MIT Open Access Articles

On Capacity Scaling in Arbitrary Wireless Networks

On Capacity Scaling in Arbitrary Wireless Networks

Urs Niesen, *Student Member, IEEE*, Piyush Gupta, *Senior Member, IEEE*, and Devavrat Shah

Abstract—In recent work, Özgür, Lévêque, and Tse (2007) obtained a complete scaling characterization of throughput scaling for random extended wireless networks (i.e., n nodes are placed uniformly at random in a square region of area n). They showed that for small path-loss exponents $\alpha \in (2, 3]$, cooperative communication is order optimal, and for large path-loss exponents $\alpha > 3$, multihop communication is order optimal. However, their results (both the communication scheme and the proof technique) are strongly dependent on the regularity induced with high probability by the random node placement. In this paper, we consider the problem of characterizing the throughput scaling in extended wireless networks with arbitrary node placement. As a main result, we propose a more general novel cooperative communication scheme that works for arbitrarily placed nodes. For small path-loss exponents $\alpha \in (2, 3]$, we show that our scheme is order optimal for all node placements, and achieves exactly the same throughput scaling as in Özgür *et al.* This shows that the regularity of the node placement does not affect the scaling of the achievable rates for $\alpha \in (2, 3]$. The situation is, however, markedly different for large path-loss exponents $\alpha > 3$. We show that in this regime the scaling of the achievable per-node rates depends crucially on the regularity of the node placement. We then present a family of schemes that smoothly "interpolate" between multihop and cooperative communication, depending upon the level of regularity in the node placement. We establish order optimality of these schemes under adversarial node placement for $\alpha > 3$.

Index Terms—Arbitrary node placement, capacity scaling, cooperative communication, hierarchical relaying, multihop communication, wireless networks.

I. INTRODUCTION

CONSIDER a wireless network with n nodes placed on $[0, \sqrt{n}]^2$ (usually referred to as an *extended network*), with each node being the source for one of n source–destination pairs and the destination for another pair. The performance of this network is captured by $\rho^*(n)$, the largest uniformly achievable rate of communication between these source–destination pairs. While the scaling behavior of $\rho^*(n)$ as the number of nodes n goes to infinity is by now well understood for random node placement, little is known for the case of arbitrary node

placements. In this paper, we are interested in analyzing the impact of such arbitrary node placement on the scaling of $\rho^*(n)$.

A. Related Work

The problem of determining the scaling of $\rho^*(n)$ was first analyzed by Gupta and Kumar [1]. They show that, under random placement of nodes in the region, certain models of communication motivated by current technology, and random source–destination pairing, the maximum achievable per-node rate $\rho^*(n)$ can scale at most as $O(n^{-1/2})$. Moreover, it was shown that multihop communication can achieve essentially the same order of scaling.

Since [1], the problem has received a considerable amount of attention. One stream of work [2]–[8] has progressively broadened the conditions on the channel model and the communication model, under which multihop communication is order optimal. Specifically, with a power loss of $r^{-\alpha}$ for signals sent over distance r, it has been established that under *high* signal attenuation $\alpha > 3$ and random node placement, the best achievable per-node rate $\rho^*(n)$ for random source–destination pairing scales essentially like $\Theta(n^{-1/2})$ and that this scaling is achievable with multihop communication.

Another stream of work [8]–[12] has proposed progressively refined multiuser cooperative schemes, which have been shown to significantly outperform multihop communication in certain environments. In an exciting recent work, Özgür *et al.* [8] have shown that with nodes placed uniformly at random, and with *low* signal attenuation $\alpha \in (2, 3]$, a cooperative communication scheme can perform significantly better than multihop communication. More precisely, they show that for $\alpha \in (2, 3]$, the best achievable per-node rate for random source–destination pairing scales as $\rho^*(n) = O(n^{1-\alpha/2+\varepsilon})$ and cooperative communication achieves a per-node rate of $\Omega(n^{1-\alpha/2-\varepsilon})$ (here, $\varepsilon > 0$ is an arbitrary but fixed constant). That is, cooperative communication is essentially order optimal in the attenuation regime $\alpha \in (2, 3]$.

In summary, for random extended networks with random source–destination pairing, the optimal communication scheme exhibits the following threshold behavior: for $\alpha \in (2, 3]$, the cooperative communication scheme is order optimal, while for $\alpha > 3$, the multihop communication scheme is order optimal.

B. Our Contributions

The characterization of the scaling of $\rho^*(n)$ as a function of the path-loss exponent α mentioned in the last paragraph depends critically on the regularity induced with high probability by placing the nodes uniformly at random. However, a wireless network encountered in practice might not exhibit this amount of regularity. Our interest is therefore in understanding the impact of the node placement on the scaling of $\rho^*(n)$. To this end,

we consider wireless networks with arbitrary (i.e., determin-istic) node placement (with minimum-separation constraint).

The impact of this arbitrary node placement depends crucially on the path-loss exponent α. For small path-loss exponents $\alpha \in (2, 3]$, we show that for random source–destination pairing, the rate of the best communication scheme is upper bounded as $\rho^*(n) = O(\log^6(n) n^{1-\alpha/2})$. We then present a novel coopera-tive communication scheme that achieves for any path-loss ex-ponent $\alpha > 2$ a per-node rate of $\rho^{\mathsf{HR}}(n) \geq n^{1-\alpha/2-o(1)}$. Thus, our cooperative communication scheme is essentially order op-timal for any such arbitrary network with $\alpha \in (2, 3]$. In other words, in the small path-loss regime, the scaling of $\rho^*(n)$ is the same irrespective of the regularity of the node placement.

The situation is, however, quite different for large path-loss exponents $\alpha > 3$. We show that in this regime the scaling of $\rho^*(n)$ depends crucially on the regularity of the node place-ment, and multihop communication may not be order optimal for any value of α. In fact, for less regular networks, we need more complicated cooperative communication schemes to achieve optimal network performance. Towards that end, we present a family of communication schemes that smoothly "in-terpolate" between cooperative communication and multihop communication, and in which nodes communicate at scales that vary smoothly from local to global. The amount of "interpola-tion" between the cooperative and multihop schemes depends on the level of regularity of the underlying node placement. We establish the optimality of this family of schemes for all $\alpha > 3$ under adversarial node placement.

In summary, for $\alpha \in (2, 3]$, the regularity of the node place-ment has no impact on the scaling of $\rho^*(n)$. Cooperative com-munication is order optimal in this regime and achieves the same scaling as in the case of random node placement. For $\alpha > 3$, the regularity of the node placement strongly impacts the scaling of $\rho^*(n)$, and a communication scheme "interpolating" between multihop and cooperative communication depending on the reg-ularity of the node placement is order optimal (under adversarial node placement). In particular, simple multihop communication may not be order optimal for any $\alpha > 3$. This contrasts with the case of random node placement where multihop communication is order optimal for all $\alpha > 3$.

C. Organization

The remainder of this paper is organized as follows. Section II describes in detail the communication model. Section III pro-vides formal statements of our results. Sections IV and V describe our new cooperative communication scheme (for the $\alpha \in (2, 3]$ regime) and "interpolation" scheme (for the $\alpha > 3$ regime) for arbitrary wireless networks. Sections VI–XI contain proofs. Finally, Sections XII and XIII contain discussions and concluding remarks.

II. MODEL

In this section, we introduce some notational conventions and describe in detail the network and channel models.

We use the following conventions: K_i for different i denote strictly positive finite constants independent of n. Vectors and matrices are denoted by boldface whenever the vector or matrix

structure is of importance. We denote by $(\cdot)^T$ and $(\cdot)^{\dagger}$ trans-pose and conjugate transpose, respectively. To simplify nota-tion, we assume, when necessary, that fractions are integers and omit $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ operators.

Consider the square

$$A(n) \triangleq [0, \sqrt{n}]^2$$

of area n, and let $V(n) \subset A(n)$ be a set of $|V(n)| = n$ nodes on $A(n)$.[1] We say that $V(n)$ has *minimum-separation* $r_{\min}$ if $r_{u,v} \geq r_{\min}$ for all $u, v \in V(n)$, where $r_{u,v}$ is the Euclidean distance between nodes u and v. We use the same channel model as in [8]. Namely, the (sampled) received signal at node v is

$$y_v[t] = \sum_{u \in V(n) \setminus \{v\}} h_{u,v}[t] x_u[t] + z_v[t] \qquad (1)$$

for all $v \in V(n)$, and where $\{x_u[t]\}_{u,t}$ are the (sampled) signals sent by the nodes in $V(n)$. Here $\{z_v[t]\}_{v,t}$ are independent and identically distributed (i.i.d.) with distribution $\mathcal{N}_{\mathbb{C}}(0, 1)$ (i.e., circularly symmetric complex Gaussian with mean 0 and vari-ance 1), and

$$h_{u,v}[t] = r_{u,v}^{-\alpha/2} \exp(\sqrt{-1}\theta_{u,v}[t])$$

for *path-loss exponent* $\alpha > 2$. We assume that for each $t \in \mathbb{N}$, the phases $\{\theta_{u,v}[t]\}_{u,v}$ are i.i.d.[2] with uniform distribution on $[0, 2\pi)$. We either assume that for each $u, v \in V(n)$ the random process $\{\theta_{u,v}[t]\}_t$ is stationary ergodic in t, which is called *fast fading* in the following, or that for each $u, v \in V(n)$ the random process $\{\theta_{u,v}[t]\}_t$ is constant in t, which is called *slow fading* in the following. In either case, we assume that full channel state information (CSI) is available at all the nodes, i.e., each node knows all $\{\theta_{u,v}[t]\}_{u,v}$ at time t. While the full CSI assumption is quite strong, it can be shown that availability of a 2-bit *quantized* version of $\{\theta_{u,v}[t]\}_{u,v}$ at all nodes is sufficient for the achievable schemes presented here (see Section XII-A for the details). We also impose an average power constraint of 1 on the signal $\{x_u[t]\}_t$ for every node $u \in V(n)$.

Each node $u \in V(n)$ wants to transmit information at uniform rate $\rho(n)$ to some other node $w \in V(n)$. We call u the *source* and w the *destination* node of this communication pair. The set of all communication pairs can be described by a *traffic matrix* $\lambda(n) \in \{0, 1\}^{n \times n}$, where the entry in $\lambda(n)$ corresponding to (u, w) is equal to 1 if node u is a source for node w. We say that $\lambda(n)$ is a *permutation traffic matrix* if it is a permutation matrix (i.e., every node is a source for exactly one communication pair and a destination for exactly one

---

[1] The setting considered here with n nodes placed on a square of area n is called an *extended* network. If the n nodes are placed on a square of unit area, we speak of a *dense* network. While dense networks are not treated in detail in this paper, we briefly discuss implications of the results for the dense setting in Section XII-C.

[2] It is worth pointing out that recent work [13] suggests that, under certain assumptions on scattering elements, for $\alpha \in (2, 3)$, and for very large values of n, the i.i.d. phase assumption as a function of $u, v \in V(n)$ used here is too optimistic. However, subsequent work by the same authors [14] shows that under different assumptions on the scatterers, the channel model used here is still valid even for $\alpha \in (2, 3)$, and for very large values of n. This indicates that the question of channel modeling for very large networks in the low path-loss regime is somewhat delicate and requires further investigation. We point out that for $\alpha \geq 3$ this issue does not arise.

communication pair). For a traffic matrix $\lambda(n)$, let $\rho^*(n)$ be the highest rate of communication that is uniformly achievable for each source–destination pair. For a permutation traffic matrix $\lambda(n)$, $\rho^*(n)$ can also be understood as the maximal achievable per-node rate.

III. MAIN RESULTS

This section presents the formal statement of our results. The results are divided into two parts. In Section III-A, we consider low path-loss exponents, i.e., $\alpha \in (2, 3]$. We present a cooperative communication scheme for arbitrary node placement and for either fast or slow fading. We show that this communication scheme is order optimal for all node placements when $\alpha \in (2, 3]$. In Section III-B, we consider high path-loss exponents, i.e., $\alpha > 3$. We present a communication scheme that "interpolates" between the cooperative and multihop communication schemes, depending on the regularity of the node placement. We show that this communication scheme is order optimal under adversarial node placement with regularity constraint when $\alpha > 3$.

A. Low Path-Loss Regime $\alpha \in (2, 3]$

The first result proposes a novel communication scheme, called *hierarchical relaying* in the following, and bounds the per-node rate $\rho^{\mathsf{HR}}(n)$ that it achieves. This provides a lower bound to $\rho^*(n)$, the largest achievable per-node rate. The hierarchical relaying scheme enables cooperative communication on the scale of the network size. In the random node placement case, this cooperation could be enabled in a cluster around the source node (cooperatively transmitting) and in a cluster around its destination node (cooperatively receiving). With arbitrary node placement, such an approach does no longer work, as both the source as well as the destination nodes may be isolated. The hierarchical relaying scheme circumvents this issue by relaying data between each source–destination pair over a densely populated region in the network. A detailed description of this scheme is provided in Section IV, and the proof of Theorem 1 is contained in Section VII.

