How does the Information Capacity of Ad Hoc Networks Scale?

Ayfer Özgür, Olivier Lévêque, David Tse

***Abstract*— n source and destination pairs randomly located in an area want to communicate with each other. Signals transmitted from one user to another at distance r apart is subject to a power attenuation of $r^{-\alpha}$ as well as a random phase. We identify exactly the scaling laws of the information theoretic capacity of the network. In the case of dense networks, where the area is fixed and the density of nodes increasing, we show that the total capacity of the network scales *linearly* with n. In the case of extended networks, where the density of nodes is fixed and the area increasing linearly with n, we show that the sum capacity scales as $n^{2-\alpha/2}$ for $\alpha < 3$ and $\sqrt{n}$ for $\alpha \geq 3$. Thus, much better scaling than multihop can be achieved in dense networks, as well as in extended networks with low attenuation. The performance gain is achieved by intelligent node cooperation and distributed MIMO communication. The key ingredient is a *hierarchical* and *digital* architecture for nodal exchange of information for realizing the cooperation.**

I. Introduction

The seminal paper by Gupta and Kumar [1] initiated the study of scaling laws in large ad-hoc wireless networks. Their by-now-familiar model considers n nodes randomly located in the unit disk, each of which wants to communicate to a random destination node at a rate $R(n)$ bits/second. They ask what is the maximally achievable scaling of the total throughput $T(n) = nR(n)$ with the system size n. They showed that classical multihop architectures with conventional single-user decoding and forwarding of packets cannot achieve a scaling of better than $O(\sqrt{n})$, and that a scheme that uses only nearest-neighbor communication can achieve a throughput that scales as $\Theta(\sqrt{n/\log n})$. This gap was later closed by Franceschetti et al [2], who showed using percolation theory that the $\Theta(\sqrt{n})$ scaling is indeed achievable.

A. Özgür is with the School of Computer and Communication Sciences, EPF Lausanne, Switzerland, ayfer.ozgur@epfl.ch

O. Lévêque is with the Electrical Engineering Dept., Stanford University, Stanford, CA 94305-9510, USA, oleveque@stanford.edu

D.Tse is with the Department of EECS, University of California at Berkeley, Berkeley, CA 94720, USA, dtse@eecs.berkeley.edu

Gupta-Kumar model makes certain assumptions on the physical-layer communication technology. In particular, it assumes that the signals received from nodes other than one particular transmitter are interference to be regarded as noise degrading the communication link. Given this assumption, long-range communication between nodes is not preferable, as the interference generated would preclude most of the other nodes from communicating. Instead, the optimal strategy is to confine to nearest neighbor communication and maximize the number of simultaneous transmissions (frequency-reuse). However, this means that each packet has to be retransmitted many times before getting to the final destination, leading to a sub-linear scaling of system throughput. Thus, fundamentally, the Gupta-Kumar result is an *interference-limited* result.

A natural question is whether this result is a consequence of the physical-layer assumptions or whether one can do better using more sophisticated physical-layer processing. In a recent work [3], Aeron and Saligrama have showed that the answer is the latter: they exhibited a scheme which yields a throughput scaling of $\Theta(n^{2/3})$ bits/second. However, it is not clear if one can do even better. In fact, how does the information theoretic capacity of the network scale? The first main result in this paper is that one can in fact achieve arbitrarily close to *linear* scaling: for any $\epsilon > 0$, we present a scheme that achieves an aggregate rate of $\Theta(n^{1-\epsilon})$. This is a surprising result: a linear scaling means the rate for *each* source-destination pair does not degrade significantly even as one puts more and more users in the network. It is easy to show that one cannot get a better capacity scaling than $O(n \log n)$, so up to logarithmic terms, our scheme is optimal.

To achieve linear scaling, one must be able to perform *many* simultaneous long-range communications. A physical-layer technique which achieves this is MIMO (multi-input multi-output): the use of multiple transmit and receive antennas to multiplex several streams of data and transmit them simultaneously. MIMO was originally developed in the point-to-point setting, where the transmit antennas are co-located at a single transmit node, each transmitting one data stream, and the receive antennas are co-located at a single receive node, jointly processing the vector of received observations at the antennas. A natural approach to apply this concept to the network setting is to have both source nodes and destination nodes cooperate in *clusters* to form distributed transmit and receive antenna arrays respectively. In this way, mutually interfering signals can be turned into useful ones that can be jointly decoded at the receive cluster and spatial multiplexing gain can be realized. In fact, if *all* the nodes in the network could cooperate for free, then a classical MIMO result [4], [5] says that a sum rate scaling proportional to n could be achieved. However, this may be over-optimistic : communication between nodes is required to set up the cooperation and this may drastically reduce the useful throughput. The Aeron-Saligrama scheme is MIMO-based and its performance is precisely limited by the cooperation overhead between receive nodes. Our main contribution is a *multi-scale, hierarchical* cooperation architecture without significant overhead. Cooperation first takes place between nodes within very small local clusters to facilitate MIMO communication over a larger spatial scale. This can then be used as a communication infrastructure for cooperation within larger clusters at the next level of the hierarchy. Continuing on this fashion, cooperation can be achieved at an almost global scale.

