$ has an upper bound that is independent of N. We named this quantity *the simultaneous transmission capacity of the network domain* and denoted it by N_t^{Ω}. Also, we found certain asymptotic rules regarding the growth of N_t^{Ω} with other parameters of the network such as A, γ, or G/β.

Moreover, we generalized the network models of [6] and [5], by (i) making no restrictions on the mobility pattern of the nodes, (ii) allowing the nodes to maintain multiple transmissions and/or receptions at the same time, (iii) making no restrictive assumptions on the source-destination associations or the end-to-end delay. Therefore, the results presented in this paper hold for a broader class of network scenarios. In particular, both immobile and mobile networks with an arbitrary mobility pattern are in this class, as well as networks whose nodes are able to maintain any number of simultaneous transmissions and/or receptions. It followed from this generalized network model that λ_e is $\Theta(1/N)$ even when the mobility pattern of the nodes, the temporal variation of transmission powers, the source-destination pairs, and the possibly multi-path routes between them are optimally chosen. Moreover, this result holds even when the communication bandwidth is divided into sub-channels of smaller bandwidth.

Furthermore, the results that we presented hold for any nonnegative value of γ as opposed to the results of [6] and [5] which hold only for values of γ that are greater than 2. This let us analytically verify the intuitive expectation that lack of space and lack of attenuation are equivalent when the transmitter and the receiver antennas are omnidirectional.

In addition, we found certain rules regarding the variation of $N_t^{\max}$, λ_e, and λ_m with other parameters of the network such as A, γ, or G/β. In particular, we have shown the number of simultaneously successful transmissions that each node can receive cannot be more than $1+G/\beta$, from which it followed that, for a given N, $N_t^{\max}$, λ_e, and λ_m are $O(1)$ with respect to A and γ. On the other hand, we observed that when N is large, the dominant factor that limits $N_t^{\max}$, λ_e, and λ_m is the shortage of space and attenuation.

Finally, we found certain required conditions to achieve a non-vanishing per-node end-to-end throughput as the number of nodes in the network grows large. For practical systems, we concluded that, to achieve a non-vanishing per node end-to-end throughput as the number of nodes grows large, it is essential to keep the average number of hops between a source and a destination bounded, and it is also essential to increase the volume of the network domain at an appropriate rate, which depends on the path loss exponent.

Certainly, a question that remains to be answered is whether or not the upper bound results presented here are tight; in other words, which of the $O(\cdot)$ results are also $\Theta(\cdot)$ results. For fixed area, we have already shown in this paper that λ_e and λ_m are $\Theta(1/N)$. In addition to this, we anticipate that all of the remaining $O(\cdot)$ results are also $\Theta(\cdot)$ results. We are currently in the process of verifying this assertion with the derivation of the lower bounds. Future work will also address extension of the results to network domains that have arbitrary shape and dimension.

9. REFERENCES

[1] Andersen J. B. and Rappaport T. S., Yoshida S. Propagation measurements and models for wireless communications channels. *IEEE Communications Magazine*, 42–49, January 1995

[2] Arpacioglu O., Haas Z. J. On the Scalability and Capacity of Planar Wireless Networks with Omnidirectional Antennas. To appear in Wireless Communications and Mobile Computing (WCMC) Journal, Special Issue on Scalability Issues in Wireless Networks, 2004

[3] Bacceli F. and Blaszczyszyn B. On a coverage range process ranging from the boolean model to the Poisson

Voroni tessellation with applications to wireless communications. *Adv. Appl. Prob.*, 33(2): 293–323, 2001

[4] Dousse O., Baccelli F., Thiran P. Impact of Interferences on Connectivity in Ad Hoc Networks. In *IEEE INFOCOM*, San Francisco, CA, April 2003

[5] Grossglauser M. and Tse D., Mobility increases the capacity of wireless adhoc networks. *IEEE/ACM Trans. Networking*, 10(4): 477–486, August 2002

[6] Gupta P. and Kumar P. R. The capacity of wireless networks. *IEEE Trans. Inform. Theory*, 46(2): 388–404, March 2000

[7] Gupta P. and Kumar P. R. Towards an information theory of large networks: an achievable rate region. *IEEE Trans. Inform. Theory*, 49(8): 1877–1894, August 2003

[8] Li J., Blake C., Couto D. D., Lee H. and Morris R. Capacity of ad hoc wireless networks. In *ACM Mobicom*, Rome, Italy, July 2001

[9] Liu B., Liu Z. and Towsley D. On the capacity of hybrid wireless networks. In *IEEE INFOCOM*, San Francisco, CA, April 2003

[10] Panik M. J. *Classical optimization : foundations and extensions*. American Elsevier Publishing Company, 1976, 227–306

[11] Peraki C. and Servetto S. D. On the maximum stable throughput problem in random networks with directional antennas. In *Proceedings of ACM MobiHoc*, Annapolis, MD, June 2003

[12] Rappaport T. S. *Wireless Communications, Principles and Practice*. NJ: Prentice Hall, 1996, 104, 276–278

[13] Toumpis S. and Goldsmith A. J. Capacity regions for wireless ad hoc networks. *IEEE Trans. Wireless Communications* 2(4): 736–748, July 2003

[14] Xie L.-L. and Kumar P. R. A network information theory for wireless communication: scaling laws and optimal operation. To appear in *IEEE Trans. Inform. Theory*

[15] Yi S., Pei Y. and Kalyanaraman S. On the capacity improvement of ad hoc wireless networks using directional antennas. *In Proceedings of ACM MobiHoc*, Annapolis, MD, June 2003