Theorem 1: Under fast fading, for any $\alpha > 2, r_{\min} \in (0, 1)$, and $\delta \in (0, 1/2)$, there exists

$$b_1(n) \geq n^{-O(\log^{\delta - 1/2}(n))}$$

such that for any n, node placement $V(n)$ with minimum separation $r_{\min}$, and permutation traffic matrix $\lambda(n)$, we have

$$\rho^*(n) \geq \rho^{\mathsf{HR}}(n) \geq b_1(n)n^{1-\alpha/2}.$$

The same conclusion holds for slow fading with probability at least

$$1 - \exp\left(-2^{\Omega(\log^{1/2+\delta}(n))}\right) = 1 - o(1)$$

as $n \to \infty$.

Theorem 1 shows that the per-node rate $\rho^{\mathsf{HR}}(n)$ achievable by the hierarchical relaying scheme is at least $n^{1-\alpha/2-\beta(n)}$, where the "loss" term $\beta(n)$ converges to zero as $n \to \infty$ at a rate arbitrarily close to $O(\log^{-1/2}(n))$ (by choosing δ small). The

performance of the hierarchical relaying scheme can intuitively be understood as follows. As mentioned before, the scheme achieves cooperation on a global scale. This leads to a multi-antenna gain of order n. On the other hand, communication is over a distance of order $n^{1/2}$, leading to a power loss of order $n^{-\alpha/2}$. Combining these two factors results in a per-node rate of $n^{1-\alpha/2}$.

We note that Theorem 1 remains valid under somewhat weaker conditions than having minimum separation $r_{\min} \in (0, 1)$. Specifically, we show that the result of Özgür *et al.* [8] can be recovered through Theorem 1 as the random node placement satisfies these weaker conditions. We discuss this in more detail in Section XII-D.

The next theorem establishes optimality of the hierarchical relaying scheme in the range of $\alpha \in (2, 3]$ for arbitrary node placement. The proof of the theorem is presented in Section VIII.

Theorem 2: Under either fast or slow fading, for any $\alpha \in (2, 3], r_{\min} \in (0, 1)$, there exists $b_2(n) = O(\log^6(n))$ such that for any n, node placement $V(n)$ with minimum separation $r_{\min}$, and for $\lambda(n)$ chosen uniformly at random from the set of all permutation traffic matrices, we have

$$\rho^*(n) \leq b_2(n)n^{1-\alpha/2}$$

with probability $1 - o(1)$ as $n \to \infty$.

Note that Theorem 2 holds only with probability $1 - o(1)$ for different reasons for the slow- and fast-fading case. For fast fading, this is due to the randomness in the selection of the permutation traffic matrix. In other words, for fast fading, with high probability, we select a traffic matrix for which the theorem holds. For the slow-fading case, there is additional randomness due to the fading realization. Here, with high probability, we select a traffic matrix and we experience a fading for which the theorem hold.

Comparing Theorems 1 and 2, we see that for $\alpha \in (2, 3]$, the proposed hierarchical relaying scheme is order optimal, in the sense that

$$\lim_{n \to \infty} \frac{\log(\rho^{\mathsf{HR}}(n))}{\log(n)} = \lim_{n \to \infty} \frac{\log(\rho^*(n))}{\log(n)} = 1 - \alpha/2.$$

Moreover, the rate it achieves is the same order as is achievable in the case of randomly placed nodes. Hence, in the low path-loss regime $\alpha \in (2, 3]$, the heterogeneity caused by the arbitrary node placement has no effect on achievable communication rates.

B. High Path-Loss Regime $\alpha > 3$

We now turn to the high path-loss regime $\alpha > 3$. In the case of *randomly* placed nodes, multihop communication achieves a per-node rate of $\rho^{\mathsf{MH}}(n) = \Omega(n^{-1/2})$ with probability $1 - o(1)$ and is order optimal for $\alpha > 3$. For *arbitrarily* placed nodes, the situation is quite different as Theorem 3 shows. The proof of Theorem 3 is contained in Section IX.

Theorem 3: Under either fast or slow fading, for any $\alpha > 3$, for any n, there exists a node placement $V(n)$ with minimum

separation $1/2$ such that for $\lambda(n)$ chosen uniformly at random from the set of all permutation traffic matrices, we have

$$\rho^*(n) \leq 2^{2+5\alpha} n^{1-\alpha/2}$$
$$\rho^{\mathsf{MH}}(n) \leq 4^\alpha n^{-\alpha/2}$$

as $n \to \infty$ with probability $1 - o(1)$.

Comparing Theorem 3 with Theorem 1 shows that under adversarial node placement with minimum-separation constraint the hierarchical relaying scheme is order optimal even when $\alpha > 3$. Moreover, Theorem 3 shows that there exist node placements satisfying a minimum separation constraint for which hierarchical relaying achieves a rate of at least a factor of order n higher than multihop communication for any $\alpha > 3$. In other words, for those node placements, cooperative communication is necessary for order optimality also for any $\alpha > 3$, in stark contrast to the situation with random node placement, where multihop communication is order optimal for all $\alpha > 3$.

Theorem 3 suggests that it is the level of regularity of the node placement that decides what scheme to choose for path-loss exponent $\alpha > 3$. So far, we have seen two extreme cases. For random node placement, resulting in very regular node placements with high probability, only local cooperation is necessary and multihop is an order-optimal communication scheme. For adversarial arbitrary node placement, resulting in a very irregular node placement, global cooperation is necessary and hierarchical relaying is an order-optimal communication scheme. We now make this notion of regularity precise, and show that, depending on the regularity of the node placement, an appropriate "interpolation" between multihop and hierarchical relaying is required for $\alpha > 3$ to achieve the optimal performance. We refer to this "interpolation" scheme as *cooperative multihop* communication in the following.

Before we state the result, we need to introduce some notation. Consider again a node placement $V(n) \subset A(n)$ with minimum separation $r_{\min} \in (0, 1)$. Divide $A(n)$ into squares of sidelength $d(n) \leq \sqrt{n}$, and fix a constant $\mu \in (0, 1]$. We say that $V(n)$ *is μ-regular at resolution $d(n)$* if every such square contains at least $\mu d^2(n)$ nodes. Note that every node placement is trivially 1-regular at resolution $\sqrt{n}$; a random node placement can be shown to be μ-regular at resolution $\log(n)$ with probability $1 - o(1)$ as $n \to \infty$ for any $\mu < 1$; and nodes that are placed on each point in the integer lattice inside $A(n)$ are 1-regular at resolution 1.

The cooperative multihop scheme enables cooperative communication on the scale of regularity $d(n)$. Neighboring squares of sidelength $d(n)$ cooperatively communicate with each other. To transmit between a source and its destination, we use multihop communication over those squares. In other words, we use cooperative communication at small scale $d(n)$, and multihop communication at large scale $\sqrt{n}$. For regular node placements, i.e., $d(n) = 1$, the cooperative multihop scheme becomes the classical multihop scheme. For very irregular node placement, i.e., $d(n) = n^{1/2}$, the cooperative multihop scheme becomes the hierarchical relaying scheme discussed in the last section.

The next theorem provides a lower bound on the per-node rate $\rho^{\mathsf{CMH}}(n)$ achievable with the cooperative multihop scheme. The proof of the theorem can be found in Section X.

Theorem 4: Under fast fading, for any $\alpha > 2, r_{\min} \in (0, 1), \mu \in (0, 1)$, and $\delta \in (0, 1/2)$, there exists

$$b_3(n) \geq n^{-O(\log^{\delta - 1/2}(n))}$$

such that for any n, node placement $V(n)$ with minimum separation $r_{\min}$, and permutation traffic matrix $\lambda(n)$, we have

$$\rho^*(n) \geq \rho^{\mathsf{CMH}}(n) \geq b_3(n) d^{*3-\alpha}(n) n^{-1/2}$$

where

$$d^*(n) \triangleq \min\{h : V(n) \text{ is } \mu \text{ regular at resolution } h\}.$$

The same conclusion holds for slow fading with probability $1 - o(1)$ as $n \to \infty$.

Theorem 4 shows that if $V(n)$ is regular at resolution $d^*(n)$, then a per-node rate of at least $\rho^{\mathsf{CMH}}(n) \geq d^{*3-\alpha}(n) n^{-1/2-\beta(n)}$ is achievable, where, as before, the "loss" term $\beta(n)$ converges to zero as $n \to \infty$ at a rate arbitrarily close to $O(\log^{-1/2}(n))$. The performance of the cooperative multihop scheme can intuitively be understood as follows. The scheme achieves cooperation on a scale of $d^2(n)$. This leads to a multiantenna gain of order $d^2(n)$. On the other hand, communication is over a distance of order $d(n)$, leading to a power loss of order $d^{-\alpha}(n)$. Moreover, each source–destination pair at a distance of order $n^{1/2}$ must transmit their data over order $n^{1/2} d^{-1}(n)$ many hops, leading to a multihop loss of $n^{-1/2} d(n)$. Combining these three factors results in a per-node rate of $d^{3-\alpha}(n) n^{-1/2}$.

The next theorem shows that Theorem 4 is tight under adversarial node placement under a constraint on the regularity. The proof of the theorem is presented in Section XI.

Theorem 5: Under either fast or slow fading, for any $\alpha > 3$, there exists $b_4(n) = O(\log^6(n))$, such that for any n, and $d^*(n)$, there exists a node placement $V(n)$ with minimum separation $1/2$ and $1/2$-regular at resolution $d^*(n)$ such that for $\lambda(n)$ chosen uniformly at random from the set of all permutation traffic matrices, we have

$$\rho^*(n) \leq b_4(n) d^{*3-\alpha}(n) n^{-1/2}$$

with probability $1 - o(1)$ as $n \to \infty$.

As an example, assume that

$$d^*(n) = n^\eta$$

for some $\eta \geq 0$. Then, Theorem 4 shows that for any node placement of regularity $d^*(n)$ and $\alpha > 3$

$$\rho^{\mathsf{CMH}}(n) \geq n^{(3-\alpha)\eta - 1/2 - \beta(n)}$$

where $\beta(n)$ converges to zero as $n \to \infty$ at a rate arbitrarily close to $O(\log^{-1/2}(n))$. In other words

$$\lim_{n \to \infty} \frac{\log(\rho^{\mathsf{CMH}}(n))}{\log(n)} \geq (3 - \alpha)\eta - 1/2.$$
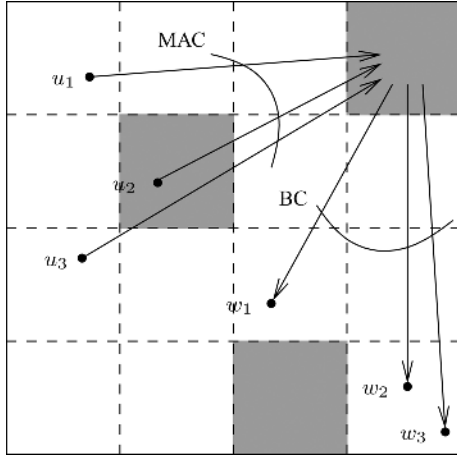
Fig. 1. Sketch of one level of the hierarchical relaying scheme. Here $\{(u_i, w_i)\}_{i=1}^3$ are three source–destination pairs. Groups of source–destination pairs relay their traffic over dense squarelets, which contain a number of nodes proportional to their area (shaded). We time share between the different dense squarelets used as relays. Within all these relay squarelets, the scheme is used recursively to enable joint decoding and encoding at each relay.

Moreover, by Theorem 5, there exist node placements with the same regularity such that for random permutation traffic with high probability $\rho^*(n)$ is (essentially) of the same order, in the sense that

$$\lim_{n \to \infty} \frac{\log(\rho^*(n))}{\log(n)} \leq (3 - \alpha)\eta - 1/2.$$

In particular, for $\eta = 0$ (i.e., regular node placement), and for $\eta = \log\log(n)/\log(n)$ (i.e., random node placement), we obtain the order $n^{-1/2}$ scaling as expected. For $\eta = 1/2$ (i.e., completely irregular node placement), we obtain the order $n^{1-\alpha/2}$ scaling as in Theorems 1 and 3.

IV. HIERARCHICAL RELAYING SCHEME

This section describes the architecture of our hierarchical relaying scheme. On a high level, the construction of this scheme is as follows. Consider n nodes $V(n)$ placed arbitrarily on the square region $A(n)$ with a minimum separation $r_{\min}$. Divide $A(n)$ into squarelets of equal size. Call a squarelet *dense*, if it contains a number of nodes proportional to its area. For each source–destination pair, choose such a dense squarelet as a *relay*, over which it will transmit information (see Fig. 1).

