

Information-Theoretic Operating Regimes of Large Wireless Networks

Ayfer Özgür, Ramesh Johari, *Member, IEEE*, David N. C. Tse, *Fellow, IEEE*, and Olivier Lévêque, *Member, IEEE*

Abstract—In analyzing the point-to-point wireless channel, insights about two qualitatively different operating regimes—bandwidth- and power-limited—have proven indispensable in the design of good communication schemes. In this paper, we propose a new scaling law formulation for wireless networks that allows us to develop a theory that is analogous to the point-to-point case. We identify fundamental operating regimes of wireless networks and derive architectural guidelines for the design of optimal schemes.

Our analysis shows that in a given wireless network with arbitrary size, area, power, bandwidth, etc., there are three parameters of importance: the short-distance signal-to-noise ratio (SNR), the long-distance SNR, and the power path loss exponent of the environment. Depending on these parameters, we identify four qualitatively different regimes. One of these regimes is especially interesting since it is fundamentally a consequence of the heterogeneous nature of links in a network and does not occur in the point-to-point case; the network capacity is both power and bandwidth limited. This regime has thus far remained hidden due to the limitations of the existing formulation. Existing schemes, either multihop transmission or hierarchical cooperation, fail to achieve capacity in this regime; we propose a new hybrid scheme that achieves capacity.

Index Terms—Ad hoc wireless networks, distributed MIMO, hierarchical cooperation, multihopping, operating regimes, scaling laws.

I. INTRODUCTION

THE classic capacity formula $C = W \log_2(1 + P_r/N_0W)$ bits per second of a point-to-point additive white Gaussian noise (AWGN) channel with bandwidth W hertz, received power P_r watts, and white noise with power spectral density $N_0/2$ watts per hertz plays a central role in communication system design. The formula not only quantifies exactly the performance limit of communication in terms of system parameters, but perhaps more importantly also identifies two

fundamentally different operating regimes. In the power-limited (or low-signal-to-noise ratio (SNR)) regime, where $\text{SNR} := P_r/N_0W \ll 0$ dB, the capacity is approximately linear in the power, and the performance depends critically on the power available but not so much on the bandwidth. In the bandwidth-limited (or high-SNR) regime, where $\text{SNR} \gg 0$ dB, the capacity is approximately linear in the bandwidth and the performance depends critically on the bandwidth but not so much on the power. The regime is determined by the interplay between the amount of power and degrees of freedom available. The design of good communication schemes is primarily driven by the parameter regime one is in.

Can analogous operating regimes be identified for *ad hoc* wireless networks, with multiple source and destination pairs and nodes relaying information for each other? To address this question, we are confronted with several problems. First, we have no exact formula for the capacity of networks, even in the simplest case of a single source–destination pair plus one relay. Second, unlike in the point-to-point case, there is no single received SNR parameter in a network. The channels between nodes closer together can be in the high-SNR regime while those between nodes farther away can be in the low-SNR regime.

One approach to get around the first problem is through the scaling law formulation. Pioneered by Gupta and Kumar [1], this approach seeks not the exact capacity of the network but only how it *scales* with the number of nodes in the network and the number of source–destination pairs. The capacity scaling turns out to depend critically on how the area of the network scales with the number of nodes. Two network models have been considered in the literature. In *dense* networks [1], [2], [9], the area is fixed while the density of the nodes increases linearly with the number of nodes. In *extended* networks [3]–[9], the area grows linearly with the number of nodes while the density is fixed. For a given path loss exponent, the area of the network determines the amount of power that can be transferred across the network and so these different scalings couple the power transferred and the number of nodes in different ways.

There are two significant limitations in using the existing scaling law results to identify fundamental operating regimes of *ad hoc* networks. First, the degrees of freedom available in a network depend on the number of nodes in addition to the amount of bandwidth available. By *a priori* coupling the power transferred in the network with the number of nodes in specific ways, the existing formulations may be missing out on much of the interesting parameter space. Second, neither dense nor extended networks allow us to model the common scenario where the channels between different node pairs can be in different SNR regimes. More concretely, let us interpret a channel to be

Manuscript received April 22, 2008; revised March 30, 2009. Current version published December 23, 2009. The work of A. Özgür was supported by Swiss NSF under Grant 200020-118076. The work of R. Johari was supported by the Defense Advanced Research Projects Agency under the ITMANET program, and the U.S. National Science Foundation under Grant 0644114. The work of D. N. C. Tse was supported by the U.S. National Science Foundation via an ITR Grant: “The 3Rs of Spectrum Management: Reuse, Reduce and Recycle.” The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Toronto, ON, Canada, July 2008.

A. Özgür and O. Lévêque are with the Faculté Informatique et Communications, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland (e-mail: ayfer.ozgur@epfl.ch; olivier.leveque@epfl.ch).

R. Johari is with the Department of Management Science and Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: ramesh.johari@stanford.edu).

D. N. C. Tse is with the Department of Electrical Engineering and Computer Science, University of California, Berkeley, Berkeley, CA 94720 USA (e-mail: dtse@eecs.berkeley.edu).

Communicated by L. Zheng, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2009.2034819

in high SNR in a large network if the SNR goes to infinity with n , and in low SNR if the SNR goes to zero with n .¹ Then it can be readily verified that in dense networks, the channels between all node pairs are in the high-SNR regime, while in extended networks, the channels between all node pairs are in the low-SNR regime.

In this paper, we consider a generalization that allows us to overcome these two limitations of the existing formulation. Instead of considering a fixed area or a fixed density, we let the area of the network scale like n^ν where ν can take on any real value. Dense networks correspond to $\nu = 0$ and extended networks correspond to $\nu = 1$. By analyzing the problem for all possible values of ν , we are now considering all possible interplay between power and degrees of freedom. Note that in networks where ν is strictly between 0 and 1, channels between nodes that are far away will be at low SNR while nodes that are closer by will be at high SNR. Indeed, the distance between nearest neighbors is of the order of $\sqrt{A/n} = n^{(\nu-1)/2}$ and, assuming a path loss exponent of α , the received SNR of the transmitted signal from the nearest neighbor scales like $n^{\alpha(1-\nu)/2}$, growing with n . On the other hand, the received SNR of the transmitted signal from the *farthest* nodes scales like $(\sqrt{A})^{-\alpha} = n^{-\alpha\nu/2}$, going to zero with n . Note that scaling the area by n^ν is completely equivalent to scaling the nearest neighbor SNR as n^β , where $\beta := \alpha(1-\nu)/2$. Since SNR is a physically more relevant parameter in designing communication systems, we will formulate the problem as scaling directly the nearest neighbor SNR.

The main result of this paper is as follows. Consider $2n$ nodes randomly located in an area $2A$ such that the received SNR for a transmission over the typical nearest neighbor distance of $\sqrt{A/n}$ is $\text{SNR}_s := n^\beta$. The path loss exponent is $\alpha \geq 2$. Each transmission goes through an independent uniform phase rotation. There are n source and destination pairs, randomly chosen, each demanding the same rate. Let $C_n(\alpha, \beta)$ denote the total capacity of the network, which is the highest achievable sum rate, in bits per second per hertz (bits/s/Hz) and its scaling exponent be defined as

$$e(\alpha, \beta) := \lim_{n \rightarrow \infty} \frac{\log C_n(\alpha, \beta)}{\log n}. \quad (1)$$

The following theorem is the main result of this paper.

Theorem 1: The scaling exponent $e(\alpha, \beta)$ of the total capacity $C_n(\alpha, \beta)$ is given by

$$e(\alpha, \beta) = \begin{cases} 1, & \beta \geq \alpha/2 - 1 \\ 2 - \alpha/2 + \beta, & \beta < \alpha/2 - 1 \text{ and } 2 \leq \alpha \leq 3 \\ 1/2 + \beta, & \beta \leq 0 \text{ and } \alpha > 3 \\ 1/2 + \beta/(\alpha - 2), & 0 < \beta < \alpha/2 - 1 \text{ and } \alpha > 3. \end{cases} \quad (2)$$

Note that dense networks correspond to $\beta = \alpha/2$, with an exponent $e(\alpha, \alpha/2) = 1$ (first case), and extended networks correspond to $\beta = 0$, with an exponent equal to

$$e(\alpha, 0) = \begin{cases} 2 - \alpha/2, & 2 \leq \alpha \leq 3 \\ 1/2, & \alpha > 3 \end{cases}$$

¹We interpret a channel in both high and low SNR, if the SNR does not depend on n .