Since the publication of [1], there have been several works dealing with information theoretic scaling laws of wireless adhoc networks [6], [7], [8], [9], [10]. All of them deal with *extended* networks, which scale to cover an increasing geographical extent with the density of nodes fixed and the source-destination distances increasing large. The best result to date [9] shows that whenever the power path loss exponent α of the environment is greater than 4 so that signal attenuates fast enough, the nearest-neighbor multihop scheme is in fact order-optimal. No better scheme than multihop is known for $\alpha \leq 4$. In contrast, the linear scaling result discussed above is for *dense* networks, where the total area is fixed and the number of nodes is increasing. Extended networks are more complicated to analyze since, in addition to interference, performance is also limited by how much energy can be transferred across long geographical distances. Nevertheless, we show that a simple modification of our hierarchical scheme can be applied to extended networks and achieves a throughput scaling of $n^{2-\alpha/2}$. Thus, for $\alpha < 3$, our scheme performs strictly better than multihop. Moreover, by evaluating a cutset upper bound, we show that our scheme meets the upper bound for $\alpha < 3$, while multihop meets the bound for $\alpha \geq 3$. The scaling law for the extended case is thus completely resolved.

The dense scaling is not only a useful step in studying extended networks, by isolating the issue of interference, but it is of interest on its own right. It is relevant whenever one wants to design networks to serve many nodes, all within communication range of each other (within a campus, an urban block, etc.). This scaling is also a reasonable model to study problems such as *spectrum sharing*, where many users in a geographical area are sharing a wide band of spectrum. Consider the scenario where we segregate the total bandwidth into many orthogonal bands, one for each separate network supporting a *fixed* number of users. As we increase the number of users, the number of such segregated networks increases but the *spectral efficiency*, in bits/s/Hz, does not scale with the *total* number of users. In contrast, if we build one large ad hoc network for all the users on the entire bandwidth, then our result says that the spectral efficiency actually increases *linearly* with the number of users. The gain is coming from a *network* effect via cooperation between the many nodes in the system.

The rest of the paper is summarized as follows. In Section II, we present the model. Section III contains the main result for dense networks and an outline of the proposed architecture together with a back-of-the-envelope analysis of its performance. Section IV characterizes the scaling law for extended networks. Section V contains our conclusions.

II. MODEL

There are n nodes uniformly and independently distributed in a square of unit area (dense scaling). $n/2$ are sources and $n/2$ are destinations. The sources and destinations are paired up one-to-one in an arbitrary way. Each source has the same traffic rate $R(n)$ to send to its destination node and a common average transmit power budget of P Watts. The total throughput of the system is $T(n) = nR(n)$.[1]

We assume that communication takes place over a flat channel of bandwidth W Hz 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:

$$h_{ik}[m] = \sqrt{G} r_{ik}^{-\alpha/2} \exp(j\theta_{ik}[m]) \tag{1}$$

[1]In the sequel, whenever we say a total throughput $T(n)$ is achievable, we implicitly mean that that a rate of $T(n)/n$ is achievable for every source-destination pair.

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 $\{\theta_{ik}[m]\}$ are i.i.d random processes across all i and k. 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 path loss exponent. For example, under free-space line-of-sight propagation, Friis' formula applies and

$$|h_{ik}[m]|^2 = \frac{G_{Tx} \cdot G_{Rx}}{(4\pi r_{ik}/\lambda_c)^2} \qquad (2)$$

so that

$$G = \frac{G_{Tx} \cdot G_{Rx} \cdot \lambda_c^2}{16\pi^2}, \qquad \alpha = 2.$$

where G_{Tx} and G_{Tx} are the transmitter and receiver antenna gains respectively and λ_c is the carrier wavelength.

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).[2] We assume that the channel gains are known at all the nodes. The received signal is a sum of the received signals plus white circular symmetric Gaussian noise of variance N_0 per symbol.

Several comments about the model are in order:

- The path loss model is based on a *far-field* assumption: the distance r_{ik} is assumed to be much larger than the carrier wavelength. When the distance is of the order or shorter than the carrier wavelength, the simple path loss model obviously does not hold anymore as path loss can potentially become path "gain". The reason is that near-field electromagnetics now come into play.
- The phase $\theta_{ik}[m]$ depends on the distance between the nodes modulo the carrier wavelength [11]. The random phase model is thus also based on a far-field assumption: we are assuming that the nodes' separation is at a much larger spatial scale compared to the carrier wavelength, so that the phases can be modelled as completely random and independent of the actual positions.
- It is realistic to assume the variation of the phases since they vary significantly when users move a distance of the order of the carrier wavelength (fractions of a meter). The positions determine the path losses and they on the other hand vary over a much larger spatial scale. So the positions are assumed to be fixed.
- We essentially assume a line-of-sight type environment and ignore multipath effects. The randomness in phases is sufficient for the long range MIMO transmissions needed in our scheme. With multipaths, there is a further randomness due to random constructive and destructive interference of these paths. It can be seen that our result easily extends to the multipath case.