Consider now one such relay squarelet and the nodes that are transmitting information over it. If we assume for the moment that all the nodes within the same relay squarelet could cooperate, then we would have a multiple-access channel (MAC) between the source nodes and the relay squarelet, where each of the source nodes has one transmit antenna, and the relay squarelet (acting as one node) has many receive antennas. Between the relay squarelet and the destination nodes, we would have a broadcast channel (BC), where each destination node has one receive antenna, and the relay squarelet (acting again as one node) has many transmit antennas. The cooperation gain from using this kind of scheme arises from the use of multiple antennas for these multiple access and broadcast channels.

To actually enable this kind of cooperation at the relay squarelet, local communication within the relay squarelets

is necessary. It can be shown that this local communication problem is actually the same as the original problem, but at a smaller scale. Hence, we can use the same scheme recursively to solve this subproblem. We terminate the recursion after several iterations, at which point we use simple time-division multiple access (TDMA) to bootstrap the scheme.

The construction of the hierarchical relaying scheme is presented in detail in Section IV-A. A back-of-the-envelope calculation of the per-node rate it achieves is presented in Section IV-B. A detailed analysis of the hierarchical relaying scheme is presented in Sections VI and VII.

A. Construction

Recall that

$$A(b) \triangleq [0, \sqrt{b}]^2$$

is the square region of area b. The scheme described here assumes that n nodes are placed arbitrarily in $A(n)$ with minimum separation $r_{\min} \in (0, 1)$. We want to find some rate, say ρ_0, that can be supported for all n source–destination pairs of a given permutation traffic matrix $\lambda(n)$. The scheme that is described below is "recursive" (and hence hierarchical) in the following sense. In order to achieve rate ρ_0 for n nodes in $A(n)$, it will use as a building block a scheme for supporting rate ρ_1 for a network of

$$n_1 \triangleq \frac{n}{2\gamma(n)}$$

nodes over $A(a_1)$ (square of area a_1) with

$$a_1 \triangleq \frac{n}{\gamma(n)}$$

for any permutation traffic matrix $\lambda(n_1)$ of n_1 nodes. Here the *branching factor* $\gamma(n)$ is a function such that $\gamma(n) \to \infty$ as $n \to \infty$. We will optimize over the choice of $\gamma(n)$ later. The same construction is used for the scheme over $A(a_1)$, and so on. In general, our scheme does the following at level $\ell \geq 0$ of the hierarchy (or recursion). In order to achieve rate ρ_ℓ for any permutation traffic matrix $\lambda(n_\ell)$ over

$$n_\ell \triangleq \frac{n}{2^\ell \gamma^\ell(n)}$$

nodes in $A(a_\ell)$, with

$$a_\ell \triangleq \frac{n}{\gamma^\ell(n)}$$

use a scheme achieving rate $\rho_{\ell+1}$ over $n_{\ell+1}$ nodes in $A(a_{\ell+1})$ for any permutation traffic matrix $\lambda(n_{\ell+1})$. The recursion is terminated at some level $L(n)$ to be chosen later.

We now describe how the hierarchy is constructed between levels ℓ and $\ell + 1$ for $0 \leq \ell < L(n)$. Each source–destination pair chooses some squarelet as a relay over which it transmits its message. This relaying of messages takes place in two phases—a multiple-access phase and a broadcast phase. We first describe the selection of relay squarelets, then the operation of the network during the multiple-access and broadcast phases, and finally, the termination of the hierarchical construction.
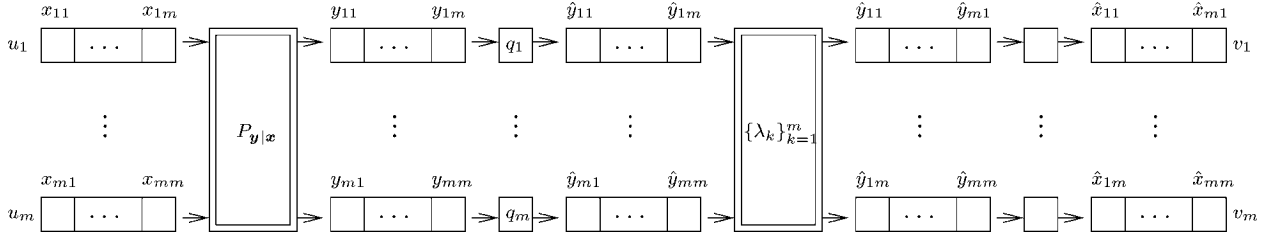
Fig. 2. Description of the multiple access phase at level ℓ in the hierarchy with $m \triangleq n_{\ell+1}$. The first system block represents the wireless channel, connecting source nodes $\{u_i\}_{i=1}^{n_{\ell+1}}$ with relay nodes $\{v_i\}_{i=1}^{n_{\ell+1}}$. The second system block represents quantizers $\{q_i\}_{i=1}^{n_{\ell+1}}$ used at the relay nodes. The third system block represents using $n_{\ell+1}$ times the communication scheme at level $\ell+1$ (organized as $n_{\ell+1}$ permutation traffic matrices $\{\lambda_k(n_{\ell+1})\}_{k=1}^{n_{\ell+1}}$) to "transpose" the matrix of quantized observations $\{\hat{y}_{ij}\}_{i,j=1}^{n_{\ell+1}}$. In other words, before the third system block, node v_1 has access to $\{\hat{y}_{1j}\}_{j=1}^{n_{\ell+1}}$, and after the third system block, node v_1 has access to $\{\hat{y}_{i1}\}_{i=1}^{n_{\ell+1}}$. The fourth system block represents matched filters used at the relay nodes.

1) Setting up Relays: Given n_ℓ nodes in $A(a_\ell)$, divide the square region $A(a_\ell)$ into $\gamma(n)$ equal sized squarelets. Denote them by $\{A_k(a_{\ell+1})\}_{k=1}^{\gamma(n)}$. Call a squarelet *dense* if it contains at least $n_\ell/2\gamma(n) = n_{\ell+1}$ nodes. In other words, a dense squarelet contains a number of nodes of at least a $1/2^{\ell+1}$ fraction of its area. We show that since the nodes in $A(a_\ell)$ have constant minimum separation $r_{\min}$, a squarelet can contain at most $O(a_{\ell+1}) = O(a_\ell/\gamma(n))$ nodes, and hence, that there are at least $\Theta(2^{-\ell}\gamma(n))$ dense squarelets. Each source–destination pair chooses a dense squarelet such that both the source and the destination are at a distance $\Omega(\sqrt{a_{\ell+1}})$ from it. We call this dense squarelet the *relay* of this source–destination pair. We show that the relays can be chosen such that each relay squarelet has at most $n_{\ell+1}$ communication pairs that use it as relay, and we assume this worst case in the following discussion.

2) Multiple-Access Phase: Source nodes that are assigned to the same (dense) relay squarelet send their messages simultaneously to that relay. We time share between the $\Theta(2^{-\ell}\gamma(n))$ different relay squarelets. If the nodes in the relay squarelet could cooperate, we would be dealing with a MAC with at most $n_{\ell+1}$ transmitters, each with one antenna, and one receiver with at least $n_{\ell+1}$ antennas. In order to achieve this cooperation, we use a hierarchical (i.e., recursive) construction. For this recursive construction, assume that we have access to a communication scheme to transmit data according to a permutation traffic matrix $\lambda(n_{\ell+1})$ between $n_{\ell+1}$ nodes located in a square of area $a_{\ell+1}$. We now show how this scheme at scale $a_{\ell+1}$ can be used to construct a scheme for scale a_ℓ (see Fig. 2).

Suppose there are $n_{\ell+1}$ source nodes $u_1, \ldots, u_{n_{\ell+1}}$ [located anywhere in $A(a_\ell)$] that relay their message over the $n_{\ell+1}$ relay nodes $v_1, \ldots, v_{n_{\ell+1}}$ (located in the same dense squarelet of area $a_{\ell+1}$). Each source node u_i divides its message bits into $n_{\ell+1}$ parts of equal length. Denote by x_{ij} the encoded part j of the message bits of node u_i (x_{ij} is really a large sequence of channel symbols; to simplify the exposition, we will, however, assume it is only a single symbol). The message parts corresponding to $\{x_{ij}\}_{i=1}^{n_{\ell+1}}$ will be relayed over node v_j, as will become clear in the following. Sources $\{u_i\}_{i=1}^{n_{\ell+1}}$ transmit $\{x_{ij}\}_{i=1}^{n_{\ell+1}}$ at time j for $j \in \{1, \ldots n_{\ell+1}\}$.

Let y_{kj} be the observed channel output at relay v_k at time j. Note that y_{kj} depends only on channel inputs $\{x_{ij}\}_{i=1}^{n_{\ell+1}}$. In order to decode the message parts corresponding to $\{x_{ij}\}_{i=1}^{n_{\ell+1}}$ at relay node v_j, it needs to obtain the observations $\{y_{ij}\}_{i=1}^{n_{\ell+1}}$ from all other relay nodes. In other words, all relays need to exchange

information. For this, each relay v_k quantizes its observation $\{y_{kj}\}_{j=1}^{n_{\ell+1}}$ at an appropriate rate K independent of n to obtain $\{\hat{y}_{kj}\}_{j=1}^{n_{\ell+1}}$. Quantized observation $\hat{y}_{kj}$ is to be sent from relay v_k to relay v_j. Thus, each of the $n_{\ell+1}$ relay nodes now has a message of size K for every other relay node.

This communication demand within the relay squarelet can be organized as $n_{\ell+1}$ permutation traffic matrices $\{\lambda_j(n_{\ell+1})\}_{j=1}^{n_{\ell+1}}$ between the $n_{\ell+1}$ relay nodes. Note that these relay nodes are located in the same square of area $a_{\ell+1}$. In other words, we are now faced with the original problem, but at smaller scale $a_{\ell+1}$. Therefore, using $n_{\ell+1}$ times the assumed scheme for transmitting according to a permutation traffic matrix for $n_{\ell+1}$ nodes in $A(a_{\ell+1})$, relay v_j can obtain all quantized observations $\{\hat{y}_{ij}\}_{i=1}^{n_{\ell+1}}$. Now v_j uses $n_{\ell+1}$ matched filters on $\{\hat{y}_{ij}\}_{i=1}^{n_{\ell+1}}$ to obtain estimates $\{\hat{x}_{ij}\}_{i=1}^{n_{\ell+1}}$ of $\{x_{ij}\}_{i=1}^{n_{\ell+1}}$. In other words, each node v_j computes[3]

$$\hat{x}_{ij} = \sum_{k=1}^{n_{\ell+1}} \frac{h_{u_i,v_k}^\dagger[j]}{\sqrt{\sum_k |h_{u_i,v_k}[j]|^2}} \hat{y}_{kj}$$

for every $i \in \{1, \ldots, n_{\ell+1}\}$. Using these estimates it then decodes the messages corresponding to $\{x_{ij}\}_{i=1}^{n_{\ell+1}}$.

3) Broadcast Phase: Nodes in the same relay squarelet then send their decoded messages simultaneously to the destination nodes corresponding to this relay. We time share between the different relay squarelets. If the nodes in the relay squarelet could cooperate, we would be dealing with a BC with one transmitter with at least $n_{\ell+1}$ antennas and with at most $n_{\ell+1}$ receivers, each with one antenna. In order to achieve this cooperation, a similar hierarchical construction as for the MAC phase is used. As in the MAC phase, assume that we have access to a scheme to transmit data according to a permutation traffic matrix $\lambda(n_{\ell+1})$ between $n_{\ell+1}$ nodes located in a square of area $a_{\ell+1}$. We again use this scheme at scale $a_{\ell+1}$ in the construction of the scheme for scale a_ℓ (see Fig. 3).