(second and third cases, respectively). These special cases are the main results of [9]. Observe that in the general case the scaling exponent $e(\alpha, \beta)$ depends on the path loss exponent α and the nearest neighbor SNR exponent β *separately*, so the general result cannot be obtained by a simple rescaling of distances in the dense or extended model.

To interpret the general result (2) and to compare it to the point-to-point scenario, let us re-express the result in terms of system quantities. Recall that SNR_s is the SNR over the smallest scale in the network, which is the typical nearest neighbor distance. Thus

$$\text{SNR}_s = n^\beta = \frac{P_r}{N_0 W} \quad (3)$$

where P_r is the received power from a node at the typical nearest neighbor distance $\sqrt{A/n}$ and W hertz is the channel bandwidth. Let us also define the SNR over the largest scale in the network, the diameter \sqrt{A} , to be

$$\text{SNR}_l := n \frac{n^{-\alpha/2} P_r}{N_0 W} = n^{1-\alpha/2+\beta} \quad (4)$$

where $n^{-\alpha/2} P_r$ is the received power from a node at distance diameter of the network. The result (2) can be used to give the following approximation to the total capacity C , in bits per second:²

$$C \approx \begin{cases} nW, & \text{SNR}_l \gg 0 \text{ dB} \\ n^{2-\alpha/2} P_r / N_0, & \text{SNR}_l \ll 0 \text{ dB and } 2 \leq \alpha \leq 3 \\ \sqrt{n} P_r / N_0, & \text{SNR}_s \ll 0 \text{ dB and } \alpha > 3 \\ \sqrt{n} W^{\frac{\alpha-3}{\alpha-2}} (P_r / N_0)^{\frac{1}{\alpha-2}}, & \text{SNR}_l \ll 0 \text{ dB, SNR}_s \gg 0 \text{ dB,} \\ & \text{and } \alpha > 3. \end{cases} \quad (5)$$

Note two immediate observations in (5). First, there are two SNR parameters of interest in networks, the short and the long distance SNRs, as opposed to the point-to-point case where there is a single SNR parameter. Second, the most natural way to measure the long-distance SNR in networks is not the SNR of a pair separated by a distance equal to the diameter of the network, but it is n times this quantity as defined in (4). Note that there are order n nodes in total located at a diameter distance to any given node in the network, hence n times the SNR between farthest nodes is the total SNR that can be transferred to this node across this large scale. On the other hand, a node has only a constant number of nearest neighbors, and hence the short-distance SNR in (3) is simply the SNR between a nearest neighbor pair.

The four regimes in (5) are shown in Fig. 1. In Regime-I, the performance is achieved by hierarchical cooperation and long range multiple-input multiple-output (MIMO) transmission, the scheme introduced in [9]. At the highest level of hierarchy, clusters of size almost order n communicate via MIMO, at distance the diameter of the network. The quantity SNR_l corresponds to the total received SNR at a node during these MIMO transmissions. Since this quantity is larger than 0 dB, the long-range MIMO transmissions, and hence the performance of the network, are in the bandwidth-limited regime, with performance roughly linear in the bandwidth W . The performance is linear

²Note that $C = W C_n(\alpha, \beta)$.

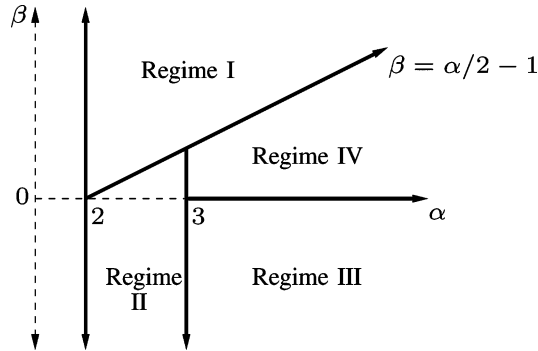


Fig. 1. The four operating regimes. The optimal schemes in these regimes are I-Hierarchical Cooperation, II-Bursty Hierarchical Cooperation, III-Multihop, IV- Multihop MIMO Hierarchical Cooperation.

in the number of nodes, implying that interference limitation is removed by cooperation, at least as far as scaling is concerned. Performance in this regime is qualitatively the same as that in dense networks.

In all the other regimes, the total long-range received SNR is less than 0 dB. Hence, we are power-limited and the transfer of power becomes important in determining performance. In Regime-II, i.e., when $\alpha \leq 3$, signal power decays slowly with distance, and the total power transfer is maximized by long-range MIMO transmission. This performance can be achieved by bursty hierarchical cooperation with long-range MIMO, much like in extended networks.

When $\alpha > 3$, signal power decays fast with distance, and the transfer of power is maximized by short-range communications. If the nearest neighbor SNR $\ll 0$ dB (Regime-III), these transmissions are in the power-efficient regime and this power gain translates linearly into capacity, so nearest neighbor multihop is optimal. This is indeed the case in extended networks, and hence nearest neighbor multihop is optimal for extended networks when $\alpha > 3$.

The most interesting case is the fourth regime, when $\alpha > 3$ and $0 < \beta < \alpha/2 - 1$. This is the case when $\text{SNR}_s \gg 0$ dB, so nearest neighbor transmissions are bandwidth-limited and not power-efficient in translating the power gain into capacity gain. There is the potential of increasing throughput by spatially multiplexing transmission via cooperation within clusters of nodes and performing distributed MIMO. Yet, the clusters cannot be as large as the size of the network since power attenuates rapidly for $\alpha > 3$.

Indeed, it turns out that the optimal scheme in this regime is to cooperate hierarchically within clusters of an intermediate size, perform MIMO transmission between adjacent clusters, and then multihop across several clusters to get to the final destination (see Fig. 2). The optimal cluster size is chosen such that the received SNR in the MIMO transmission is at 0 dB. Any smaller cluster size results in power inefficiency. Any larger cluster size reduces the amount of power transfer because of the attenuation. Note that the two extremes of this architecture are precisely the traditional multihop scheme, where the cluster size is 1 and the number of hops is \sqrt{n} , and the long-range cooperative scheme, where the cluster size is of order n and the number of hops is one. Note also that because short-range links are bandwidth-limited and long-range links are power-limited, the net-

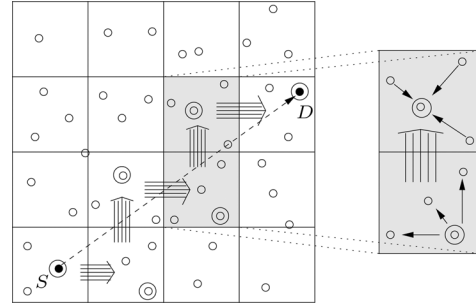


Fig. 2. The figure illustrates the optimal scheme in Regime IV which is based on cooperating locally and multihopping globally. Note that packets are transmitted by multihopping on the network level and each hop is realized with distributed MIMO transmissions combined with hierarchical cooperation.

work capacity is *both* bandwidth- and power-limited. Thus, the capacity is sensitive to both the amount of bandwidth and the amount of power available. This regime is fundamentally a consequence of the heterogeneous nature of links in a network and does not occur in point-to-point links, nor in dense or extended networks.

The organization of the paper is as follows. In the following section, we present our model in more detail. Section III derives a tight upper bound on the scaling exponent in (1). Section IV introduces schemes that achieve this upper bound in different parameter ranges. The two sections together prove our main result in Theorem 1. Section V contains our conclusions.

II. MODEL

There are $2n$ nodes uniformly and independently distributed in a rectangle of area $2\sqrt{A} \times \sqrt{A}$. Half of the nodes are sources and the other half are destinations. The sources and destinations are randomly paired up one-to-one without any consideration on node locations. Each source has the same traffic rate R in bits/s/Hz to send to its destination node and a common average transmit power budget of P watts. The total throughput of the system is $T = nR$.