Theoretically, as the number of nodes increases, the far-field assumption eventually becomes invalid as nodes become closer. In reality, the typical separation between nodes is so much larger than the carrier wavelength that the number of nodes when the far-field assumption fails is humongous, i.e. there is a clear separation between the large and the small spatial scales. Consider the following numerical example. Suppose the area of interest is 1 sq. km, well within the communication range of many radio devices. With a carrier frequency of 3 GHz, the carrier wavelength is 0.1m. Even with a very large system size of $n = 10000$ nodes, the typical separation between nearest neighbors is 10 m, very much in the far-field. Under free-space propagation and assuming unit transmit and receive antenna gains, the attenuation given by Friis' formula (2) is about 10^{-6}, much smaller than unity. To have a nearest-neighbor distance of $0.1m$ (the carrier wavelength), 10^8 nodes would be needed in the area! Hence, there is a wide range of system parameters for which simultaneously the number of nodes is large and the far-field assumption holds.

In most of the following discussions, we will simplify the notation by suppressing the dependency of the channel gains on the time index m.

III. Main Result for Dense Networks

We first give an information-theoretic upper bound on the achievable scaling law for the aggregate throughput in the network. Before starting to look for good communication strategies, Theorem 3.1 establishes the best we can hope for.

Theorem 3.1: The aggregate throughput in the network with n nodes is bounded above by

$$T(n) \leq K' n \log n$$

[2]With more technical efforts, we believe our results can be extended to the slow fading setting where the phases are fixed as well.

with high probability[3] for some constant $K' > 0$ and independent of n.

Proof: Consider a source-destination pair (s, d) in the network. The transmission rate $R(n)$ from source node s to destination node d is upper bounded by the capacity of the single-input multiple-output (SIMO) channel between source node s and the rest of the network. Using a standard formula for this channel (see eg. [11]), we get:

$$\begin{aligned} R(n) &\leq \log\left(1 + \frac{P}{N_0}\sum_{\substack{i=1\\ i\neq s}}^{n}|h_{is}|^2\right) \\ &= \log\left(1 + \frac{P}{N_0}\sum_{\substack{i=1\\ i\neq s}}^{n}\frac{G}{r_{is}^{\alpha}}\right). \end{aligned}$$

It is a well known fact that in a random network with n nodes uniformly distributed on a fixed two-dimensional area, the minimum distance between any two nodes in the network is larger than $\frac{1}{n^{1+\delta}}$ with high probability, for any $\delta > 0$. Using this fact, we obtain

$$R(n) \leq \log\left(1 + \frac{GP}{N_0}n^{\alpha(1+\delta)+1}\right) \leq K'\log n$$

for some constant $K' > 0$ and independent of n for all-source destination pairs in the network with high probability. The theorem follows. □

In the view of what is ultimately possible, established by Theorem 3.1, we are now ready to state the main result of this paper.

Theorem 3.2: Let $\alpha \geq 2$. For any $\epsilon > 0$, with high probability an aggregate throughput

$$T(n) \geq Kn^{1-\epsilon}$$

is achievable in the network for all possible pairings between sources and destinations. $K > 0$ is a constant independent of n and the source-destination pairing.

Theorem 3.2 states that it is actually possible to perform arbitrarily close to the bound given in Theorem 3.1. The two theorems together establish the capacity scaling for the network up to logarithmic terms. Note how dramatically different is this new linear capacity scaling law from the well-known throughput scaling of $\Theta(\sqrt{n})$ implied by [1], [2] for the same model. Note also that the upper bound in Theorem 3.1 assumes a genie-aided removal of interference between simultaneous transmissions from different sources. By proving Theorem 3.2, we will show that it is possible to mitigate such interference without a genie but with cooperation between the nodes.

The proof of Theorem 3.2 relies on the construction of an explicit scheme that realizes the promised scaling law. The construction is based on recursively using the following key lemma, which addresses the case when $\alpha > 2$.

Lemma 3.1: Consider $\alpha > 2$ and a network with n nodes subject to interference from external sources. Let the interference signals received by different nodes in the network be uncorrelated and the interference power received by each node be upper bounded by

$$P_I \leq K_I$$

for some $K_I > 0$ and independent of n. Let us assume there exists a scheme such that for each n, with probability at least $1 - e^{-n^{c_1}}$ it achieves an aggregate throughput

$$T(n) \geq K_1 n^b$$

for every possible source-destination pairing in a network of n nodes. K_1 and c_1 are positive constants independent of n and the source-destination pairing, and $0 \leq b < 1$. Let us also assume that the per node average power budget required to realize this scheme is:

$$P \leq \frac{K_p}{n} \tag{3}$$

for some $K_p > 0$ and independent of n.

Then one can construct another scheme that achieves a *higher* aggregate throughput

$$T(n) \geq K_2 n^{\frac{1}{2-b}}$$

for every source-destination pairing in a network of n nodes under the same interference conditions, where $K_2 > 0$ is another constant independent of n and the pairing. Moreover, the failure rate for the new scheme is upper bounded by $e^{-n^{c_2}}$ for another positive constant c_2 while the per node average power needed to realize the scheme is also bounded above by (3).