Suppose there are $n_{\ell+1}$ relay nodes $v_1, \ldots, v_{n_{\ell+1}}$ (located in the same dense squarelet of area $a_{\ell+1}$) that relay traffic for $n_{\ell+1}$ destination nodes $w_1, \ldots, w_{n_{\ell+1}}$ [located anywhere in $A(a_\ell)$]. Recall that at the end of the MAC phase, each relay node v_j has (assuming decoding was successful) access to parts j of the message bits of all source nodes $\{u_i\}_{i=1}^{n_{\ell+1}}$. Node v_j re-encodes

---

[3]Note that, since we assume full CSI, node v_j has access to the channel gains $\{h_{u_i,v_k}[j]\}_{i,k}$ at any time $t \geq j$. In particular, this is the case at the time the matched filtering is performed.
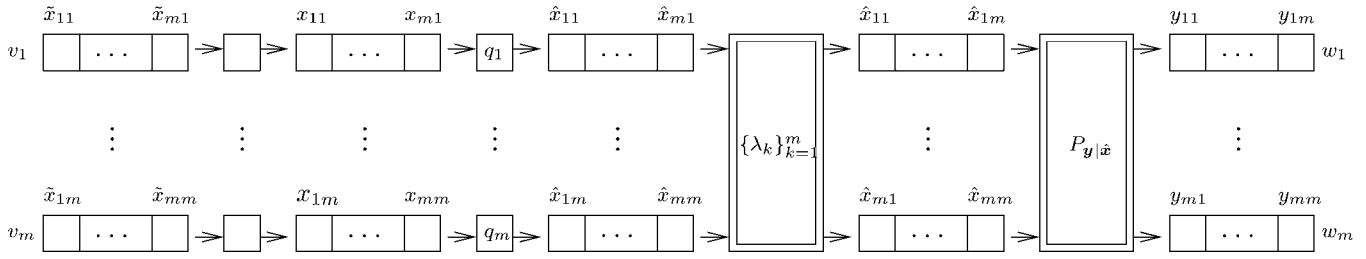
Fig. 3. Description of the broadcast phase at level ℓ in the hierarchy with $m \triangleq n_{\ell+1}$. The first system block represents transmit beamforming at each of the relay nodes $\{v_i\}_{i=1}^{n_{\ell+1}}$. The second system block represents quantizers $\{q_i\}_{i=1}^{n_{\ell+1}}$ used at the relay nodes. The third system block represents using $n_{\ell+1}$ times the communication scheme at level $\ell + 1$ (organized as $n_{\ell+1}$ permutation traffic matrices $\{\lambda_k(n_{\ell+1})\}_{k=1}^{n_{\ell+1}}$) to "transpose" the matrix of quantized beamformed channel symbols $\{\hat{x}_{ij}\}_{i,j=1}^{n_{\ell+1}}$. In other words, before the third system block, node v_1 has access to $\{\hat{x}_{i1}\}_{i=1}^{n_{\ell+1}}$, and after the third system block, node v_1 has access to $\{\hat{x}_{1j}\}_{j=1}^{n_{\ell+1}}$. The fourth system block is the wireless channel, connecting relay nodes $\{v_i\}_{i=1}^{n_{\ell+1}}$ with destination nodes $\{w_i\}_{i=1}^{n_{\ell+1}}$.

these parts independently; call $\{\tilde{x}_{ij}\}_{i=1}^{n_{\ell+1}}$ the encoded channel symbols (as before, we assume $\tilde{x}_{ij}$ is only a single symbol to simplify exposition). Relay node v_j then performs transmit beamforming on $\{\tilde{x}_{ij}\}_{i=1}^{n_{\ell+1}}$ for the $n_{\ell+1}$ transmit antennas of $\{v_k\}_{k=1}^{n_{\ell+1}}$ to be sent at time $T + j$ (for some appropriately chosen $T > 0$ not depending on j). Call x_{kj} the resulting channel symbol to be sent from relay node v_k. Then[4]

$$x_{kj} = \sum_i \frac{h_{v_k,w_i}^\dagger[T+j]}{\sqrt{\sum_k |h_{v_k,w_i}[T+j]|^2}} \tilde{x}_{ij}.$$

In order to actually send this channel symbol, relay node v_k needs to obtain x_{kj} from node v_j. Thus, again all relay nodes need to exchange information.

To enable local cooperation within the relay squarelet, each relay node v_j quantizes its beamformed channel symbols $\{x_{kj}\}_{k=1}^{n_{\ell+1}}$ at an appropriate rate $K \log(n)$ with K independent of n to obtain $\{\hat{x}_{kj}\}_{k=1}^{n_{\ell+1}}$. Now, quantized value $\hat{x}_{kj}$ is sent from relay v_j to relay v_k. Thus, each of the $n_{\ell+1}$ relay nodes now has a message of size $K \log(n)$ for every other relay node.

This communication demand within the relay squarelet can be organized as $n_{\ell+1}$ permutation traffic matrices $\{\lambda_k(n_{\ell+1})\}_{k=1}^{n_{\ell+1}}$ between the $n_{\ell+1}$ relay nodes. Note that these relay nodes are located in the same square of area $a_{\ell+1}$. Hence, we are again faced with the original problem, but at smaller scale $a_{\ell+1}$. Using $n_{\ell+1}$ times the assumed scheme for transmitting according to a permutation traffic matrix for $n_{\ell+1}$ nodes in $A(a_{\ell+1})$, relay v_k can obtain all quantized beamformed channel symbols $\{\hat{x}_{kj}\}_{j=1}^{n_{\ell+1}}$. Now each v_k sends $\hat{x}_{kj}$ over the wireless channel at time instance $T + j$ (with T chosen to account for the preceding MAC phase and the local cooperation in the BC phase). Call y_{ij} the received channel output at destination node w_i at time instance $T + j$. Using y_{ij}, destination node w_i can now decode part j of the message bits of its source node u_i.

4) Spatial Reuse and Termination of Recursion: The scheme does appropriately weighted time-division among different levels $0 \leq \ell \leq L(n)$. Within any level $\ell \geq 1$, multiple regions of the original square $A(n)$ of area n are being operated in parallel. The details related to the effects of interference between

different regions operating at the same level of hierarchy are discussed in the proofs.

The recursive construction terminates at some large enough level $L = L(n)$ (to be chosen later). At this scale, we have n_L nodes in area $A(a_L)$. A permutation traffic matrix at this level comprises n_L source–destination pairs. These transmissions are performed using TDMA. Again, multiple regions in the original square of area n at level L are active simultaneously.

B. Achievable Rates

Here we present a back-of-the-envelope calculation of the per-node rate $\rho^{\mathsf{HR}}(n)$ achievable with the hierarchical relaying scheme described in the previous section. The complete proof is stated in Section VII. We assume throughout that long block codes and corresponding optimal decoders are used for transmission.

Instead of computing the rate achieved by hierarchical relaying, it will be convenient to instead analyze its inverse, i.e., the time utilized for transmission of a single message bit from each source to its destination under a permutation traffic matrix $\lambda(n)$. Using the hierarchical relaying scheme, each message travels through L levels of the hierarchy. Call $\tau_\ell(n)$ the amount of time spent for the transmission of one message bit between each of the n_ℓ source–destination pairs at level ℓ in the hierarchy. We compute $\tau_\ell(n)$ recursively.

At any level $\ell \geq 1$, there are multiple regions of area a_ℓ operating at the same time. Due to the spatial reuse, each of these regions gets to transmit a constant fraction of time. It can be shown that the addition of interference due to this spatial reuse leads only to a constant loss in achievable rate. Hence, the time required to send one message bit is only a constant factor higher than the one needed if region $A(a_\ell)$ is considered separately. Consider now one such region $A(a_\ell)$. By the time-sharing construction, only one of its $\Theta(2^{-\ell}\gamma(n))$ dense relay squarelets of area $a_{\ell+1}$ is active at any given moment. Hence, the time required to operate all relay squarelets is a $\Theta(2^{-\ell}\gamma(n))$ factor higher than for just one relay squarelet separately. Consider now one such relay squarelet, and assume $n_{\ell+1}$ source nodes in $A(a_\ell)$ communicate each $n_{\ell+1}$ message bits to their respective destination nodes through a MAC phase and BC phase with the help of the $n_{\ell+1}$ relay nodes in this relay squarelet of area $a_{\ell+1}$.

In the MAC phase, each of the $n_{\ell+1}$ sources simultaneously sends one bit to each of the $n_{\ell+1}$ relay nodes. The total time for this transmission is composed of two terms.

---

[4]Note that, since we only assume causal CSI, relay node v_j does not actually have access to $\{h_{v_k,w_i}[T+j]\}_{k,i}$ at the time the beamforming is performed. This problem can, however, be circumvented. The details are provided in the proofs (see Lemma 10).

i) Transmission of $n_{\ell+1}$ message bits from each of the $n_{\ell+1}$ source nodes to those many relay nodes. Since we time share between $\Theta(2^{-\ell}\gamma(n))$ relay squarelets, we can transmit with an average power constraint of $\Theta(2^{-\ell}\gamma(n))$ during the time a relay squarelet is active, and still satisfies the overall average power constraint of 1. With this "bursty" transmission strategy, we require a total of

$$O\left(n_{\ell+1}\frac{a_\ell^{\alpha/2}}{2^{-\ell}\gamma(n)n_{\ell+1}}\right)$$
$$= O\left(n_{\ell+1}4^\ell\gamma^{\ell(1-\alpha/2)}(n)n^{\alpha/2-1}\right) \quad (2)$$

channel uses to transmit $n_{\ell+1}$ bits per source node. The terms on the left-hand side of (2) can be understood as follows: $n_{\ell+1}$ is the number of bits to be transmitted; $a_\ell^{\alpha/2}$ is the power loss since most nodes communicate over a distance of $\Theta(a_\ell^{1/2})$; $2^{-\ell}\gamma(n)$ is the average transmit power; $n_{\ell+1}$ is the multiple-antenna gain, since we have that many transmit and receive antennas.

ii) We show that constant rate quantization of the received observations at the relays is sufficient. Hence, the $n_{\ell+1}$ bits for all sources generate $O(n_{\ell+1})$ transmissions at level $\ell+1$ of the hierarchy. Therefore

$$O(n_{\ell+1}\tau_{\ell+1}(n)) \quad (3)$$

channel uses are needed to communicate all quantized observations to their respective relay nodes.

Combining (2) and (3), accounting for the factor $2^{-\ell}\gamma(n)$ loss due to time division between relay squarelets, we obtain that the transmission time for one message bit from each source to the relay squarelet in the MAC phase at level ℓ is

$$\tau_\ell^{\mathsf{MAC}}(n) = O\left(2^\ell\gamma^{1+\ell(1-\alpha/2)}(n)n^{\alpha/2-1} + \tau_{\ell+1}(n)\right). \quad (4)$$

Next, we compute the number of channel uses per message bit received by the destination nodes in the BC phase. Similar to the MAC phase, each of the $n_{\ell+1}$ relay nodes has $n_{\ell+1}$ message bits out of which one bit is to be transmitted to each of the $n_{\ell+1}$ destination nodes. Since there are $n_{\ell+1}$ relay nodes, each destination node receives $n_{\ell+1}$ message bits. As before the required transmission time has two components.

i) Transmission of the encoded and quantized message bits from each of the $n_{\ell+1}$ relay nodes to all other relay nodes at level $\ell+1$ of the hierarchy. We show that each message bit results in $O((\ell+1)\log n)$ quantized bits. Therefore, $O(n_{\ell+1}(\ell+1)\log n)$ bits need to be transmitted from each relay node. This requires

$$O(n_{\ell+1}(\ell+1)\log(n)\tau_{\ell+1}(n)) \quad (5)$$

channel uses.

ii) Transmission of $n_{\ell+1}$ message bits from the relay nodes to each destination node. As before, we use bursty transmission with an average power constraint of $\Theta(2^{-\ell}\gamma(n))$ during the fraction $\Theta(2^\ell\gamma^{-1}(n))$ of time each relay squarelet is active (this satisfies the overall

average power constraint of 1). Using this bursty strategy requires

$$O\left(n_{\ell+1}\frac{a_\ell^{\alpha/2}}{2^{-\ell}\gamma(n)n_{\ell+1}}\right)$$
$$= O\left(n_{\ell+1}4^\ell\gamma^{\ell(1-\alpha/2)}(n)n^{\alpha/2-1}\right) \quad (6)$$

channel uses for transmission of $n_{\ell+1}$ bits per destination node. As in the MAC phase, $n_{\ell+1}$ on the left-hand side of (6) can be understood as the number of bits to be transmitted, $a_\ell^{\alpha/2}$ as the power loss for communicating over distance $\Theta(a_\ell^{1/2})$, $2^{-\ell}\gamma(n)$ as the average transmit power, and $n_{\ell+1}$ as the multiple-antenna gain.