We assume that communication takes place over a flat channel of bandwidth W hertz around a carrier frequency of f_c , $f_c \gg W$. The complex baseband-equivalent channel gain between node i and node k at time m is given by

$$H_{ik}[m] = \sqrt{G} r_{ik}^{-\alpha/2} \exp(j\theta_{ik}[m]) \quad (6)$$

where r_{ik} is the distance between the nodes, $\theta_{ik}[m]$ is the random phase at time m , uniformly distributed in $[0, 2\pi]$, and $\{\theta_{ik}[m]; 1 \leq i \leq 2n, 1 \leq k \leq 2n\}$ is a collection of independent and identically distributed (i.i.d.) random processes. The $\theta_{ik}[m]$'s and the r_{ik} 's are also assumed to be independent. The parameters G and $\alpha \geq 2$ are assumed to be constants; α is called the power path loss exponent.

The path-loss model is based on the standard far-field assumption: we assume that the distance r_{ik} is much larger than the carrier wavelength λ_c . When the distances are comparable or shorter than the carrier wavelength, the simple path-loss model obviously does not hold anymore as path loss can potentially become path “gain.” Moreover, the phases $\theta_{ik}[m]$ depend on the distance between the nodes modulo the carrier wavelength and

they can only be modeled as completely random and independent of the actual positions of the nodes if the nodes' separation is large enough. Indeed, a recent result [10] showed that, without making an *a priori* assumption of i.i.d. phases, the degrees of freedom are limited by the diameter of the network (normalized by the carrier wavelength λ_c). This is a *spatial* limitation and holds regardless of how many communicating nodes there are in the network. This result suggests that as long as the number of nodes n is smaller than this normalized diameter, the spatial degrees of freedom limitation does not kick in, the degrees of freedom are still limited by the number of nodes, and the i.i.d. phase model is still reasonable. For example, in a network with diameter 1 km and carrier frequency 3 GHz, the number of nodes should be of the order of 10^4 or less for the i.i.d. phase model to be valid. While the present paper deals exclusively with the standard i.i.d. random phase model, it would be interesting to apply our new scaling law formulation to incorporate regimes where there is a spatial degrees of freedom limitation as well.

Note that the channel is random, depending on the location of the users and the phases. The locations are assumed to be fixed over the duration of the communication. The phases are assumed to vary in a stationary ergodic manner (fast fading). We assume that the phases $\{\theta_{ik}[m]\}$ are known in a casual manner at all the nodes in the network. The signal received by node i at time m is given by

$$Y_i[m] = \sum_{k \neq i} H_{ik}[m] X_k[m] + Z_i[m]$$

where $X_k[m]$ is the signal sent by node k at time m and $Z_i[m]$ is white circularly symmetric Gaussian noise of variance N_0 per symbol.

III. CUT-SET UPPER BOUND

We consider a cut dividing the network area into two equal halves. We are interested in upper-bounding the sum of the rates of communications $T_{L \rightarrow R}$ passing through the cut from left to right. These communications with source nodes located on the left and destination nodes located on the right half domain are depicted in bold lines in Fig. 3. Since the source–destination (S-D) pairs in the network are formed uniformly at random, $T_{L \rightarrow R}$ is equal to 1/4th of the total throughput T with high probability (w.h.p.).³ The maximum achievable $T_{L \rightarrow R}$ in bits/s/Hz is bounded above by the capacity of the MIMO channel between all nodes S located to the left of the cut and all nodes D located to the right. Under the fast fading assumption, we have

$$T_{L \rightarrow R} \leq \max_{\substack{Q(H) \geq 0 \\ \mathbb{E}[Q_{kk}(H)] \leq P, \forall k \in S}} \mathbb{E} \left(\log \det \left(I + \frac{1}{N_0 W} H Q(H) H^* \right) \right) \quad (7)$$

where

$$H_{ik} = \frac{\sqrt{G} e^{j\theta_{ik}}}{r_{ik}^{\alpha/2}}, \quad k \in S, i \in D.$$

The mapping $Q(\cdot)$ is from the set of possible channel realizations H to the set of positive semi-definite transmit covariance matrices. The diagonal element $Q_{kk}(H)$ corresponds to

³with high probability, i.e., with probability going to 1 as n grows.

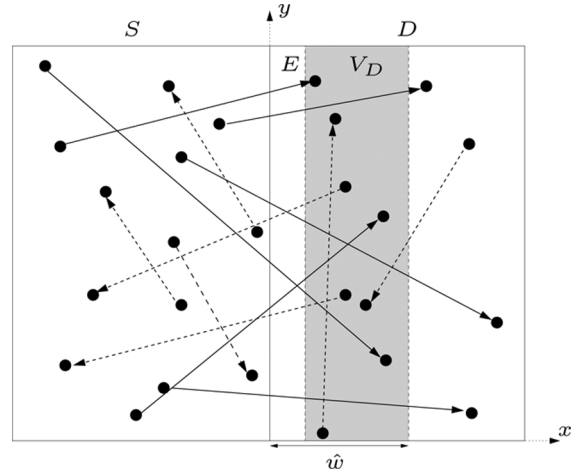


Fig. 3. The cut-set considered in Section III. The communication requests that pass across the cut from left to right are depicted in bold lines.

the power allocated to the k th node for channel state H . Let us simplify notation by introducing

$$\text{SNR}_s := \frac{GP}{N_0 W (A/n)^{\alpha/2}} \quad (8)$$

which can be interpreted as the average SNR between nearest neighbor nodes since $\sqrt{A/n}$ is the typical nearest neighbor distance in the network. Let us also rescale the distances in the network by this nearest neighbor distance, defining

$$\hat{r}_{ik} := \frac{1}{\sqrt{A/n}} r_{ik} \quad \text{and} \quad \hat{H}_{ik} := \frac{e^{j\theta_{ik}}}{\hat{r}_{ik}^{\alpha/2}}. \quad (9)$$

Note that the first transformation rescales space and maps our original network of area $2\sqrt{A} \times \sqrt{A}$ to a network of area $2\sqrt{n} \times \sqrt{n}$, referred to as an extended network in the literature. Consequently, the matrix \hat{H} defined in terms of the rescaled distances relates to such an extended network with area $2n$. We can rewrite (7) in terms of these new variables as

$$T_{L \rightarrow R} \leq \max_{\substack{Q(\hat{H}) \geq 0 \\ \mathbb{E}[Q_{kk}(\hat{H})] \leq 1, \forall k \in S}} \mathbb{E}(\log \det(I + \text{SNR}_s \hat{H} Q(\hat{H}) \hat{H}^*)). \quad (10)$$

In order to upper-bound (10), we will use an approach similar to the one developed in [9, Sec. V-B] for analyzing the capacity scaling of extended networks. Note that although due the rescaling in (9), \hat{H} in (10) governs an extended network, the problem in (10) is not equivalent to the classical extended setup since here we do not necessarily assume $\text{SNR}_s = 1$. Indeed, we want to keep full generality and avoid such arbitrary assumptions on SNR_s in the current paper. Formally, we are interested in characterizing the whole regime $\text{SNR}_s = n^\beta$ where β can be any real number.

One way to upper-bound (10) is through upper-bounding the capacity by the total received SNR, formally using the relation

$$\log \det(I + \text{SNR}_s \hat{H} Q(\hat{H}) \hat{H}^*) \leq \text{Tr}(\text{SNR}_s \hat{H} Q(\hat{H}) \hat{H}^*). \quad (11)$$

The upper bound is tight only if the SNR received by each right-hand side node (each diagonal entry of the matrix $\text{SNR}_s \hat{H}Q(\hat{H})\hat{H}^*$) is small. (Note that the relation in (11) relies on the inequality $\log(1+x) \leq x$ which is only tight if x is small.) In the extended setup, where $\text{SNR}_s = 1$, the network is highly power-limited and the received SNR is small, that is, decays to zero with increasing n , for every right-hand side node. Using (11) yields a tight upper bound in that case. However, in the general case SNR_s can be arbitrarily large which can result in high received SNR for certain right-hand side nodes that are located close to the cut or even for all nodes, depending on how large exactly SNR_s is. Hence, before using (11) we need to distinguish between those right-hand side nodes that receive high SNR and those that have poor power connections to the left-hand side.