Lemma 3.1 is the key step to build a hierarchical architecture. Since $\frac{1}{2-b} > b$ for $0 \leq b < 1$, the new scheme is always better than the old. We will now give a rough description of how the new scheme can be constructed given the old scheme, as well as a back-of-the-envelope analysis of the scaling law it achieves.

[3] i.e. probability going to 1 as system size grows.

The constructed scheme is based on clustering and long-range MIMO transmissions between clusters. We divide the network into clusters of M nodes. Let us focus for now on a particular source node s and its destination node d. s will send M bits to d in 3 steps:

- 1) Node s will distribute its M bits among the M nodes in its cluster, one for each node;
- 2) These nodes together can then form a distributed transmit antenna array, sending the M bits *simultaneously* to the destination cluster where d lies;
- 3) Each node in the destination cluster obtained one observation from the MIMO transmission, and it quantizes and ships the observation back to d, which can then do joint MIMO processing of all the observations and decode the M transmitted bits.

From the network point of view, all source-destination pairs have to eventually accomplish these three steps. Step 2 is long-range communication and only one source-destination pair can operate at the same time. Steps 1 and 3 involve local communication and can be parallelized across source-destination pairs. Combining all this leads to three phases in the operation of the network:

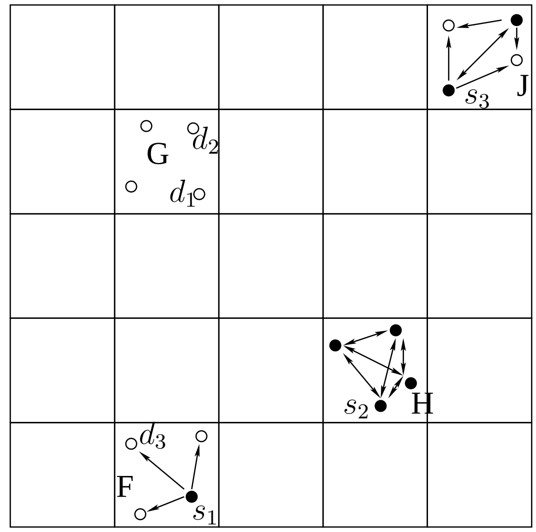

Fig. 1. Source nodes inside clusters F, G, H and J are illustrated while distributing bits in Phase 1. Note that the clusters work in parallel. In this and the following figures Fig. 2 and Fig. 3, we highlight three source-destination pairs $s_1 - d_1$, $s_2 - d_2$ and $s_3 - d_3$, such that nodes s_1 and d_3 are located in F, nodes s_2 and s_3 are located in H and J respectively, and nodes d_1 and d_2 are located in G. Source nodes in the network are depicted in black and destination nodes are depicted in white.

Phase 1: Setting Up Transmit Cooperation Clusters work in parallel. Within a cluster, each source node has to distribute M bits to the other nodes, 1 bit for each node, such that at the end of the phase each node has 1 bit from each of the source nodes in the same cluster. Since there can be at most M source nodes in each cluster, this gives a traffic demand of exchanging at most M^2 bits. The key observation is that this is similar to the original problem of communicating between n source and destination pairs, but on a network of size M. More specifically, this traffic demand of exchanging M^2 bits is handled by setting up M sub-phases, and assigning M source-destination pairs for each sub-phase. Since our channel model is scale invariant, note that the scheme given in the hypothesis of the lemma can be used in each sub-phase by simply scaling down the power with cluster area. Having aggregate throughput M^b, each sub-phase is completed in M^{1-b} time slots while the whole phase takes M^{2-b} time slots. See Figure 1.

Phase 2: MIMO Transmissions We perform successive long-distance MIMO transmissions between source-destination pairs, one at a time. In each one of the MIMO transmissions , say one between s and d, the M bits of s are simultaneously transmitted by the M nodes in its cluster to the M nodes in the cluster of d. Each of the long-distance MIMO transmissions are repeated for each source node in the network, hence we need n time slots to complete the phase. See Figure 2.

Phase 3: Cooperate to Decode Clusters work in parallel. Since there are at most M destination nodes inside the clusters, each cluster received at most M MIMO transmissions in phase 2, one intended for each of the destination nodes in the cluster. Thus, each node in the cluster has at most M received observations, one from each of the MIMO transmissions, and each observation is to be conveyed to a different destination node in its cluster. Nodes quantize each observation into fixed Q bits so there are now a total of at most QM^2 bits to exchange inside each cluster. Using exactly the same scheme as in Phase 1, we conclude the phase in QM^{2-b} time slots. See Figure 3.