Combining (5) and (6), accounting for a factor $2^{-\ell}\gamma(n)$ loss due to time division between relay squarelets, the transmission time for one message bit from the relays to each destination node in the BC phase at level ℓ is

$$\tau_\ell^{\mathsf{BC}}(n) = O\Big(2^\ell\gamma^{1+\ell(1-\alpha/2)}(n)n^{\alpha/2-1}$$
$$+ (\ell+1)\log(n)\tau_{\ell+1}(n)\Big). \quad (7)$$

From (4) and (7), we obtain the following recursion:

$$\tau_\ell(n)$$
$$= \tau_\ell^{\mathsf{MAC}}(n) + \tau_\ell^{\mathsf{BC}}(n)$$
$$= O\left(2^\ell\gamma^{\ell(1-\alpha/2)+1}(n)n^{\alpha/2-1} + (\ell+1)\log(n)\tau_{\ell+1}(n)\right)$$
$$= O(2^L\gamma(n)n^{\alpha/2-1} + L\log(n)\tau_{\ell+1}(n)) \quad (8)$$

where we have used $\alpha > 2$. This recursion holds for all $0 \leq \ell < L$. At level L, we use TDMA among n_L nodes in region $A(a_L)$ with a permutation traffic matrix $\lambda(n_L)$. Each of the n_L source–destination pairs uses the wireless channel for $1/n_L$ fraction of the time at power $O(n_L)$, satisfying the average power constraint. Assuming the received power is less than 1 for all n (so that we operate in the power limited regime), we can achieve a rate of at least $\Omega(a_L^{-\alpha/2})$ between any source–destination pair. Equivalently

$$\tau_L(n) = O\left(a_L^{\alpha/2}\right)$$
$$= O\left(n^{\alpha/2}\gamma^{-L\alpha/2}(n)\right)$$
$$= O\left(n^{\alpha/2}\gamma^{-L}(n)\right). \quad (9)$$

Combining (8) and (9), we have

$$\tau_0(n)$$
$$= O(n^{\alpha/2-1}2^L\gamma(n) + L\log(n)\tau_1(n))$$
$$= \cdots$$
$$= O(n^{\alpha/2-1}(L\log(n))^L 2^L\gamma(n) + (L\log(n))^L\tau_L(n))$$
$$= O(n^{\alpha/2-1}(L\log(n))^L(2^L\gamma(n) + n\gamma^{-L}(n))). \quad (10)$$

The term

$$(L\log(n))^L(2^L\gamma(n) + n\gamma^{-L}(n))$$

is the "loss" factor over the desired order $n^{\alpha/2-1}$ scaling, and we now choose the branching factor $\gamma(n)$ and the hierarchy depth $L \triangleq L(n)$ to make it small. Fix a $\delta \in (0, 1/2)$ and set

$$L(n) \triangleq \log^{1/2-\delta}(n)$$
$$\gamma(n) \triangleq n^{1/(L(n)+1)}.$$

With this

$$(L(n)\log(n))^{L(n)} \leq n^{2\log^{-1/2-\delta}(n)\log\log(n)}$$
$$2^{L(n)}\gamma(n) \leq n^{\log^{-1/2-\delta}(n)+\log^{\delta-1/2}(n)}$$
$$n\gamma^{-L(n)}(n) \leq n^{\log^{\delta-1/2}(n)}.$$

Since $\delta > 0$, the $n^{\log^{\delta-1/2}(n)}$ term dominates in (10), and we obtain

$$\tau_0(n) \leq \tilde{b}(n)n^{\alpha/2-1}$$

where

$$\tilde{b}(n) \leq n^{O(\log^{\delta-1/2}(n))}.$$

Hence, the per-node rate of the hierarchical relaying scheme is lower bounded as

$$\rho^{\mathsf{HR}}(n) = 1/\tau_0(n) \geq b(n)n^{1-\alpha/2}$$

with

$$b(n) \geq n^{-O(\log^{\delta-1/2}(n))}.$$

Note that to minimize the loss term, we should choose $\delta > 0$ to be small.

V. COOPERATIVE MULTIHOP SCHEME

In this section, we provide a brief description of the cooperative multihop scheme. The details of the construction and the analysis of its performance can be found in Section X.

Recall that a node placement $V(n)$ is μ-regular at resolution $d(n)$ if every square $[id(n), (i+1)d(n)] \times [jd(n), (j+1)d(n)]$ for some $i, j \in \mathbb{N}$ contains at least $\mu d^2(n)$ nodes. Given such a node placement $V(n)$, divide it into squares of sidelength $d(n)$. Consider four adjacent squares, combined into a bigger square of sidelength $2d(n)$. By the regularity assumption on $V(n)$, this bigger square contains at least $4\mu d^2(n)$ nodes. Hence, we can apply the hierarchical relaying scheme introduced in the last section to support any permutation traffic within this bigger square at a per-node rate of

$$b(n)(d^2(n))^{1-\alpha/2} = b(n)d^{2-\alpha}(n)$$

where $b(n)$ is essentially of order $n^{-\log^{-1/2}(n)}$. By properly choosing the permutation traffic matrices within every possible such bigger square of sidelength $2d(n)$, this creates an equivalent communication graph with $n/d^2(n)$ nodes each corresponding to a square of sidelength $d(n)$ in $A(n)$, and with edges between nodes corresponding to neighboring squares. With the above communication procedure and appropriate spatial reuse, each such edge has a capacity of

$$d^2(n)b(n)d^{2-\alpha}(n) = b(n)d^{4-\alpha}(n).$$

The resulting communication graph is depicted in Fig. 4.

Now, to send a message from a source node in $V(n)$ to its destination node, we first locate the squares of sidelength $d(n)$ they are located in. We then route the message over the edges of the communication graph constructed above in a multihop fashion.
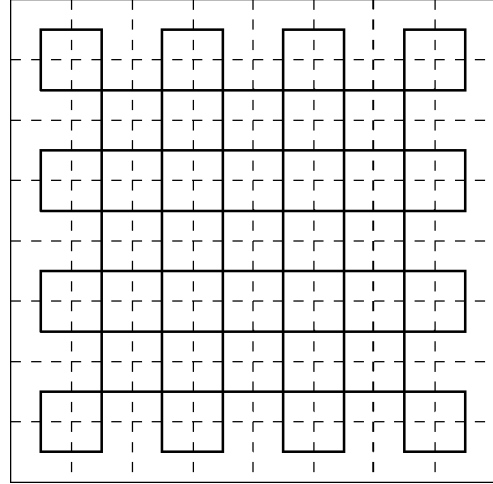

Fig. 4. Communication graph (in bold) resulting from the construction of the cooperative multihop scheme. The entire square has sidelength $\sqrt{n}$, and the dashed squares have sidelength $d(n)$. Each (bold) edge in the communication graph corresponds to using the hierarchical relaying scheme between the nodes in the adjacent squares of sidelength $d(n)$.

By the construction of the communication graph, each such edge is implemented using the hierarchical relaying scheme. In other words, we perform multihop communication over distance $\sqrt{n}$ with hop length $d(n)$, and each such hop is implemented using hierarchical relaying over distance $d(n)$. Since each edge in the communication graph has a capacity of $b(n)d^{4-\alpha}(n)$ and has to support roughly $n^{1/2}d(n)$ source–destination pairs, we obtain a per-node rate of

$$\rho^{\mathsf{CMH}}(n) \geq b(n)d^{4-\alpha}(n)n^{-1/2}d^{-1}(n)$$
$$= b(n)d^{3-\alpha}(n)n^{-1/2}$$

per source–destination pair.

VI. ANALYSIS OF THE HIERARCHICAL RELAYING SCHEME

In this section, we analyze in detail the hierarchical relaying scheme. Throughout Sections VI-A–VI-C, we consider communication at level $\ell, 0 \leq \ell < L = L(n)$, of the hierarchy. All constants K_i are independent of ℓ.

Recall that at level ℓ, we have a square region $A(a_\ell)$ of area

$$a_\ell \triangleq \frac{n}{\gamma^\ell(n)}$$

containing

$$n_\ell \triangleq \frac{n}{2^\ell \gamma^\ell(n)}$$

nodes $V(n_\ell)$. We divide $A(a_\ell)$ into $\gamma(n)$ squarelets of area $a_{\ell+1}$. Recall that a squarelet of area $a_{\ell+1}$ in level ℓ of the hierarchy is called dense if it contains at least $n_{\ell+1}$ nodes. We impose a power constraint of $P_\ell(n) = \Theta(2^{-\ell}\gamma(n))$ during the time any particular relay squarelet is active. Since we time share between $\Theta(2^{-\ell}\gamma(n))$ relay squarelets, this satisfies the overall average power constraint (by choosing constants appropriately).

Since other regions of area a_ℓ are active at the same time as the one under consideration, we have to deal with interference. To this end, we consider a slightly more general noise model that includes the experienced interference at the relay squarelets.

More precisely, we assume that, for all $u \in V(n_\ell)$, the additive noise term $\{z_u[t]\}_t$ is independent of the signal $\{x_u[t]\}_t$ and of the channel gains $\{h_{u,v}[t]\}_{v,t}$; that the noise term is stationary and ergodic across time t, but with arbitrary dependence across nodes u; and that the noise has zero mean and bounded power N_0 independent of n. Note that we do not require the additive noise term to be Gaussian. In the above, N_0 accounts for both noise (which has power 1 in the original model), as well as interference. We show in Section VII that these assumptions are valid.

Recall the following choice of $\gamma(n)$ and $L(n)$:
$$L(n) \triangleq \log^{1/2-\delta}(n)$$
$$\gamma(n) \triangleq n^{1/(L(n)+1)} \tag{11}$$
with $\delta \in (0, 1/2)$ independent of n. This choice satisfies
$$\gamma(n) \le \gamma(\tilde{n}), \qquad \text{if } n \le \tilde{n}$$
$$\gamma^{L(n)}(n) \le n, \qquad \text{for all } n$$
$$2^{-L(n)}\gamma(n) \to \infty, \qquad \text{as } n \to \infty. \tag{12}$$

The first condition in (12) implies that the number of squarelets $\gamma(n)$ we divide $A(n)$ into increases in n. The second condition implies the squarelet area $a_{L(n)}$ at the last level of the hierarchy is bigger than 1. As we will see, the third condition implies that the number of dense squarelets at the last level (and hence at every level) grows unbounded as $n \to \infty$ (see Lemma 6).

Throughout Section VI, we consider the fast-fading channel model. Slow fading is discussed in Section VII-B.

A. Setting Up Relays

The first lemma states that the minimum-separation requirement $r_{\min} \in (0, 1)$ implies that a constant fraction of squarelets must be dense. We point out that this is the only consequence of the minimum-separation requirement used to prove Theorem 1. Thus, Theorem 1 remains valid if we just assume that Lemma 6 holds directly. See also Section XII-D for further details.

Lemma 6: For any $V(n_\ell) \subset A(a_\ell)$ with $|V(n_\ell)| \ge n_\ell$ and with minimum separation $r_{\min} \in (0, 1)$, each of its squarelets of area $a_{\ell+1}$ contains at most $K_1 a_\ell / \gamma(n)$ nodes, and there are at least $K_2 2^{-\ell} \gamma(n)$ dense squarelets.

Proof: Put a circle of radius $r_{\min}/2$ around each node. By the minimum-separation requirement, these circles do not intersect. Each node covers an area of $\pi r_{\min}^2/4$. Increasing the side-length of each squarelet by $r_{\min}$, this provides a total area of
$$(\sqrt{a_\ell/\gamma(n)} + r_{\min})^2 \le \frac{a_\ell}{\gamma(n)}(1 + r_{\min})^2$$
in which the circles around these nodes are packed. Here we have used that $\gamma^{\ell+1}(n) \le n$ by (12), and therefore
$$\gamma(n) \le n/\gamma^\ell(n) = a_\ell.$$

Hence, there can be at most $K_1 a_\ell / \gamma(n)$ nodes per squarelet with
$$K_1 \triangleq 4\frac{(1 + r_{\min})^2}{\pi r_{\min}^2}.$$

Note that, since $r_{\min} < 1$, we have $K_1 > 1$.

Let $d(n_\ell)$ be the number of dense squarelets in $A(a_\ell)$, and therefore, $\gamma(n) - d(n_\ell)$ is the number of squarelets that

are not dense. By the argument in the last paragraph, each dense squarelet contains at most $K_1 a_\ell / \gamma(n)$ nodes, and those squarelets that are not dense contain less than $n_{\ell+1}$ nodes by the definition of dense squarelets. Hence, $d(n_\ell)$ must satisfy
$$d(n_\ell)K_1 a_\ell/\gamma(n) + (\gamma(n) - d(n_\ell))n_{\ell+1} \ge |V(n_\ell)| \ge n_\ell.$$
Thus, using $a_\ell = 2^\ell n_\ell, n_{\ell+1} = n_\ell/2\gamma(n)$, we have
$$d(n_\ell)K_1 2^\ell + (\gamma(n) - d(n_\ell))/2 \ge \gamma(n).$$

As $K_1 2^\ell > 1$, this yields
$$d(n_\ell) \ge \frac{1 - 1/2}{K_1 2^\ell - 1/2}\gamma(n)$$
$$\ge \frac{2^{-\ell}}{2K_1}\gamma(n)$$
$$= K_2 2^{-\ell}\gamma(n)$$
with
$$K_2 \triangleq \frac{1}{2K_1}. \qquad \square$$

Consider $V(n_\ell) \subset A(a_\ell)$ with $|V(n_\ell)| \ge n_\ell$, and choose arbitrary $K_2 2^{-\ell}\gamma(n)$ dense squarelets of area $a_{\ell+1}$ (as guaranteed by Lemma 6). Call those squarelets $\{A_k(a_{\ell+1})\}_{k=1}^{K_2 2^{-\ell}\gamma(n)}$. For each source–destination pair, we now select one such dense squarelet to relay traffic over. To avoid bottlenecks, this selection has to be done such that all relay squarelets carry approximately the same amount of traffic. Moreover, for technical reasons, the distances from the source and the destination to the relay squarelet cannot be too small.