For the sake of simplicity in presentation, we assume in this section that there is a rectangular region located immediately to the right of the cut that is cleared of nodes. Formally, we assume that the set of nodes $E = \{i \in D : 0 \leq \hat{x}_i \leq 1\}$ is empty, where \hat{x}_i denotes the horizontal coordinate of the rescaled position $\hat{r}_i = (\hat{x}_i, \hat{y}_i)$ of node i . In fact, w.h.p. this property does not hold in a random realization of the network. However, making this assumption allows us to exhibit the central ideas of the discussion in a simpler manner. The extension of the analysis to the general case (without this particular assumption) is given in Appendix I.

Let V_D denote the set of nodes located on a rectangular strip immediately to the right of the empty region E . Formally, $V_D = \{i \in D : 1 \leq \hat{x}_i \leq \hat{w}\}$ where $1 \leq \hat{w} \leq \sqrt{n}$ and $\hat{w} - 1$ is the rescaled width of the rectangular strip V_D . See Fig. 3. We would like to tune \hat{w} so that V_D contains the right-hand side nodes with high received SNR; i.e., those with received SNR larger than a threshold, say 1. Note, however, that we do not yet know the covariance matrix Q of the transmissions from the left-hand side nodes, which is to be determined from the maximization problem in (10). Thus, we cannot compute the received SNR of a right-hand side node. For the purpose of determining V_D , however, let us arbitrarily look at the case when Q is the identity matrix and define the received SNR of a right-hand side node $i \in D$ when left-hand side nodes are transmitting *independent* signals at full power to be

$$\begin{aligned} \text{SNR}_i &:= \frac{P}{N_0W} \sum_{k \in S} |H_{ik}|^2 \\ &= \text{SNR}_s \sum_{k \in S} |\hat{H}_{ik}|^2 \\ &= \text{SNR}_s \hat{d}_i \end{aligned} \quad (12)$$

where we have defined

$$\hat{d}_i := \sum_{k \in S} |\hat{H}_{ik}|^2. \quad (13)$$

Later, we will see that this arbitrary choice of identity covariance matrix is indeed a reasonable one. A good approximation for \hat{d}_i is

$$\hat{d}_i \approx \hat{x}_i^{2-\alpha} \quad (14)$$

where \hat{x}_i denotes the rescaled horizontal coordinate of node i (see [9, Lemma 5.4]). Using (12) and (14), we can identify three different regimes and specify \hat{w} accordingly.

- 1) If $\text{SNR}_s \geq n^{\alpha/2-1}$, then $\text{SNR}_i \gtrsim 1, \forall i \in D$. Thus, let us choose $\hat{w} = \sqrt{n}$ or equivalently $V_D = D$.
- 2) If $\text{SNR}_s < 1$, then $\text{SNR}_i \lesssim 1, \forall i \in D$. Thus, let us choose $\hat{w} = 1$ or equivalently $V_D = \emptyset$.
- 3) If $1 \leq \text{SNR}_s < n^{\alpha/2-1}$, then let us choose

$$\hat{w} = \begin{cases} \sqrt{n}, & \text{if } \alpha = 2 \\ \text{SNR}_s^{\frac{1}{\alpha-2}}, & \text{if } \alpha > 2 \end{cases}$$

so that we ensure $\text{SNR}_i \gtrsim 1, \forall i \in V_D$.

We now would like to break the information transfer from the left-half domain S to the right-half domain D in (10) into two terms. The first term governs the information transfer from S to V_D . The second term governs the information transfer from S to the remaining nodes on the right-half domain, i.e., $D \setminus V_D$. Recall that the characteristic of the nodes V_D is that they have good power connections to the left-hand side, that is, the information transfer from S to V_D is not limited in terms of power but can be limited in degrees of freedom. Thus, it is reasonable to bound the rate of this first information transfer by the cardinality of the set V_D rather than the total received SNR. On the other hand, the remaining nodes $D \setminus V_D$ have poor power connections to the left-half domain and the information transfer to these nodes is limited in power, hence using (11) is tight. Formally, we proceed by applying the generalized Hadamard's inequality which yields

$$\begin{aligned} &\log \det(I + \text{SNR}_s \hat{H}Q(\hat{H})\hat{H}^*) \\ &\leq \log \det(I + \text{SNR}_s \hat{H}_1Q(\hat{H})\hat{H}_1^*) \\ &\quad + \log \det(I + \text{SNR}_s \hat{H}_2Q(\hat{H})\hat{H}_2^*) \end{aligned}$$

where \hat{H}_1 and \hat{H}_2 are obtained by partitioning the original matrix \hat{H} : \hat{H}_1 is the rectangular matrix with entries $\hat{H}_{ik}, k \in S, i \in V_D$ and \hat{H}_2 is the rectangular matrix with entries $\hat{H}_{ik}, k \in S, i \in D \setminus V_D$. In turn, (10) is bounded above by

$$\begin{aligned} T_{L \rightarrow R} &\leq \max_{\substack{Q(\hat{H}_1) \geq 0 \\ \mathbb{E}(Q_{kk}(\hat{H}_1)) \leq 1, \forall k \in S}} \mathbb{E}(\log \det(I + \text{SNR}_s \hat{H}_1Q(\hat{H}_1)\hat{H}_1^*)) \\ &\quad + \max_{\substack{Q(\hat{H}_2) \geq 0 \\ \mathbb{E}(Q_{kk}(\hat{H}_2)) \leq 1, \forall k \in S}} \mathbb{E}(\log \det(I \\ &\quad + \text{SNR}_s \hat{H}_2Q(\hat{H}_2)\hat{H}_2^*)). \end{aligned} \quad (15)$$

The first term in (15) can be bounded by considering the sum of the capacities of the individual multiple-input single-output (MISO) channels between nodes in S and each node in V_D

$$\begin{aligned} &\max_{\substack{Q(\hat{H}_1) \geq 0 \\ \mathbb{E}(Q_{kk}(\hat{H}_1)) \leq 1, \forall k \in S}} \mathbb{E}(\log \det(I + \text{SNR}_s \hat{H}_1Q(\hat{H}_1)\hat{H}_1^*)) \\ &\leq \sum_{i \in V_D} \log \left(1 + n \text{SNR}_s \sum_{k \in S} |\hat{H}_{ik}|^2 \right) \\ &\leq (\hat{w} - 1) \sqrt{n} \log n \log(1 + n^{1+\alpha(1/2+\delta)} \text{SNR}_s) \end{aligned} \quad (16)$$

w.h.p. for any $\delta > 0$, where we use the fact that for any covariance matrix Q of the transmissions from the left-hand side, the SNR received by each node $i \in V_D$ is smaller than $n \text{SNR}_s \hat{d}_i$ and $\hat{d}_i \leq n^{\alpha(1/2+\delta)}$ since the rescaled minimal separation between any two nodes in the network is larger than $\frac{1}{n^{1/2+\delta}}$ w.h.p.

⁴Note that this is when we use the earlier assumption of an empty strip E of width 1. Without the assumption, we would need to choose $\hat{w} < 1$ in this part.

for any $\delta > 0$. The number of nodes in V_D is upper-bounded by $(\hat{w} - 1)\sqrt{n} \log n$ w.h.p.