Assuming that each destination node is able to decode the transmitted bits from its source node from the M quantized signals it gathers by the end of Phase 3, we can calculate the rate of the scheme as follows: Each source node is able to transmit M bits to its destination node, hence nM bits in total are delivered to their destinations in $M^{2-b} + n + QM^{2-b}$ time slots,

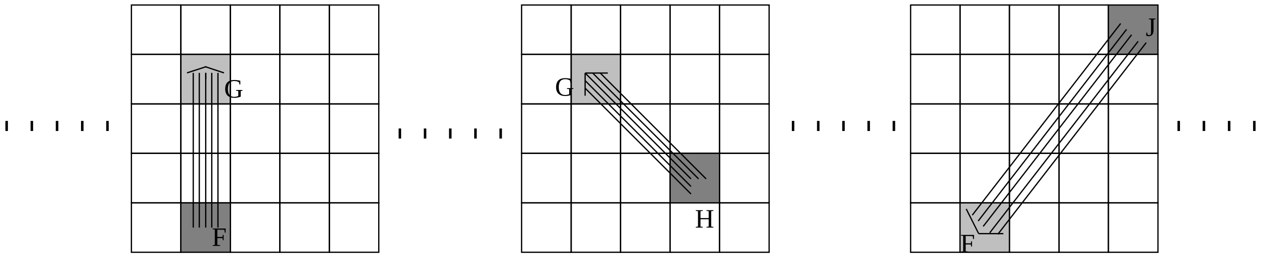

Fig. 2. Successive MIMO transmissions are performed between clusters. The first figure depicts MIMO transmission from cluster F to G, where bits originally belonging to s_1 are simultaneously transmitted by all nodes in F to all nodes in G. The second MIMO transmission is from H to G, while now bits of source node s_2 are transmitted from nodes in H to nodes in G. The third picture illustrates MIMO transmission from cluster J to F.

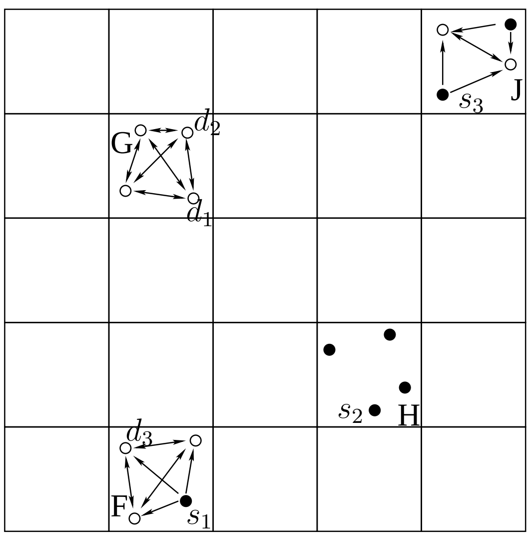

Fig. 3. Observations received by nodes in F, G, H and J during Phase 2, are conveyed to their destination nodes in parallel. Note that the picture is completely symmetric to Fig. 1 except that the characteristic of the traffic is not "from source nodes" but is "to destination nodes" in this case.

yielding an aggregate throughput of

$$\frac{nM}{M^{2-b} + n + QM^{2-b}}$$

bits per time slot. Maximizing this throughput by choosing $M = n^{\frac{1}{2-b}}$ yields $T(n) = \frac{1}{2+Q} n^{\frac{1}{2-b}}$ for the aggregate throughput which is the result in Lemma 3.1.

Clusters can work in parallel in phases 1 and 3 because for $\alpha > 2$, the aggregate interference at a particular cluster caused by other active nodes is bounded. For $\alpha = 2$, the aggregate interference scales like $\log n$, leading to a slightly different version of the lemma.

Lemma 3.2: Consider $\alpha = 2$ and a network with n nodes subject to interference from external sources. Let the interference signals received by different nodes in the network be uncorrelated and the interference power received by each node be upper bounded by

$$P_I \leq K_I \log n$$

for some $K_I \geq 0$ and independent of n. Let us assume there exists a scheme such that for each n with failure probability at most $e^{-n^{c_1}}$, achieves an aggregate throughput

$$T(n) \geq K_1 \frac{n^b}{\log n}$$

for every source-destination pairing in a network with n nodes. K_1 and c_1 are positive constants independent of n and the source-destination pairing, and $0 \leq b < 1$. Let us also assume that the per node average power budget required to realize this scheme is:

$$P \leq \frac{K_p}{n} \tag{4}$$

for some $K_p > 0$ and independent of n.

Then one can construct another scheme that achieves a *higher* aggregate throughput

$$T(n) \geq K_2 \frac{n^{\frac{1}{2-b}}}{\log n}$$

for every source-destination pairing in a network of n nodes under the same interference conditions, where $K_2 > 0$ is another constant independent of n and the pairing. Moreover, the failure rate for the new scheme is upper bounded by $e^{-n^{c_2}}$ for another positive constant c_2 while the per node average power needed to realize the scheme is also bounded above by (4).

We can now use Lemma 3.1 and 3.2 to prove Theorem 3.2.

Proof of Theorem 3.2: We only focus on the case of $\alpha > 2$. The case of $\alpha = 2$ proceeds similarly, differing only with a reduction of a factor of $\log n$ in the throughputs.

We start by observing that the simple scheme of transmitting directly between the source-destination pairs one at a time (TDMA) satisfies the requirements of the lemma. The throughput is $\Theta(1)$, so $b = 0$. The failure probability is 0. Since each source is only transmitting $\frac{1}{n}$th of the time and the distance between the source and its destination is bounded, the average power consumed per node is of the order of $\frac{1}{n}$.