Formally, the selection of relay squarelets can be described by the *schedules* $S \in \{0,1\}^{n_\ell \times K_2 2^{-\ell}\gamma(n)}$ with $s_{u,k} = 1$ if source node u relays traffic over dense squarelet k, and $\tilde{S} \in \{0,1\}^{K_2 2^{-\ell}\gamma(n) \times n_\ell}$ with $\tilde{s}_{k,w} = 1$ if destination node w receives traffic from dense squarelet k. With slight abuse of notation, let $r_{u,A_k(a_{\ell+1})}$ be the distance between node $u \in V(n_\ell)$ and the closest point in $A_k(a_{\ell+1})$, i.e.,
$$r_{u,A_k(a_{\ell+1})} \triangleq \min_{v \in A_k(a_{\ell+1})} r_{u,v}. \tag{13}$$

Define the sets
$$\mathcal{S}(n_\ell) \triangleq \left\{ S \in \{0,1\}^{n_\ell \times K_2 2^{-\ell}\gamma(n)} : \right.$$
$$0 \le \sum_{u=1}^{n_\ell} s_{u,k} \le n_{\ell+1} \, \forall k,$$
$$0 \le \sum_{k=1}^{K_2 2^{-\ell}\gamma(n)} s_{u,k} \le 1 \, \forall u,$$
$$\left. s_{u,k} = 1 \to r_{u,A_k(a_{\ell+1})} \ge \sqrt{2a_{\ell+1}} \, \forall u, k \right\} \tag{14}$$
and
$$\tilde{\mathcal{S}}(n_\ell) \triangleq \left\{ \tilde{S} \in \{0,1\}^{K_2 2^{-\ell}\gamma(n) \times n_\ell} : \tilde{S}^T \in \mathcal{S}(n_\ell) \right\}.$$

The sets $\mathcal{S}(n_\ell)$ and $\tilde{\mathcal{S}}(n_\ell)$ are the collection of schedules satisfying the conditions mentioned in the last paragraph. More precisely, the first condition in (14) ensures that at most $n_{\ell+1}$ source–destination pairs relay over the same dense squarelet,

the second condition ensures that each source–destination pair chooses at most one relay squarelet, and the third condition ensures that sources and destinations are at least at distance $\sqrt{2a_{\ell+1}}$ from the chosen relay squarelet.

Next, we prove that any node placement that satisfies Lemma 6 allows for a decomposition of any permutation traffic matrix $\lambda(n_\ell)$ into a small number of schedules belonging to $\mathcal{S}(n_\ell)$ and $\tilde{\mathcal{S}}(n_\ell)$.

Lemma 7: There exist K_3 such that for all n large enough (independent of ℓ), and every permutation traffic matrix $\lambda(n_\ell) \in \{0,1\}^{n_\ell \times n_\ell}$, we can find $K_3 2^\ell$ schedules $\{S^{(i)}(n_\ell)\}_{i=1}^{K_3 2^\ell} \subset \mathcal{S}(n_\ell), \{\tilde{S}^{(i)}(n_\ell)\}_{i=1}^{K_3 2^\ell} \subset \tilde{\mathcal{S}}(n_\ell)$ satisfying

$$\lambda(n_\ell) = \sum_{i=1}^{K_3 2^\ell} S^{(i)}(n_\ell) \tilde{S}^{(i)}(n_\ell).$$

Proof: Pick an arbitrary source–destination pair in $\lambda(n_\ell)$, and consider the squarelets containing the source and the destination node. Since each squarelet has side length $\sqrt{a_{\ell+1}}$, there are at most 50 squarelets at distance less than $\sqrt{2a_{\ell+1}}$ from either of those two squarelets. As $2^{-L(n)}\gamma(n) \to \infty$ as $n \to \infty$ by (12), there exists K (independent of ℓ) such that, for $n \geq K$, we have $50 \leq K_2 2^{-\ell-1}\gamma(n)$. Since there are at least $K_2 2^{-\ell}\gamma(n)$ dense squarelets by Lemma 6, there must exist at least $K_2 2^{-\ell-1}\gamma(n)$ dense squarelets that are at distance at least $\sqrt{2a_{\ell+1}}$ from both squarelets containing the source and the destination node.

In order to construct a decomposition of $\lambda(n_\ell)$, we use the following procedure. Sequentially, each of the n_ℓ source–destination pairs chooses one of the (at least) $K_2 2^{-\ell-1}\gamma(n)$ dense squarelets at distance at least $\sqrt{2a_{\ell+1}}$ that has not already been chosen by $n_{\ell+1}$ other pairs. If any source–destination pair cannot select such a squarelet, then stop the procedure and use the source–destination pairs matched with dense squarelets so far to define matrices $S^{(1)}(n_\ell)$ and $\tilde{S}^{(1)}(n_\ell)$. Now, remove all the matched source–destination pairs, forget that dense squarelets were matched to any source–destination pair, and redo the above procedure, going through the remaining source–destination pairs.

Let

$$K_3 \triangleq 4/K_2.$$

We claim that by repeating this process of generating matrices $S^{(i)}(n_\ell)$ and $\tilde{S}^{(i)}(n_\ell)$, we can match all source–destination pairs to some dense squarelet with at most $K_3 2^\ell$ such matrices. Indeed, a new pair of matrices is generated only when a source–destination pair cannot be matched to any of its available (at least) $K_2 2^{-\ell-1}\gamma(n)$ dense squarelets. If this happens, all these dense squarelets are matched by $n_{\ell+1} = n_\ell/2\gamma(n)$ pairs. Hence, at least $K_2 2^{-\ell-2}n_\ell$ source–destination pairs are matched in each "round." Since there are n_ℓ total pairs, we need at most

$$\frac{n_\ell}{K_2 2^{-\ell-2}n_\ell} = K_3 2^\ell$$

matrices $S^{(i)}(n_\ell)$ and $\tilde{S}^{(i)}(n_\ell)$. □

For a permutation traffic matrix $\lambda(n_\ell)$, communication proceeds as follows. Write

$$\lambda(n_\ell) = \sum_{i=1}^{K_3 2^\ell} S^{(i)}(n_\ell) \tilde{S}^{(i)}(n_\ell)$$

as in Lemma 7. Split time into $K_3 2^\ell$ equal length time slots. In slot i, we use $S^{(i)}(n_\ell)\tilde{S}^{(i)}(n_\ell)$ as our traffic matrix. Consider without loss of generality $i = 1$ in the following. Write

$$S^{(1)}(n_\ell)\tilde{S}^{(1)}(n_\ell) = \sum_{k=1}^{K_2 2^{-\ell}\gamma(n)} S^{(1,k)}(n_{\ell+1})\tilde{S}^{(1,k)}(n_{\ell+1})$$

where $S^{(1,k)}(n_{\ell+1})\tilde{S}^{(1,k)}(n_{\ell+1})$ is the traffic relayed over the dense squarelet $A_k(a_{\ell+1})$. We time share between the schedules for $k \in \{1, \ldots, K_2 2^{-\ell}\gamma(n)\}$. Consider now any such k. In the worst case, there are exactly $n_{\ell+1}$ communication pairs to be relayed over $A_k(a_{\ell+1})$, and the relay squarelet $A_k(a_{\ell+1})$ contains exactly $n_{\ell+1}$ nodes. We will assume this worst case in the following.

We focus on the transmission according to the traffic matrix $S^{(1,1)}(n_{\ell+1})\tilde{S}^{(1,1)}(n_{\ell+1})$. Let $V(n_{\ell+1})$ be the nodes in $A_1(a_{\ell+1})$, and let $U(n_{\ell+1})$ and $W(n_{\ell+1})$ be the source and destination nodes of $S^{(1,1)}(n_{\ell+1})\tilde{S}^{(1,1)}(n_{\ell+1})$, respectively. In other words, the source nodes $U(n_{\ell+1})$ communicate to their respective destination nodes $W(n_{\ell+1})$ using the nodes $V(n_{\ell+1})$ as relays.

B. Multiple-Access Phase

Each source node in $U(n_{\ell+1})$ splits its message into $n_{\ell+1}$ equal length parts. Part j at every node $u \in U(n_{\ell+1})$ is to be relayed over the jth node in $V(n_{\ell+1})$. Each part is separately encoded at the source and separately decoded at the destination. After the source nodes are done transmitting their messages, the nodes in the relay squarelet quantize their (sampled) observations corresponding to part j and communicate the quantized values to the jth node in the relay squarelet. This node then decodes the jth message parts of all source nodes. Note that this induces a uniform traffic pattern between the nodes in the relay squarelet, i.e., every node needs to transmit quantized observations to every other node. While this traffic pattern does not correspond to a permutation traffic matrix, it can be written as a sum of $n_{\ell+1}$ permutation traffic matrices. A fraction $1/n_{\ell+1}$ of the traffic within the relay squarelet is transmitted according to each of these permutation traffic matrices. This setup is depicted in Fig. 2 in Section IV-A.

Assuming for the moment that we have a scheme to send the quantized observations to the dedicated node in the relay squarelet, the traffic matrix $S^{(1,1)}(n_{\ell+1})$ between $U(n_{\ell+1})$ and $V(n_{\ell+1})$ describes then a MAC with $n_{\ell+1}$ transmitters, each with one antenna, and one receiver with $n_{\ell+1}$ antennas. We call this the *MAC induced by* $S^{(1,1)}(n_{\ell+1})$ in the following. Before we analyze the rate achievable over this induced MAC, we need an auxiliary result on quantized channels.

Consider the quantized channel in Fig. 5. Here, f is the channel encoder, φ is the channel decoder, and $\{q_k\}_{k=1}^m$ are quantizers. All these have to be chosen. $P_{\boldsymbol{y}|x}$ and $P_{\hat{x}|\hat{\boldsymbol{y}}}$, on the other hand, represent fixed stationary ergodic channels
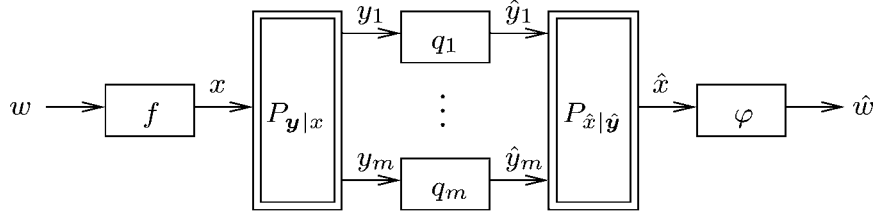
Fig. 5. Sketch of the quantized channel: f and φ are the channel encoder and decoder, respectively; $\{q_k\}_{k=1}^m$ are quantizers; and $P_{\boldsymbol{y}\,|\,x}$ and $P_{\hat{x}\,|\,\hat{\boldsymbol{y}}}$ represent stationary ergodic channels with the indicated marginal distributions.

with the indicated marginal distributions. We call R the rate of the channel code (f, φ) and $\{R_k\}_{k=1}^m$ the rates of quantizers $\{q_k\}_{k=1}^m$.

Lemma 8: If there exist distributions P_x and $\{P_{\hat{y}_k\,|\,y_k}\}_{k=1}^m$ such that $R < I(x; \hat{x})$ and $R_k > I(y_k; \hat{y}_k)$, $\forall k$, then $(R, \{R_k\}_{k=1}^m)$ is achievable over the quantized channel.

Proof: The proof follows from a simple extension of Theorem 1 in Appendix II of [8]. $\quad\square$

Lemma 9: Let the additive noise $\{z_v\}_{v \in V(n_{\ell+1})}$ be uncorrelated (over v). For the MAC induced by $S^{(1,1)}(n_{\ell+1})$ with per-node average power constraint $P_\ell(n) \le n_{\ell+1}^{-1} a_\ell^{\alpha/2}$, a rate of

$$\rho_\ell^{\text{MAC}}(n) \ge K_4 P_\ell(n) n_{\ell+1} a_\ell^{-\alpha/2}$$

per source node is achievable, and the number of bits required at each relay node to quantize the observations is at most K_5 bits per $n_{\ell+1}$ total message bits[5] sent by the source nodes.

Proof: The source nodes send signals with a power of (essentially) $n_{\ell+1}^{-1} a_\ell^{\alpha/2}$ for a fraction $P_\ell(n) n_{\ell+1} a_\ell^{-\alpha/2} \le 1$ of time and are silent for the remaining time. To ensure that interference is uniform, the time slots during which the nodes send signals are chosen randomly as follows. Generate independently for each region $A(a_\ell)$ a Bernoulli process $\{B[t]\}_{t \in \mathbb{N}}$ with parameter $P_\ell(n) n_{\ell+1} a_\ell^{-\alpha/2}/(1+\eta) \le 1$ for some small $\eta > 0$. The nodes in $A(a_\ell)$ are active whenever $B[t] = 1$ and remain silent otherwise. Since the blocklength of the codes used is assumed to be large, this satisfies the average power constraint of $P_\ell(n)$ with high probability for any $\eta > 0$. Since we are interested only in the scaling of capacity, we ignore the additional $1/(1 + \eta)$ term in the following to simplify notation. Clearly, we only need to consider the fraction of time during which $B[t] = 1$.