The second term in (15) is the capacity of the MIMO channel between nodes in S and nodes in $D \setminus V_D$. Using (11), we get

$$\begin{aligned} & \max_{\substack{Q(\hat{H}_2) \geq 0 \\ \mathbb{E}(Q_{kk}(\hat{H}_2)) \leq 1, \forall k \in S}} \mathbb{E}(\log \det(I + \text{SNR}_s \hat{H}_2 Q(\hat{H}_2) \hat{H}_2^*)) \\ & \leq \max_{\substack{Q(\hat{H}_2) \geq 0 \\ \mathbb{E}(Q_{kk}(\hat{H}_2)) \leq 1, \forall k \in S}} \mathbb{E}(\text{Tr}(\text{SNR}_s \hat{H}_2 Q(\hat{H}_2) \hat{H}_2^*)) \\ & \leq n^\epsilon \text{SNR}_{\text{tot}} \end{aligned} \quad (17)$$

for any $\epsilon > 0$ w.h.p., where

$$\text{SNR}_{\text{tot}} = \sum_{i \in D \setminus V_D} \text{SNR}_i = \text{SNR}_s \sum_{i \in D \setminus V_D} \hat{d}_i. \quad (18)$$

Inequality (17) is proved in [9, Lemma 5.2] and is precisely showing that an identity covariance matrix is good enough for maximizing the power transfer from the left-hand side. Recall that SNR_i in (18) has already been defined in (12) to be the received SNR of node i under *independent* signaling from the left-hand side. Note that (18) is equal to zero when $D \setminus V_D = \emptyset$ or equivalently when $\hat{w} = \sqrt{n}$. If $D \setminus V_D \neq \emptyset$, the last summation in (18) can be approximated with an integral since nodes are uniformly distributed on the network area. Using also (14), it is easy to derive the following approximation for the summation

$$\sum_{i \in D \setminus V_D} \hat{d}_i \approx \int_0^{\sqrt{n}} \int_{\hat{w}}^{\sqrt{n}} \hat{x}^{(2-\alpha)} d\hat{x} d\hat{y}.$$

Here we state a precise result that can be found by straightforward modifications of the analysis in [9]. If $\hat{w} \neq \sqrt{n}$, we have

$$\text{SNR}_{\text{tot}} \leq \begin{cases} K_1 \text{SNR}_s n (\log n)^3, & \alpha = 2 \\ K_1 \text{SNR}_s n^{2-\alpha/2} (\log n)^2, & 2 < \alpha < 3 \\ K_1 \text{SNR}_s \sqrt{n} (\log n)^3, & \alpha = 3 \\ K_1 \text{SNR}_s \hat{w}^{3-\alpha} \sqrt{n} (\log n)^2, & \alpha > 3. \end{cases} \quad (19)$$

where $K_1 > 0$ is a constant independent of SNR_s and n .

Combining the upper bounds (16) and (17) together with our choices for \hat{w} specified earlier, one can get an upper bound on $T_{L \rightarrow R}$ in terms of SNR_s and n . Here, we state the final result in terms of scaling exponents: Let us define

$$\beta := \lim_{n \rightarrow \infty} \frac{\log \text{SNR}_s}{\log n}$$

and

$$e(\alpha, \beta) := \lim_{n \rightarrow \infty} \frac{\log T}{\log n} = \lim_{n \rightarrow \infty} \frac{\log T_{L \rightarrow R}}{\log n}. \quad (20)$$

We have

$$e(\alpha, \beta) \leq \begin{cases} 1, & \beta \geq \alpha/2 - 1 \\ 2 - \alpha/2 + \beta, & \beta < \alpha/2 - 1 \text{ and } 2 \leq \alpha < 3 \\ 1/2 + \beta, & \beta \leq 0 \text{ and } \alpha \geq 3 \\ 1/2 + \beta/(\alpha - 2), & 0 < \beta < \alpha/2 - 1 \text{ and } \alpha \geq 3 \end{cases} \quad (21)$$

where we identify four different operating regimes depending on α and β .

Note that in the first regime the upper bound (16) is active with $\hat{w} = \sqrt{n}$ (or equivalently $V_D = D$) while (17) is zero. The capacity of the network is limited by the degrees of freedom in an $n \times n$ MIMO transmission between the left- and the right-hand side nodes. In the second regime, (17), with the corresponding upper bound being the second line in (19), yields a larger contribution than (16). The capacity is limited by the total received SNR in a MIMO transmission between the left-hand side nodes and $D \setminus V_D$. Note that this total received SNR is equal (in order) to the power transferred in a MIMO transmission between two groups of n nodes separated by a distance of the order of the diameter of the network, i.e., $n^2 \times (\sqrt{n})^{-\alpha} \times \text{SNR}_s$.

In the third regime, (17) is active with $\hat{w} = 1$ (or equivalently $V_D = \emptyset$) while (16) is zero. The corresponding upper bound is the fourth line in (19). Note that this is where we make use of the assumption that there are no nodes located at rescaled distance smaller than 1 to the cut. Due to this assumption, the choice $\hat{w} = 1$ vanishes the upper bound (16) and simultaneously yields $K_1 \text{SNR}_s \sqrt{n} (\log n)^2$ in the last line in (19). If there were nodes closer than rescaled distance 1 to the cut, we would need to choose $\hat{w} < 1$ to vanish the contribution from (16) which would yield a larger value for the term $K_1 \text{SNR}_s \hat{w}^{3-\alpha} \sqrt{n} (\log n)^2$. The capacity in the third regime is still limited by the total SNR received by nodes in $D \setminus V_D$ ($= D$ now) but in this case, the total is dominated by the SNR transferred between the nearest nodes to the cut, i.e., \sqrt{n} pairs separated by the nearest neighbor distance, yielding $\sqrt{n} \times \text{SNR}_s$.

The most interesting regime is the fourth one. Both (16) and (17) with the choice $\hat{w} = \text{SNR}_s^{\frac{1}{\alpha-2}}$ yield the same contribution. Note that (16) upper-bounds the information transfer to V_D , the set of nodes that have bandwidth-limited connections to the left-hand side. This information transfer is limited in degrees of freedom. On the other hand, (17) upper-bounds the information transfer to $D \setminus V_D$, the set of nodes that have power-limited connections to the left-hand side. This second information transfer is power-limited. Eventually in this regime, the network capacity is both limited in degrees of freedom and power, since increasing the bandwidth increases the first term (16) and increasing the power increases the second term (17).

IV. ORDER OPTIMAL COMMUNICATION SCHEMES

In this section, we search for communication schemes whose performance meets the upper bound derived in the previous section. The derivation of the upper bound already provides hints on what these schemes can be: In the first two regimes, the capacity of the network is limited by the degrees of freedom and received SNR, respectively, in a network-wide MIMO transmission. The recently proposed hierarchical cooperation scheme in [9] is based on such MIMO transmissions so it is a natural candidate for optimality in these regimes.

In the third regime, the information transfer between the two halves of the network is limited by the power transferred between the closest nodes to the cut. This observation suggests the following idea: if the objective is to transfer information from the left-half network to the right-half, then it is enough

to employ only those pairs that are located closest to the cut and separated by the nearest neighbor distance. (The rest of the nodes in the network can undertake simultaneous transmissions suggesting the idea of spatial reuse.) In other words, the upper bound derivation suggests that efficient transmissions in this regime are the point-to-point transmissions between nearest neighbors. Indeed, this is how the well-known multihop scheme transfers power across the network so the multihop scheme arises as a natural candidate for optimality in the third regime.

In the derivation of the upper bound for the fourth regime, we have seen that the two terms (16) and (17), governing the information transfer to V_D and $D \setminus V_D$, respectively, yield the same contribution with the particular choice $\hat{w} = \text{SNR}_s^{\frac{1}{\alpha-2}}$. Since the contributions of the two terms are equal (and since we are interested in order here) the derivation of the upper bound suggests the following idea: information can be transferred optimally from the left-half network to the right-half by performing MIMO transmission only between those nodes on both sides of the cut that are located up to $\hat{w} = \text{SNR}_s^{\frac{1}{\alpha-2}}$ rescaled distance to the cut. Note that (16) corresponds to the degrees of freedom in such a MIMO transmission. As in the case of multihop, we can have spatial reuse and allow the rest of the nodes in the network to perform simultaneous transmissions. Thus, the derivation of the upper bound suggests that efficient transmissions in the fourth regime are MIMO transmissions at the scale $\hat{w} = \text{SNR}_s^{\frac{1}{\alpha-2}}$. Combined with the idea of spatial reuse, this understanding suggests to transfer information in the network by performing MIMO transmissions at the particular (local) scale of $\hat{w} = \text{SNR}_s^{\frac{1}{\alpha-2}}$ and then multihopping at the global scale. This new scheme is introduced in Section IV-B.⁵

A. Known Schemes in the Literature

There are two fundamentally different communication schemes suggested for wireless networks in the literature: The multihop scheme and the hierarchical cooperation scheme. The multihop scheme is based on multihopping packets via nearest neighbor transmissions. Its aggregate throughput is well known to be

$$T_{\text{multihop}} = K_1 \sqrt{n} \log \left(1 + \frac{\text{SNR}_s}{1 + K_2 \text{SNR}_s} \right)$$

w.h.p. where $\log(1 + \frac{\text{SNR}_s}{1 + K_2 \text{SNR}_s})$ is the throughput achieved in the nearest neighbor transmissions. $K_2 \text{SNR}_s$ is the interference from simultaneous transmissions in the network-to-noise ratio where $K_2 > 0$ is a constant independent of n and SNR_s . The factor \sqrt{n} is the number of nearest neighbor transmissions that can be parallelized over a given cut. The scaling exponent $e_{\text{multihop}}(\alpha, \beta)$ of the multihop scheme (defined analogously to (20)) is given by

$$e_{\text{multihop}}(\alpha, \beta) = \begin{cases} 1/2, & \beta > 0 \\ 1/2 + \beta, & \beta \leq 0. \end{cases} \quad (22)$$

As can be expected, multihop only achieves the upper bound in (21) in the third regime when $\beta \leq 0$ and $\alpha \geq 3$. In other words, when even the nearest neighbor transmissions in the network are

power limited and signals attenuate sufficiently fast so that pairs located farther apart than the nearest neighbor distance cannot contribute to the power transfer effectively, the optimal strategy is to confine to nearest neighbor transmissions.