As soon as we have a scheme to start with, Lemma 3.1 can be applied recursively, yielding a scheme that achieves higher throughput at each step of the recursion. More precisely, starting with a TDMA scheme with $b = 0$ and applying Lemma 3.1 recursively h times, one gets a scheme achieving $O(n^{\frac{h}{h+1}})$ aggregate throughput. Given any $\epsilon > 0$, we can now choose h such that $\frac{h}{h+1} \geq 1 - \epsilon$ and we get a scheme that achieves $O(n^{1-\epsilon})$ aggregate throughput scaling with high probability. This concludes the proof of Theorem 3.2. □

Gathering everything together, we have built a hierarchical scheme to achieve the desired throughput. At the lowest level of the hierarchy, we use the simple TDMA scheme to exchange bits for cooperation among small clusters. Combining this with longer range MIMO transmissions, we get a higher throughput scheme for cooperation among nodes in larger clusters at the next level of the hierarchy. Finally, at the top level of the hierarchy, the cooperation clusters are almost the size of the network and the MIMO transmissions are over the global scale to meet the desired traffic demands. Figure 4 shows the resulting hierarchical scheme with a focus on the top two levels.

IV. Extended Networks

A. Bursty Hierarchical Scheme does better than Multihop for $\alpha < 3$

So far, we have considered *dense* networks, where the total geographical area is fixed and the density of nodes increasing. Another natural scaling is the *extended* case, where the density of nodes is fixed and the area is increasing, a $\sqrt{n} \times \sqrt{n}$ square. This models the situation where we want to scale the network to cover an increasing geographical area.

As compared to dense networks, the distance between nodes is increased by a factor of $\sqrt{n}$, and hence for the same transmit powers, the received powers are all decreased by a factor of $n^{\alpha/2}$. Equivalently, by rescaling space, an extended network can just be considered as a dense network on a unit area but with the average power constraint per node reduced to $P/n^{\alpha/2}$ instead of P.

Lemmas 3.1 and 3.2 state that the average power per node required to run our hierarchical scheme in dense networks is not the full power P but P/n. In light of the observation above, this immediately implies that when $\alpha = 2$, we can directly apply our scheme to extended networks and achieve a *linear* scaling. For extended networks with $\alpha > 2$, our scheme would not satisfy the equivalent power constraint $P/n^{\alpha/2}$ and we are now in the power-limited regime (as opposed to the degrees-of-freedom limited regime). However, we can consider a simple "bursty" modification of the hierarchical scheme which runs the hierarchical scheme a fraction

$$\frac{1}{n^{\alpha/2-1}}$$

of the time with power P/n per node and remains silent for the rest of the time.[4] This meets the given average power constraint of $P/n^{\alpha/2}$, and achieves an aggregate throughput of

$$\frac{1}{n^{\alpha/2-1}} \cdot n^{1-\epsilon} = n^{2-\alpha/2-\epsilon} \qquad \text{bits/second.}$$

Note that the quantity $n^{2-\alpha/2} = n^2 \cdot n^{-\alpha/2}$ can be interpreted as the total power transferred between a size n transmit cluster and a size n receive cluster, n^2 node pairs in all, with a power attenuation of $n^{-\alpha/2}$ for each node pairs. This power transfer is taking place at the top level of the hierarchy. The fact that the achievable rate is proportional to the power transfer further emphasizes that our scheme is power-limited rather than degrees-of-freedom limited in extended networks.

Let us compare our scheme to multihop. For $\alpha < 3$, it performs strictly better than multihop, while for $\alpha > 3$, it performs worse. Summarizing these observations, we have the following achievability theorem for extended networks, the counterpart to Theorem 3.2 for dense networks.

Theorem 4.1: Consider an extended network on a $\sqrt{n} \times \sqrt{n}$ square. There are two cases.

- $2 \leq \alpha < 3$: For every $\epsilon > 0$, with high probability, an aggregate throughput:

$$T(n) \geq K n^{2-\alpha/2-\epsilon}$$

[4]We talk in terms of time but such burstiness can just as well be implemented over frequency with only a fraction of the total bandwidth W used.