Let $\boldsymbol{y}$ be the received vector at the relay squarelet, and let $\hat{\boldsymbol{y}}$ be the (componentwise) quantized observations. We use a matched filter at the relay squarelet, i.e.,

$$\hat{x}_u = \frac{\boldsymbol{h}_u^\dagger}{\|\boldsymbol{h}_u\|} \hat{\boldsymbol{y}}$$

where column vector $\boldsymbol{h}_u = \{h_{u,v}\}_{v \in V(n_{\ell+1})}$ are the channel gains between node $u \in U(n_{\ell+1})$ and the nodes in the relay squarelet $V(n_{\ell+1})$. The use of a matched filter is possible since we assume full CSI is available at all the nodes.

We now use Lemma 8 to show that we can design quantizers $\{q_v\}_{v \in V(n_{\ell+1})}$ of constant rate and achieve a per-node communication rate of at least $K_4 P_\ell(n) n_{\ell+1} a_\ell^{-\alpha/2}$. The first channel

in Lemma 8 (see Fig. 5) will correspond to the wireless channel between a source node u and its relay squarelet $V(n_{\ell+1})$. The second "channel" in Lemma 8 will correspond to the matched filter used at the relay squarelet. To apply Lemma 8, we need to find a distribution for x_u and for $\hat{y}_v \,|\, y_v$. Define

$$\tilde{r}_u \triangleq r_{u, A_1(a_{\ell+1})}/\sqrt{2a_\ell} \le 1$$

with $r_{u, A_1(a_{\ell+1})}$ as in (13), to be the normalized distance of the source node $u \in U(n_{\ell+1})$ to the relay squarelet $A_1(a_{\ell+1})$. For each $u \in U(n_{\ell+1})$, let $x_u \sim \mathcal{N}_\mathbb{C}(0, \tilde{r}_u^\alpha n_{\ell+1}^{-1} a_\ell^{\alpha/2})$ independent of $x_{\tilde{u}}$ for $u \ne \tilde{u}$, and let $\hat{y}_{\tilde{v}} = y_v + \tilde{z}_v$ for $\tilde{z}_v \sim \mathcal{N}_\mathbb{C}(0, \Delta^2)$ independent of $\boldsymbol{y}$ and for some $\Delta^2 > 0$. Note that the channel input x_u has power that depends on the normalized distance $\tilde{r}_u$ (i.e., only nodes $u \in U(n_{\ell+1})$ that are at maximal distance $\sqrt{2a_\ell}$ from the relay squarelet transmit at full available power). This is to ensure that all signals are received at roughly the same strength by the relays.

We proceed by computing the mutual informations $I(y_v; \hat{y}_v|\{h_{\tilde{u},\tilde{v}}\})$ and $I(x_u; \hat{x}_u|\{h_{\tilde{u},\tilde{v}}\})$ as required in Lemma 8 (the conditioning on $\{h_{\tilde{u},\tilde{v}}\}$ being due to the availability of full CSI). Note first that by construction of $S^{(1,1)}(n_{\ell+1})$ [see (14)], we have for $u \in U(n_{\ell+1})$ and $v \in V(n_{\ell+1})$

$$r_{u, A_1(a_{\ell+1})} \le r_{u,v} \le 2r_{u, A_1(a_{\ell+1})}$$

and, hence

$$\frac{1}{2\sqrt{2a_\ell}} \le \frac{\tilde{r}_u}{r_{u,v}} \le \frac{1}{\sqrt{2a_\ell}}. \tag{15}$$

From this, and since $|h_{u,v}|^2 = r_{u,v}^{-\alpha}$, we obtain

$$2^{-3\alpha/2} a_\ell^{-\alpha/2} \le |h_{u,v}|^2 \tilde{r}_u^\alpha \le 2^{-\alpha/2} a_\ell^{-\alpha/2}$$
$$2^{-3\alpha/2} n_{\ell+1} a_\ell^{-\alpha/2} \le \|\boldsymbol{h}_u\|^2 \tilde{r}_u^\alpha \le 2^{-\alpha/2} n_{\ell+1} a_\ell^{-\alpha/2}. \tag{16}$$

We start by computing $I(y_v; \hat{y}_v|\{h_{\tilde{u},\tilde{v}}\})$. We have

$$\hat{y}_v = \sum_{u \in U(n_{\ell+1})} h_{u,v} x_u + z_v + z_{\tilde{v}}$$

and, hence, $\hat{y}_v$ has mean zero and variance

$$\mathbb{E}(|\hat{y}_v|^2) = \sum_{u \in U(n_{\ell+1})} |h_{u,v}|^2 \tilde{r}_u^\alpha n_{\ell+1}^{-1} a_\ell^{\alpha/2} + N_0 + \Delta^2$$
$$\le n_{\ell+1} 2^{-\alpha/2} a_\ell^{-\alpha/2} n_{\ell+1}^{-1} a_\ell^{\alpha/2} + N_0 + \Delta^2$$
$$= 2^{-\alpha/2} + N_0 + \Delta^2$$

---

[5]Total message bits refer to the sum of all message bits transmitted by the $n_{\ell+1}$ source nodes.

where we have used (16). Hence

$$
\begin{aligned}
I(y_v; \hat{y}_v | \{h_{\tilde{u},\tilde{v}}\}) &= h(\hat{y}_v | \{h_{\tilde{u},\tilde{v}}\}) - h(\hat{y}_v | y_v, \{h_{\tilde{u},\tilde{v}}\}) \\
&\leq \log(2\pi e \mathbb{E}(|\hat{y}_v|^2)) - \log(2\pi e \Delta^2) \\
&\leq \log(2\pi e(2^{-\alpha/2} + N_0 + \Delta^2)) - \log(2\pi e \Delta^2) \\
&= \log\left(1 + \frac{2^{-\alpha/2} + N_0}{\Delta^2}\right).
\end{aligned} \tag{17}
$$

We now compute $I(x_u; \hat{x}_u | \{h_{\tilde{u},\tilde{v}}\})$. We have

$$
\hat{x}_u = \|\boldsymbol{h}_u\| x_u + \sum_{\tilde{u} \in U(n_{\ell+1}) \setminus \{u\}} \frac{\boldsymbol{h}_u^\dagger \boldsymbol{h}_{\tilde{u}}}{\|\boldsymbol{h}_u\|} x_{\tilde{u}} + \frac{\boldsymbol{h}_u^\dagger}{\|\boldsymbol{h}_u\|}(\boldsymbol{z} + \tilde{\boldsymbol{z}}).
$$

Conditioned on $\{\boldsymbol{h}_{\tilde{u}}\}_{\tilde{u} \in U(n_{\ell+1})}$

$$
\|\boldsymbol{h}_u\| x_u \sim \mathcal{N}_{\mathbb{C}}\left(0, \|\boldsymbol{h}_u\|^2 \tilde{r}_u^\alpha n_{\ell+1}^{-1} a_\ell^{\alpha/2}\right)
$$

and

$$
\begin{aligned}
&\mathbb{E}\left(\left|\sum_{\tilde{u} \in U(n_{\ell+1}) \setminus \{u\}} \frac{\boldsymbol{h}_u^\dagger \boldsymbol{h}_{\tilde{u}}}{\|\boldsymbol{h}_u\|} x_{\tilde{u}} + \frac{\boldsymbol{h}_u^\dagger}{\|\boldsymbol{h}_u\|}(\boldsymbol{z} + \tilde{\boldsymbol{z}})\right|^2 \Bigg| \{\boldsymbol{h}_{\tilde{u}}\}\right) \\
&= n_{\ell+1}^{-1} a_\ell^{\alpha/2} \sum_{\tilde{u} \in U(n_{\ell+1}) \setminus \{u\}} \tilde{r}_{\tilde{u}}^\alpha \frac{|\boldsymbol{h}_u^\dagger \boldsymbol{h}_{\tilde{u}}|^2}{\|\boldsymbol{h}_u\|^2} + N_0 + \Delta^2
\end{aligned}
$$

where we have used the assumption that $\{z_v\}_{v \in V(n_{\ell+1})}$ are uncorrelated in the second line. Using (16), this is, in turn, upper bounded by

$$
2^{3\alpha/2} \tilde{r}_u^\alpha n_{\ell+1}^{-2} a_\ell^\alpha \sum_{\tilde{u} \in U(n_{\ell+1}) \setminus \{u\}} \tilde{r}_{\tilde{u}}^\alpha |\boldsymbol{h}_u^\dagger \boldsymbol{h}_{\tilde{u}}|^2 + N_0 + \Delta^2.
$$

Similarly, we can lower bound the received signal power as

$$
\mathbb{E}(\|\boldsymbol{h}_u\|^2 |x_u|^2) \geq 2^{-3\alpha/2}.
$$

Since Gaussian noise is the worst additive noise under a power constraint [15], and applying Jensen's inequality to the convex function $\log(1 + 1/x)$, we obtain (18), shown at the bottom of the page. We have for $u \neq \tilde{u}$

$$
\begin{aligned}
\mathbb{E}\left(|\boldsymbol{h}_u^\dagger \boldsymbol{h}_{\tilde{u}}|^2\right) &= \mathbb{E}(\boldsymbol{h}_u^\dagger \boldsymbol{h}_{\tilde{u}} \boldsymbol{h}_{\tilde{u}}^\dagger \boldsymbol{h}_u) \\
&= \sum_{v \in V(n_{\ell+1})} |h_{u,v}|^2 |h_{\tilde{u},v}|^2 \\
&= \sum_{v \in V(n_{\ell+1})} r_{u,v}^{-\alpha} r_{\tilde{u},v}^{-\alpha}
\end{aligned} \tag{19}
$$

and, hence, using (15)

$$
\begin{aligned}
&\mathbb{E}\left(\tilde{r}_u^\alpha \sum_{\tilde{u} \in U(n_{\ell+1}) \setminus \{u\}} \tilde{r}_{\tilde{u}}^\alpha |\boldsymbol{h}_u^\dagger \boldsymbol{h}_{\tilde{u}}|^2\right) \\
&= \sum_{\tilde{u} \in U(n_{\ell+1}) \setminus \{u\}} \sum_{v \in V(n_{\ell+1})} \tilde{r}_u^\alpha r_{u,v}^{-\alpha} \tilde{r}_{\tilde{u}}^\alpha r_{\tilde{u},v}^{-\alpha} \\
&\leq 2^{-\alpha} n_{\ell+1}^2 a_\ell^{-\alpha}.
\end{aligned}
$$

Therefore, we can continue (18) as

$$
I(x_u; \hat{x}_u | \{h_{\tilde{u},\tilde{v}}\}) \geq \frac{1}{2} \log\left(1 + \frac{2^{-3\alpha/2}}{2^{\alpha/2} + N_0 + \Delta^2}\right) \triangleq K_4. \tag{20}
$$

Using (17) and (20) in Lemma 8, and observing that we only communicate during a fraction

$$
P_\ell(n) n_{\ell+1} a_\ell^{-\alpha/2} \leq 1
$$

of time yields a per source node rate $\rho_\ell^{\mathsf{MAC}}(n)$ arbitrarily close to

$$
K_4 P_\ell(n) n_{\ell+1} a_\ell^{-\alpha/2}
$$

and a quantizer of rate arbitrarily close to

$$
\log\left(1 + \frac{2^{-\alpha/2} + N_0}{\Delta^2}\right)
$$

bits per observation at each relay node. Since by (20) mutual information $I(x_u; \hat{x}_u | \{h_{\tilde{u},\tilde{v}}\})$ is at least K_4 for every $u \in U(n_{\ell+1})$ during the fraction of time we actually communicate, this implies that there are at most $1/K_4$ observations at each relay node per $n_{\ell+1}$ total message bits. Thus, the number of bits per relay node required to quantize the observations is at most

$$
K_5 \triangleq \frac{1}{K_4} \log\left(1 + \frac{2^{-\alpha/2} + N_0}{\Delta^2}\right)
$$

bits per $n_{\ell+1}$ total message bits sent by the source nodes. $\quad\square$

C. Broadcast Phase

At the end of the MAC phase, all relay nodes $V(n_{\ell+1})$ received a part of the message sent by each source node. In the BC phase, each node in the relay squarelet encodes these messages together for $n_{\ell+1}$ transmit antennas. The encoded message is then quantized and communicated to all the nodes in the relay squarelet. These nodes then send the quantized encoded message to the destination nodes $W(n_{\ell+1})$. Note that this again

---

$$
\begin{aligned}
I(x_u; \hat{x}_u | \{h_{\tilde{u},\tilde{v}}\}) &\geq \mathbb{E}\left(\log\left(1 + \frac{2^{-3\alpha/2}}{2^{3\alpha/2} \tilde{r}_u^\alpha n_{\ell+1}^{-2} a_\ell^\alpha \sum_{\tilde{u} \in U(n_{\ell+1}) \setminus \{u\}} \tilde{r}_{\tilde{u}}^\alpha |\boldsymbol{h}_u^\dagger \boldsymbol{h}_{\tilde{u}}|^2 + N_0 + \Delta^2}\right)\right) \\
&\geq \log\left(1 + \frac{2^{-3\alpha/2}}{2^{3\alpha/2} \tilde{r}_u^\alpha n_{\ell+1}^{-2} a_\ell^\alpha \sum_{\tilde{u} \in U(n_{\ell+1}) \setminus \{u\}} \tilde{r}_{\tilde{u}}^\alpha \mathbb{E}\left(|\boldsymbol{h}_u^\dagger \boldsymbol{h}_{\tilde{u}}|^2\right) + N_0 + \Delta^2}\right)
\end{aligned} \tag{18}
$$

induces a uniform traffic pattern between the nodes in the relay squarelet, i.e., every node needs to transmit quantized encoded messages to every other node. While this traffic pattern does not correspond to a permutation traffic matrix it can be written as a sum of $n_{\ell+1}$ permutation traffic matrices. A fraction $1/n_{\ell+1}$ of the traffic within the relay squarelet is transmitted according to each of these permutation traffic matrices. This setup is depicted in Fig. 3 in Section IV-A.