The second scheme for wireless networks in [9] is based on a hierarchical cooperation architecture that performs distributed MIMO transmissions between clusters of nodes. The overhead introduced by the cooperation scheme is small so that the throughput achieved by the distributed MIMO transmissions is not that different (at least in scaling sense) from the throughput of a classical MIMO system where transmit and receive antennas are collocated and can cooperate for free. Indeed, the aggregate throughput achieved by the scheme is almost equal to the rate of a MIMO transmission between two clusters of the size of the network n and separated by a distance equal to the diameter of the network \sqrt{A} . More precisely

$$T_{\text{HC}} \geq K_3 n^{1-\epsilon} \log \left(1 + n \frac{GP}{N_0 W (\sqrt{A})^\alpha} \right) \quad (23)$$

for any $\epsilon > 0$ and a constant $K_3 > 0$ w.h.p., where $n^{-\epsilon}$ is the loss in performance due to cooperation overhead. The quantity $n \frac{GP}{N_0 W (\sqrt{A})^\alpha}$ is the total power received by a node in the receive cluster, when nodes in the transmit cluster are signaling independently at full power. Expressing T_{HC} in terms of SNR_s in (8), we have

$$T_{\text{HC}} \geq K_3 n^{1-\epsilon} \log(1 + n^{1-\alpha/2} \text{SNR}_s).$$

Thus, the scaling exponent of hierarchical cooperation is given by

$$e_{\text{HC}}(\alpha, \beta) = \begin{cases} 1, & \beta \geq \alpha/2 - 1 \\ 2 - \alpha/2 + \beta, & \beta < \alpha/2 - 1. \end{cases} \quad (24)$$

The performance in the second line is achieved by using a bursty version of the hierarchical cooperation scheme, where nodes operate the original scheme only a fraction $\frac{1}{n^{\alpha/2-1}}$ of the total time and stay inactive in the rest to save power. See [9, Sec. V-A]. We see that hierarchical cooperation meets the upper bound in (21) in the first regime when $\beta \geq \alpha/2 - 1$, i.e., when power is not a limitation. When power is limited but $2 \leq \alpha \leq 3$, bursty hierarchical cooperation can be used to achieve the optimal power transfer. We see that neither multihop nor hierarchical cooperation is able to meet the upper bound in the fourth regime.

B. A New Hybrid Scheme: Cooperate Locally, Multihop Globally

Let us divide our network of $2n$ nodes and area $2\sqrt{A} \times \sqrt{A}$ into square cells of area $A_c = \frac{2A}{2n} \text{SNR}_s^{1/(\alpha/2-1)}$. Note that $A_c \leq A$, hence this is a valid choice, if $\beta \leq \alpha/2 - 1$. If also $\beta > 0$, each cell contains of the order of $M = \text{SNR}_s^{1/(\alpha/2-1)}$ nodes w.h.p. We transmit the traffic between the S-D pairs in the network by multihopping from one cell to the next. More precisely, let the S-D line associated to an S-D pair be the line connecting its source node to its destination node. Let the packets of this S-D pair be relayed along adjacent cells on its S-D line just like in standard multihop. See Fig. 2. The total traffic through each cell is that due to all S-D lines passing through the cell, which is $O(\sqrt{nM})$. Let us randomly associate each of these $O(\sqrt{nM})$

⁵In a different context, a similar scheme has been suggested recently in an independent work [11].

S-D lines passing through a cell with one of the M nodes in the cell, so that each node is associated with $O(\sqrt{n/M})$ S-D lines. The only rule that we need to respect while doing this association is that if an S-D line starts or ends in a certain cell, then the node associated to the S-D line in this cell should naturally be its respective source or destination node. The nodes associated to an S-D line are those that will decode, temporarily store, and forward the packets of this S-D pair during the multihop operation. The following lemma states a key result regarding the rate of transmission between neighboring cells.

Lemma 1: There exists a strategy (based on hierarchical cooperation) that allows each node in the network to relay its packets to their respective destination nodes in the adjacent cells at a rate

$$R_{\text{relay}} \geq K_4 n^{-\epsilon}$$

for any $\epsilon > 0$ and a constant $K_4 > 0$.

In steady-state operation, the outbound rate of a relay node given in the lemma should be shared between the $O(\sqrt{n/M})$ S-D lines that the relay is responsible for. Hence, the rate per S-D pair is given by

$$R \geq K_4 \sqrt{M} n^{-1/2-\epsilon} \quad (25)$$

or, equivalently, the aggregate rate achieved by the scheme is

$$T_{\text{multihop+HC}} \geq K_4 n^{1/2-\epsilon} \text{SNR}_s^{\frac{1}{\alpha-2}}.$$

In terms of the scaling exponent, we have

$$e_{\text{multihop+HC}}(\alpha, \beta) = 1/2 + \beta/(\alpha - 2), \quad \text{if } 0 < \beta \leq \alpha/2 - 1$$

which matches the upper bound (21) in the third regime.

Note that considering (25), it is beneficial to choose M as large as possible since it reduces the relaying burden. However, Lemma 1 does not hold for any arbitrary M . The proof of the lemma reveals a key property regarding our initial choice for M (or A_c).

Proof of Lemma 1: Let us concentrate only on two neighboring cells in the network. (Consider, for example, the two cells highlighted in Fig. 2): The two neighboring cells together form a network of $2M$ nodes randomly and uniformly distributed on a rectangular area $2\sqrt{A_c} \times \sqrt{A_c}$. Let the M nodes in one of the cells be sources and the M nodes in the other cell be destinations and let these source and destination nodes be paired up randomly to form M S-D pairs. (This traffic will later be used to model the hop between two adjacent cells.) As we have already discussed in (23), using hierarchical cooperation one can achieve an aggregate rate

$$\begin{aligned} MR_{\text{relay}} &\geq K_3 M^{1-\epsilon} \log \left(1 + M \frac{GP}{N_0 W (\sqrt{A_c})^\alpha} \right) \\ &= K_3 M^{1-\epsilon} \log(1 + M^{1-\alpha/2} \text{SNR}_s) \end{aligned}$$

for these M S-D pairs. The second equation is obtained by substituting $A_c = MA/n$ and $\text{SNR}_s = \frac{GP}{N_0 W (A/n)^{\alpha/2}}$. Note that if $M^{1-\alpha/2} \text{SNR}_s \geq 1$, then

$$R_{\text{relay}} \geq K_3 M^{-\epsilon} \geq K_3 n^{-\epsilon}. \quad (26)$$

In other words, $M = \text{SNR}_s^{1/(\alpha/2-1)}$ is the largest cell size one can choose while still maintaining almost constant transmission rate for each of the M S-D pairs.