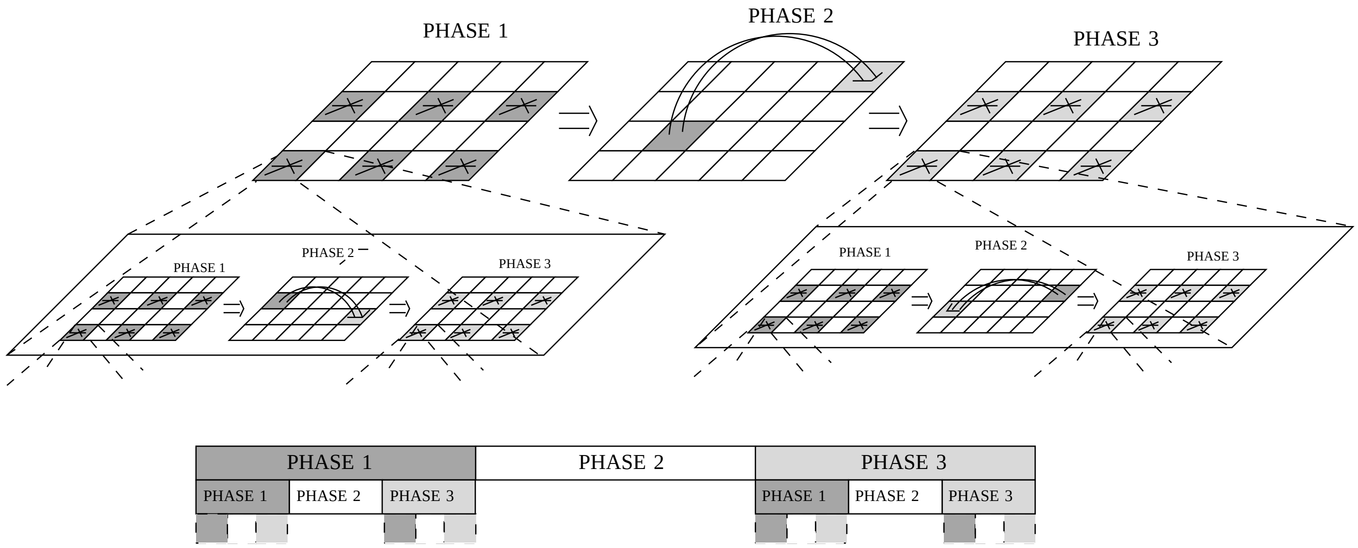

Fig. 4. The time division in a hierarchical scheme as well as the salient features of the three phases are illustrated.

is achievable in the network for all possible pairings between sources and destinations. $K > 0$ is a constant independent of n and the source-destination pairing.

- $\alpha \geq 3$: With high probability, an aggregate throughput:

$$T(n) \geq K\sqrt{n}$$

is achievable in the network for all possible pairings between sources and destinations. $K > 0$ is a constant independent of n and the source-destination pairing.

B. Cutset Upper Bound for Random S-D Pairings

Can we do better than the scaling in Theorem 4.1? So far we have been considering arbitrary source-destination pairings but clearly there are some pairings for which a much better scaling can be achieved. For example, if the source-destinations are all nearest neighbor to each other, then a linear capacity scaling can be achieved for any α. Thus, for the extended network case, we need to narrow down the class of S-D pairings to prove a sensible upper bound. In this section, we will focus on *random* S-D pairings and prove a high probability upper bound which matches the achievability result in Theorem 4.1.

Note that the hierarchical scheme is achieving near-global cooperation. In the context of dense networks, this yields a near linear number of degrees of freedom for communication. In the context of extended networks, in addition to the degrees of freedom provided, this scheme allows almost all nodes in the network to cooperate in transferring energy between any source-destination pair. In fact, we saw that in extended networks with $\alpha > 2$, our scheme is *power-limited* rather than degrees of freedom limited. A natural place to look for a matching upper bound is to consider a *cutset* bound on how much power can flow across the network. Divide the square into two equal halves. The total throughput between the S-D pairs with sources on the left half of the cut and with destinations on the right half (which w.h.p. is $1/4$ of the total throughput $T(n)$ between all S-D pairs) is bounded by the capacity of the MIMO link with all the nodes on the left cooperating as a super-source and with all the nodes on the right cooperating as a super-receiver. It is important to note that the power constraint for this MIMO link is a power constraint on each of the individual nodes, *not* a total power constraint on the whole array. We evaluate the scaling of this MIMO capacity to prove the following upper bound on the total throughput.

Theorem 4.2: Let $Q_{tot}(n)$ be the total power received by all the nodes on the right of the cut at a distance of at least 1 from the boundary, when each node on the left half is transmitting independent signals at full power. Then for every $\epsilon > 0$, the total throughput for random pairing is bounded with high probability by:

$$T(n) \leq n^{\epsilon} Q_{tot}(n).$$

Moreover, the scaling of the total received power can be evaluated to be:

$$Q_{tot}(n) = \begin{cases} \Theta(n^{2-\alpha/2}) & 2 \leq \alpha < 3 \\ \Theta(\sqrt{n}\,\log n) & \alpha = 3 \\ \Theta(\sqrt{n}) & \alpha > 3 \end{cases}$$

Theorem 4.2 together with Theorem 4.1 identify exactly the capacity scaling law in extended networks for *all* values of $\alpha \geq 2$.

Theorem 4.2 says two things of importance. First, it says that independent signalling at the transmit nodes is sufficient to achieve the cutset upper bound, as far as scaling is concerned. There is therefore no need, in order for the transmit nodes to cooperate, to do any sort of transmit beamforming. This is fortuitous since our hierarchical MIMO performs only independent signalling across the transmit nodes in the long-range MIMO phase. Second, it identifies the total received power as the fundamental quantity limiting performance. Depending on α, there is a dichotomy on how this quantity scales with the system size. This dichotomy can be interpreted as follows.