Now let us turn back to our original problem concerning the steady-state operation of the multihop scheme. At each hop, each of the M nodes in a cell needs to relay its packets to one of the four (left, right, up, and down) adjacent cells. Since the S-D lines are randomly assigned to the nodes in the cell, there are $M/4$ nodes on the average that want to transmit in each direction. These transmissions can be realized successively using hierarchical cooperation and the relaying rate in (26) can be achieved in each transmission. On the other hand, the time-division multiple access (TDMA) between the four transmissions will reduce the overall relaying rate by a factor of 4. Indeed, one should also consider a TDMA scheme between the cells allowing only those cells that are sufficiently separated in space to operate simultaneously so that the inter-cell interference in the network does not degrade the quality of the transmissions significantly. The inter-cell interference and TDMA will further reduce the rate in (26) by a constant factor however will not affect the scaling law. Such insights on scheduling and interference are standard by now and are not central to our analysis on scaling laws. We refer the reader to [9, Lemma 4.2] for more details. \square

Note that the new scheme illustrated in Fig. 2 is a combination of multihop and hierarchical cooperation. Packets are transferred by multihopping on the network level and each hop is realized via distributed MIMO transmissions. Our analysis shows that multihopping and distributed MIMO are two fundamental strategies for wireless networks. However, optimality can only be achieved if these two strategies are combined together appropriately; the optimal combination depends on the SNR level in the network. When $\alpha > 3$, we identify three different regimes in wireless networks: the high, low, and hybrid SNR regimes. The high-SNR regime ($\beta \geq \alpha/2 - 1$) is the extremal case when even the long-distance SNR in the network is large ($\text{SNR}_l \gg 0$ dB). Distributed MIMO with hierarchical cooperation achieves capacity in this case. In the hybrid SNR regime ($0 < \beta \leq \alpha/2 - 1$), the long-distance SNR in the network is low ($\text{SNR}_l \ll 0$ dB), and packets need to be transmitted by multihopping at this scale; while close by pairs are still in the high-SNR regime ($\text{SNR}_s \gg 0$ dB) and distributed MIMO provides the optimal information transfer at this smaller scale. The low-SNR regime ($\beta \leq 0$) is the other extreme when even the short distance SNR is low ($\text{SNR}_s \ll 0$ dB). The multihop MIMO scheme reduces to pure multihop in this last case.

V. CONCLUSION

Suppose you are asked to design a communication scheme for a particular network with given size, area, power budget, path loss exponent, etc. What would be the efficient strategy to operate this wireless network? In this paper, we answer this question by connecting engineering quantities that can be directly measured in the network to the design of good communication schemes. In a given wireless network, we identify two SNR parameters of importance, the short-distance and the long-distance SNRs. The short-distance SNR is the SNR between

nearest neighbor pairs. The long-distance SNR is the SNR between farthest nodes times the size of the network. If the long-distance SNR is high, then the network is in the bandwidth-limited regime. Long-distance communication is feasible and good communication schemes should exploit this feasibility. If the long-distance SNR is low, then the network is power-limited and good communication schemes need to maximize the power transfer across the network. When the power path loss exponent is small so that signals decay slowly, this power transfer is maximized by global cooperation. When the power path loss exponent is large and signals decay fast, the power transfer is maximized by cooperating in smaller scales. The cooperation scale is dictated by the power path loss exponent and the short-distance SNR in the network.

The current results in the literature, in particular [9] that provides the complete picture for the dense and the extended scaling regimes, fail to answer this engineering question because they only address two specific cases that couple the degrees of freedom and power in the network in two very particular ways. The picture is much richer than what can be delineated by these two settings. In that sense, the current paper suggests the abandonment of the existing formulation of wireless networks in terms of dense and extended scaling regimes, a formulation that has been dominant in the literature over the last decade. A better delineation is obtained by treating the power and degrees of freedom available in the network as two independent parameters and studying the interplay between them.

APPENDIX I REMOVING THE ASSUMPTION OF AN EMPTY STRIP IN SECTION III

While proving the upper bound on network capacity in Section III, we have considered a vertical cut of the network that divides the network area into two equal halves and assumed that there is an empty rectangular region to the right of this cut, of width equal to the nearest neighbor distance in the network (or of width equal to 1 in the corresponding rescaled network). With high probability, this assumption does not hold in a random realization of the network. Indeed, for any linear cut of the random network, w.h.p. there will be nodes on both sides of the cut that are located at a distance much smaller than the nearest neighbor distance to the cut. In order to prove the result in Section III rigorously for random networks, we need to consider a cut that is not necessarily linear but satisfies the property of having no nodes located closer than the nearest neighbor distance to it. Below, we show the existence of such a cut using methods from percolation theory. See [13] for a more general discussion of applications of percolation theory to wireless networks.

Lemma 2: For any realization of the random network and a constant $0 < c < 1/7\sqrt{e}$ independent of n and A , w.h.p. there exists a vertical cut of the network area that is not necessarily linear but is located in the middle of the network in a slab not wider than $L = c\sqrt{A/n}\log n$ and is such that there exists no

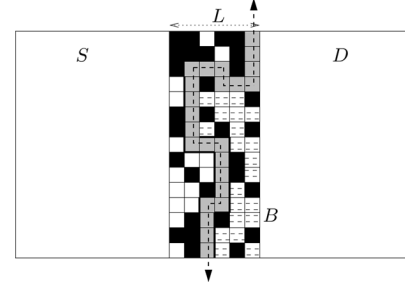


Fig. 4. The cut in Lemma 2. The cut is free of nodes up to distance $c/2$ on both sides.

nodes at distance smaller than $\frac{c}{2}\sqrt{A/n}$ to the cut on both sides. See Fig. 4.

The assumption of an empty region E in Section III, allowed us to plug in $\hat{w} = 1$ in the fourth line of (19) and conclude that when the left-hand side nodes S are transmitting independent signals, the total SNR received by all nodes D to the right of the linear cut is bounded above by

$$\begin{aligned} \text{SNR}_{\text{tot}} &= \sum_{i \in D} \text{SNR}_i \\ &\leq \begin{cases} K_1 \text{SNR}_s n (\log n)^3, & \alpha = 2 \\ K_1 \text{SNR}_s n^{2-\alpha/2} (\log n)^2, & 2 < \alpha < 3 \\ K_1 \text{SNR}_s \sqrt{n} (\log n)^3, & \alpha = 3 \\ K_1 \text{SNR}_s \sqrt{n} (\log n)^2, & \alpha > 3 \end{cases} \quad (27) \end{aligned}$$

where SNR_i is defined in (12).

The same result can be proven for the cut given in Lemma 2 without requiring any special assumption. Let B denote the set of nodes located to the right of the cut but inside the rectangular slab mentioned in the lemma (see Fig. 4). Then

$$\text{SNR}_{\text{tot}} = \sum_{i \in B} \text{SNR}_i + \sum_{i \in D \setminus B} \text{SNR}_i. \quad (28)$$

For any node $i \in B$, an approximate upper bound for SNR_i is

$$\text{SNR}_i \lesssim \text{SNR}_s \int_0^{\sqrt{2\pi}} \int_c^{\sqrt{n}} \frac{1}{\hat{r}^\alpha} \hat{r} d\hat{r} d\theta,$$

since Lemma 2 guarantees that there are no left-hand side nodes located at rescaled distance smaller than c to a right-hand side node i . Moreover, nodes are uniformly distributed on the network area so the summation in (12) over the left-hand side nodes S can be approximated by an integral. A precise upper bound on SNR_i can be found using the binning argument in [9, Lemma 5.2] which yields

$$\text{SNR}_i \leq K_1 \text{SNR}_s \log n.$$

Since there are less than $\sqrt{n} \log n$ nodes in B with high probability, the first summation in (28) can be upper-bounded by

$$\sum_{i \in B} \text{SNR}_i \leq K_1 \text{SNR}_s \sqrt{n} (\log n)^2.$$

Note that this contribution is smaller than any of the terms in (27). The second summation $\sum_{i \in D \setminus B} \text{SNR}_i$ in (28) is equal or smaller in order to (27) since when the nodes B are removed

there is a empty region of width at least c between the nodes S and remaining nodes $D \setminus B$. Hence, for the second term in (28), we are back in the situation discussed in Section III, hence the upper bound (27) applies.

Proof of Lemma 2: Let us divide our network of area $2\sqrt{A} \times \sqrt{A}$ into square cells of side length $c\sqrt{A/n}$ where $0 < c < 1$ is a constant independent of A and n . We say that a cell is closed if it contains at least one node and open if it contains no nodes. Since the $2n$ nodes are uniformly and independently distributed on the network area $2A$, the probability that a given cell is closed is upper-bounded by the union bound by

$$\mathbb{P}[\text{a cell is closed}] \leq c^2.$$

Similarly, the probability that a given set of m cells $\{c_1, \dots, c_m\}$ are simultaneously closed is upper-bounded by

$$\begin{aligned} \mathbb{P}[\{c_1, \dots, c_m\} \text{ is closed}] \\ &= \mathbb{P}[c_1 \text{ is closed}] \times \mathbb{P}[c_2 \text{ is closed} | c_1 \text{ is closed}] \times \dots \\ &\leq c^2 \times c^2 \times \dots \times c^2 = c^{2m} \end{aligned} \quad (29)$$

since by the union bound we have

$$\begin{aligned} \mathbb{P}[c_{k+1} \text{ is closed} | c_1, \dots, c_k \text{ is closed}] \\ &\leq \frac{(c^2 A/n)}{A - k(c^2 A/n)} (n - k) \\ &\leq c^2 \end{aligned}$$

when $0 < c < 1$.

Now let us consider a slab of width $c\sqrt{A/n} \log n$ in the middle of the network. Equivalently, this is a rectangle of $\log n \times \sqrt{n}/c$ cells. By choosing c properly, we will show that this slab contains at least one *open path* that crosses the network from top to bottom. Such a path is called an open top–bottom crossing. A path is called open if it is composed of neighboring cells that are open, a neighboring cell being one of the four cells located immediately to the top, bottom, left and right of a cell (see Fig. 4). On the other hand, we define a closed path in a slightly different manner: A closed path is composed of neighboring cells that are closed but a neighboring cell can now be one of the eight cells located immediately at the top, top-left, left, bottom-left, bottom, bottom-right, right, top-right of a cell. See Fig. 5. With these definitions of closed and open paths, we have

$$\begin{aligned} \mathbb{P}[\text{the slab contains an open top–bottom crossing}] \\ &= 1 - \mathbb{P}[\text{the slab contains a closed left–right crossing}] \end{aligned}$$

where a closed left–right crossing refers to a closed path that connects the left-boundary \mathcal{L} of the slab to its right boundary \mathcal{R} . Let $\mathbb{P}(i \leftrightarrow \mathcal{R})$ denote the probability that there exists a closed path starting from a particular cell $i \in \mathcal{L}$ and ending at the right-boundary. Note that such a path should be at least of length $\log n$ cells. Denoting by N_i the number of closed paths of length $\log n$ that start from the cell i , we have

$$\mathbb{P}(i \leftrightarrow \mathcal{R}) \leq \mathbb{P}(N_i \geq 1).$$

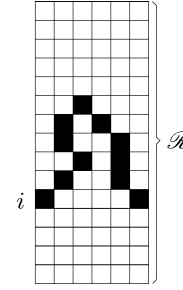


Fig. 5. A closed left–right crossing.

By (29), a given path of length $\log n$ is closed with probability less than $c^{2 \log n}$. By the union bound, we have

$$\mathbb{P}(N_i \geq 1) \leq c^{2 \log n} \sigma_i(\log n)$$

where $\sigma_i(\log n)$ denotes the number of distinct, loop-free paths of length $\log n$ starting from i . This number is obviously not larger than $\sigma_i(\log n) \leq 5 \times 7^{(\log n - 1)}$. Combining the three inequalities, we have

$$\begin{aligned} \mathbb{P}[\text{the slab contains a closed left–right crossing}] \\ &\leq \sum_{i=1}^{\sqrt{n}/c} \mathbb{P}(i \leftrightarrow \mathcal{R}) \\ &\leq \frac{5}{7c} \sqrt{n} (7c^2)^{\log n}. \end{aligned}$$

Choosing $c^2 < \frac{1}{7\sqrt{e}}$, the last probability decreases to 0 as n increases. This concludes the proof of the lemma. \square

REFERENCES

- [1] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 388–404, Mar. 2000.
- [2] S. Aeron and V. Saligrama, "Wireless ad hoc networks: Strategies and scaling laws for the fixed SNR regime," *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 2044–2059, Jun. 2007.
- [3] L.-L. Xie and P. R. Kumar, "A network information theory for wireless communications: Scaling laws and optimal operation," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 748–767, May 2004.
- [4] A. Jovicic, P. Viswanath, and S. R. Kulkarni, "Upper bounds to transport capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2555–2565, Nov. 2004.
- [5] O. Lévêque and E. Telatar, "Information theoretic upper bounds on the capacity of large, extended *ad hoc* wireless networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 858–865, Mar. 2005.
- [6] L.-L. Xie and P. R. Kumar, "On the path-loss attenuation regime for positive cost and linear scaling of transport capacity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2313–2328, Jun. 2006.
- [7] S. Ahmad, A. Jovicic, and P. Viswanath, "Outer bounds to the capacity region of wireless networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2770–2776, Jun. 2006.
- [8] A. Özgür, O. Lévêque, and E. Preissmann, "Scaling laws for one and two-dimensional random wireless networks in the low attenuation regime," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3573–3585, Oct. 2006.
- [9] A. Özgür, O. Lévêque, and D. N. C. Tse, "Hierarchical cooperation achieves optimal capacity scaling in ad-hoc networks," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3549–3572, Oct. 2007.
- [10] M. Franceschetti, M. D. Migliore, and P. Minero, "The Capacity of Wireless Networks: Information-Theoretic and Physical Limits," 2007, preprint.
- [11] U. Niesen, P. Gupta, and D. Shah, "On Capacity Scaling in Arbitrary Wireless Networks," 2007, eprint arXiv: 0711.2745, arxiv.org.

- [12] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [13] M. Franceschetti, O. Dousse, D. N. C. Tse, and P. Thiran, "Closing the gap in the capacity of wireless networks via percolation theory," *IEEE Trans. Inf. Theory*, vol. 53, no. 3, pp. 1009–1018, Mar. 2007.

Ayfer Özgür received the B.Sc. degrees in electrical engineering and physics from Middle East Technical University, Ankara, Turkey, in 2001 and the M.Sc. degree in electrical engineering from the same university in 2004.

She is now working towards the Ph.D. degree at the Laboratory of Information Theory, Swiss Federal Institute of Technology–Lausanne. Her research interests are in wireless communications and information theory.

Ramesh Johari (M'05) received the A.B. degree in mathematics from Harvard University, Cambridge, MA, in 1998, the Certificate of Advanced Study in Mathematics from the University of Cambridge, Cambridge, U.K., in 1999, and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT) in 2004.

He is currently an Assistant Professor of Management Science and Engineering, and by courtesy, Computer Science and Electrical Engineering, at Stanford University, Stanford, CA.

David N. C. Tse (M'96–SM'07–F'09) received the B.A.Sc. degree in systems design engineering from University of Waterloo, Waterloo, ON, Canada, in 1989 and the M.S. and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, in 1991 and 1994 respectively.

From 1994 to 1995, he was a postdoctoral member of technical staff at A.T. & T. Bell Laboratories. Since 1995, he has been at the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, where he is currently a Professor. His research interests are in information theory, wireless communications and networking. He is a coauthor, with P. Viswanath, of the text *Fundamentals of Wireless Communication*.

Prof. Tse received a 1967 NSERC 4-year graduate fellowship from the government of Canada in 1989, an NSF CAREER award in 1998, the Best Paper Awards at the Infocom 1998 and Infocom 2001 conferences, the Erlang Prize in 2000 from the INFORMS Applied Probability Society, the IEEE Communications and Information Theory Society Joint Paper Award in 2001, and the Information Theory Society Paper Award in 2003. He was the Technical Program Co-Chair of the IEEE International Symposium on Information Theory in 2004, and was an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY from 2001 to 2003.

Olivier Lévêque (M'07) was born in Switzerland in 1971. He received the physics diploma from EPFL, Lausanne, Switzerland, in 1995 and completed the Ph.D. degree in mathematics at EPFL in 2001.

Since then, he has been with the Laboratory of Information Theory at EPFL. He spent the academic year 2005–2006 at the Electrical Engineering Department of Stanford University, Stanford, CA, where he was appointed as Lecturer. His research interests include stochastic analysis, random matrices, wireless communications, and information theory.