The total received power is dominated either by the power transferred between nodes near the cut or by the power transferred between nodes far away from the cut. There are relatively fewer node *pairs* near the cut than away from the cut (order $\sqrt{n}$ versus order n^2), but the channels between the nodes near the cut are considerably stronger than between the nodes far away from the cut. When the attenuation parameter α is less than 3, the received power is dominated by transfer between nodes far away from the cut. The hierarchical scheme, which involves at the top level of the hierarchy MIMO transmissions between clusters of size $n^{1-\epsilon}$ at distance $\sqrt{n}$ apart, achieves arbitrarily closely the required power transfer and is therefore optimal in this regime. When $\alpha \geq 3$, the received power in the cutset bound is dominated by the power transfer by the nodes near the cut. This can be achieved by nearest neighbor multihop and multihop is therefore optimal in this regime.

It should be noted that earlier works identified thresholds on α above which nearest neighbor multihop is order-optimal ($\alpha > 5$ in [7], $\alpha > 4.5$ in [10] and $\alpha > 4$ in [9].) All of them essentially use the same cutset bound as above. The fact that they did not identify the tightest threshold (which we have shown to be 3) is because their upper bounds on the cutset bound are not tight.

V. Conclusions

In this paper, we have shown that near global MIMO cooperation between nodes can be achieved in ad hoc networks without significant cooperation overhead. This is a surprising result, as it allows the full degrees of freedom in the network to be shared among all nodes and implies that interference is not a fundamental limitation. In dense networks where all nodes are within communication range of each other, this yields a linear capacity scaling. In extended networks, such near-global MIMO cooperation also allows the maximum transfer of energy between all source-destination pairs. This leads to the identification of the optimal (power-limited) capacity scaling law of extended networks for all values of α.

The key ideas behind our scheme are:

- using MIMO for long-range communication to achieve spatial multiplexing;
- local transmit and receive cooperation to maximize spatial reuse;
- setting up the intra-cluster cooperation such that it is yet another digital communication problem, but in a smaller network, thus enabling a hierarchical cooperation architecture.

Our result is based on only very weak assumptions about the channel. It is valid for any path loss exponent $\alpha \geq 2$. It holds regardless of whether there are multi-paths, as long as nodal separation is much larger than the carrier wavelength so that the phases of the channels are random. This is sufficient to enable MIMO. We have focused on the 2-D setting, where the nodes are on the plane, but our results generalize naturally to d-dimensional networks.

Acknowledgment

David Reed raised the question of whether a linear capacity scaling is possible, and that provides part of the impetus for this research. The authors would also like to thank to Emre Telatar, Shuchin Aeron and Venkatesh Saligrama for many helpful discussions. The work of Ayfer Özgür was supported by Swiss NSF grant Nr 200021-108089; the work of Olivier Lévêque was supported by Swiss NSF grant Nr PA002-108976; the work of David Tse was supported by the U.S. National Science Foundation via an ITR grant: ”The 3R’s of Spectrum Management: Reuse, Reduce and Recycle”.

References

[1] P. Gupta and P. R. Kumar, *The Capacity of Wireless Networks*, IEEE Trans. on Information Theory 42 (2), 2000, 388-404.

[2] M. Franceschetti, O. Dousse, D. Tse, P. Thiran, *On the Throughput Capacity of Random Wireless Networks*, Preprint, 2006.

[3] S. Aeron, V. Saligrama, *Wireless Ad hoc Networks: Strategies and Scaling Laws for the Fixed SNR Regime*, Preprint, 2006.

[4] G. J. Foschini, *Layered Space-Time Architecture For Wireless Communication in a Fading Environment when Using Multi-Element Antennas*, AT&T Bell Labs. Tech. J., vol.1 (2), pp.41-59, 1996.

[5] E. Telatar, *Capacity of Multi-Antenna Gaussian Channels"* European Trans. on Telecommunications, ETT, vol.10 (6), pp. 585-596, Nov. 1999.

[6] L. -L. Xie and P. R. Kumar, *A Network Information Theory for Wireless Communications: Scaling Laws and Optimal Operation*, IEEE Trans. on Information Theory 50 (5), 2004, 748-767.

[7] A. Jovicic, P. Viswanath and S. R. Kulkarni, *Upper Bounds to Transport Capacity of Wireless Networks*, IEEE Trans. on Information Theory 50 (11), 2004, 2555-2565.

[8] O. Lévêque and E. Telatar, *Information Theoretic Upper Bounds on the Capacity of Large, Extended Ad-Hoc Wireless Networks*, IEEE Trans. on Information Theory 51 (3), March 2005.

[9] L. -L. Xie, P. R. Kumar, *On the Path-Loss Attenuation Regime for Positive Cost and Linear Scaling of Transport Capacity in Wireless Networks*, Preprint, 2006.

[10] S. Ahmad, A. Jovicic and P. Viswanath, *Outer Bounds to the Capacity Region of Wireless Networks*, IEEE Trans. Inform. Theory, vol. 52, pp. 2770-2776, June 2006.

[11] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, 2005.