Assuming for the moment that we have a scheme to send the quantized encoded messages to the corresponding nodes in the relay squarelet, the traffic matrix $\tilde{S}^{(1,1)}(n_{\ell+1})$ between $V(n_{\ell+1})$ and $W(n_{\ell+1})$ describes then a BC with one transmitter with $n_{\ell+1}$ antennas and $n_{\ell+1}$ receivers, each with one antenna. We call this the *BC induced by* $\tilde{S}^{(1,1)}(n_{\ell+1})$ in the following.

Lemma 10: For the BC induced by $\tilde{S}^{(1,1)}(n_{\ell+1})$ with per-node average power constraint $P_\ell(n) \leq n_{\ell+1}^{-1} a_\ell^{\alpha/2}$, a rate of

$$\rho_\ell^{\mathsf{BC}}(n) \geq K_6 P_\ell(n) n_{\ell+1} a_\ell^{-\alpha/2}$$

is achievable per destination node, and the number of bits required to quantize the observations is at most $K_7(\ell+1) \log(n)$ bits at each relay node per $n_{\ell+1}$ total message bits[6] received by the destination nodes.

Proof: Consider a node $v \in V(n_{\ell+1})$ in the relay squarelet, say the first one. From the MAC phase, this node received the first part of the messages of each source node $u \in U(n_{\ell+1})$. We would like to jointly encode these message parts at the relay node using transmit beamforming, and then transmit the corresponding encoded signal using all the nodes in the relay squarelet. However, this cannot be done directly, because at the encoding time, the future channel state at transmission time is unknown.

We circumvent this problem by reordering the signals to be transmitted at the relay nodes as follows. Let

$$\{\hat{\theta}_{v,w}\}_{v \in V(n_{\ell+1}), w \in W(n_{\ell+1})} \in \{0, \pi/2, \pi, 3\pi/2\}^{n_{\ell+1}^2}$$

be a "quantized" channel state. The part of the messages at node one in the relay squarelet is encoded for $n_{\ell+1}$ transmit nodes with an assumed channel gain of

$$\hat{h}_{v,w}[t] = r_{v,w}^{-\alpha/2} \exp(\sqrt{-1}\hat{\theta}_{v,w}[t])$$

where the $\{\hat{\theta}_{v,w}[t]\}_{v,w,t}$ are cycled as a function of t through all possible values in $\{0, \pi/2, \pi, 3\pi/2\}^{n_{\ell+1}^2}$. The components of the encoded messages are then quantized and each component sent to the corresponding node in the relay squarelet. Once all nodes in the relay squarelet have received the encoded message, they send in each time slot a sample of the encoded messages corresponding to the quantized channel state closest (in Euclidean distance) to the actual channel realization in that time slot. By ergodicity of $\{\theta_{u,v}[t]\}_t$, each quantized channel state is used approximately the same number of times. More precisely, as the message length grows to infinity, we can send samples of the encoded message parts a $1/(1+\eta)$ fraction of time with probability approaching 1 for any $\eta > 0$. Since we have no constraint on the encoding delay in our setup, we can

choose η arbitrarily small, and given that we are only interested in scaling laws, we will ignore this term in the following to simplify notation. Note that the destination nodes can reorder the received samples since we assume full CSI. In the following, we let $\{\hat{\theta}_{v,w}\}_{v,w}$ be the random quantized channel state induced by $\{\theta_{v,w}\}_{v,w}$ through the above procedure. Denote by $\{\hat{h}_{v,w}\}_{v,w}$ the corresponding channel gains.

As in the MAC phase, the nodes in the relay squarelet send signals at a power (essentially) $n_{\ell+1}^{-1} a_\ell^{\alpha/2}$ a fraction $P_\ell(n) n_{\ell+1} a_\ell^{-\alpha/2} \leq 1$ of time and are silent for the remaining time. To create interference at uniform power, this is done in the same randomized manner as in the MAC phase. Generate independently for each region $A(a_\ell)$ a Bernoulli process $\{B[t]\}_{t \in \mathbb{N}}$ with parameter $P_\ell(n) n_{\ell+1} a_\ell^{-\alpha/2}/(1+\eta)$ for some small $\eta > 0$. The nodes in $A(a_\ell)$ are active whenever $B[t] = 1$ and remain silent otherwise. As before, we ignore the additional $1/(1+\eta)$ term. Again we only need to consider the fraction of time during which $B[t] = 1$.

Consider the message part at a relay node for destination node $w \in W(n_{\ell+1})$. We encode this part independently; call $\tilde{x}_w$ the encoded message part. The relay node then performs transmit beamforming to construct the encoded message for all its destination nodes, i.e.,

$$\boldsymbol{x} = \sum_{w \in W(n_{\ell+1})} \frac{\hat{\boldsymbol{h}}_w^\dagger}{\|\boldsymbol{h}_w\|} \tilde{x}_w$$

where row vector $\boldsymbol{h}_w = \{h_{v,w}\}_{v \in V(n_{\ell+1})}$ contains the channel gains to node w, and where we have used $|\hat{h}_{v,w}| = |h_{v,w}|$. The relay node then quantizes the vector of encoded messages componentwise and forwards the quantized version $\hat{\boldsymbol{x}}$ to the other nodes in the relay squarelet. These nodes then send $\hat{\boldsymbol{x}}$ over the channel to the destination nodes. The received signal at destination node w is thus

$$y_w = \boldsymbol{h}_w \hat{\boldsymbol{x}} + z_w.$$

With this, we have the setup considered in Lemma 8 (with different variable names). The first "channel" in Lemma 8 (see Fig. 5) will correspond to the transmit beamforming used at the relay squarelet. The second channel in Lemma 8 will now correspond to the wireless channel between the relay squarelet $V(n_{\ell+1})$ and a destination node w. To apply Lemma 8, we need to find a distribution for $\tilde{x}_w$ and for $\hat{x}_v \mid x_v$. We also need to guarantee that $\hat{x}_v$ satisfies the power constraint at each node v in the relay squarelet. For each $w \in W(n_{\ell+1})$, let $\tilde{x}_w \sim \mathcal{N}_{\mathbb{C}}(0, K n_{\ell+1}^{-1} a_\ell^{\alpha/2})$ (for some K to be chosen later) independent of $\tilde{x}_{\tilde{w}}$ for $w \neq \tilde{w}$, and let $\hat{x}_v = x_v + \tilde{z}_v$ for $\tilde{z}_v \sim \mathcal{N}_{\mathbb{C}}(0, \Delta^2)$ independent of $\boldsymbol{x}$ and for some $\Delta^2 > 0$. We then have

$$y_w = \frac{\boldsymbol{h}_w \hat{\boldsymbol{h}}_w^\dagger}{\|\boldsymbol{h}_w\|} \tilde{x}_w + \sum_{\tilde{w} \in W(n_{\ell+1}) \setminus \{w\}} \frac{\boldsymbol{h}_w \hat{\boldsymbol{h}}_{\tilde{w}}^\dagger}{\|\boldsymbol{h}_{\tilde{w}}\|} \tilde{x}_{\tilde{w}} + \boldsymbol{h}_w \tilde{\boldsymbol{z}} + z_w.$$

We proceed by computing the mutual informations $I(x_v; \hat{x}_v \mid \{h_{\tilde{u},\tilde{v}}\})$ and $I(\tilde{x}_w; y_w \mid \{h_{\tilde{u},\tilde{v}}\})$ as required in Lemma 8 (the conditioning on $\{h_{\tilde{u},\tilde{v}}\}$ again being due to the availability of full CSI). Note first that by construction of $\tilde{S}^{(1,1)}(n_{\ell+1})$, we have for any $w \in W(n_{\ell+1})$

$$2 \min_{v \in V(n_{\ell+1})} r_{v,w} \geq \max_{v \in V(n_{\ell+1})} r_{v,w}$$

---

[6]Total message bits refers to the sum of all message bits received by the $n_{\ell+1}$ destination nodes.

and, therefore

$$\frac{|h_{v,w}|^2}{\|\boldsymbol{h}_w\|^2} \leq \frac{\left(\min_{v \in V(n_{\ell+1})} r_{v,w}\right)^{-\alpha}}{n_{\ell+1} \left(\max_{v \in V(n_{\ell+1})} r_{v,w}\right)^{-\alpha}} \leq \frac{2^\alpha}{n_{\ell+1}}. \quad (21)$$

We start by computing $I(x_v; \hat{x}_v | \{h_{\tilde{u}, \tilde{v}}\})$. $\hat{x}_v$ has mean zero and variance

$$\begin{aligned}
\mathbb{E}(|\hat{x}_v|^2) &= \sum_{w \in W(n_{\ell+1})} \frac{|h_{v,w}|^2}{\|\boldsymbol{h}_w\|^2} K n_{\ell+1}^{-1} a_\ell^{\alpha/2} + \Delta^2 \\
&\leq n_{\ell+1} \frac{2^\alpha}{n_{\ell+1}} K n_{\ell+1}^{-1} a_\ell^{\alpha/2} + \Delta^2 \\
&\leq n_{\ell+1}^{-1} a_\ell^{\alpha/2}
\end{aligned} \quad (22)$$

for

$$K \triangleq 2^{-\alpha}(1 - \Delta^2)$$

which is positive for $\Delta^2 < 1$, and where we have used (21) and that

$$n_{\ell+1}^{-1} a_\ell^{\alpha/2} \geq 2^{\ell+1}\gamma(n) \geq 1$$

by (12). Equation (22) shows that $\hat{x}_v$ satisfies the power constraint of node v in the relay squarelet $V(n_{\ell+1})$. Moreover, we obtain

$$\begin{aligned}
I(x_v; \hat{x}_v | \{h_{\tilde{u}, \tilde{v}}\}) &= h(\hat{x}_v | \{h_{\tilde{u}, \tilde{v}}\}) - h(\hat{x}_v | x_v, \{h_{\tilde{u}, \tilde{v}}\}) \\
&\leq \log(2\pi e \mathbb{E}(|\hat{x}_v|^2)) - \log(2\pi e \Delta^2) \\
&\leq \log\left(\frac{n_{\ell+1}^{-1} a_\ell^{\alpha/2}}{\Delta^2}\right).
\end{aligned} \quad (23)$$

It remains to compute $I(\tilde{x}_w; y_w | \{h_{\tilde{u}, \tilde{v}}\})$. Note that the encoding procedure guarantees that

$$\cos(\pi/4)^2 \|\boldsymbol{h}_w\|^4 \leq |\boldsymbol{h}_w \hat{\boldsymbol{h}}_w^\dagger|^2 \leq \|\boldsymbol{h}_w\|^4.$$

Moreover, for $w \neq \tilde{w}$

$$\begin{aligned}
\mathbb{E}(|\boldsymbol{h}_w \hat{\boldsymbol{h}}_{\tilde{w}}^\dagger|^2) &= \mathbb{E}(\boldsymbol{h}_w \hat{\boldsymbol{h}}_{\tilde{w}}^\dagger \hat{\boldsymbol{h}}_{\tilde{w}} \boldsymbol{h}_w^\dagger) \\
&= \sum_{v \in V(n_{\ell+1})} \mathbb{E}(|h_{vw}|^2 |\hat{h}_{v\tilde{w}}|^2) \\
&= \sum_{v \in V(n_{\ell+1})} \mathbb{E}(|h_{vw}|^2 |h_{v\tilde{w}}|^2) \\
&= \mathbb{E}(|\boldsymbol{h}_w \boldsymbol{h}_{\tilde{w}}^\dagger|^2).
\end{aligned}$$

From this, we get by a similar argument as in Lemma 9 that

$$I(\tilde{x}_w; y_w | \{h_{\tilde{u}, \tilde{v}}\}) \geq K_6. \quad (24)$$

Using (23) and (24) in Lemma 8, and observing that we only communicate during a fraction

$$P_\ell(n) n_{\ell+1} a_\ell^{-\alpha/2}$$

of time, yields a per destination node rate $\rho_\ell^{\mathsf{BC}}(n)$ arbitrarily close to

$$K_6 P_\ell(n) n_{\ell+1} a_\ell^{-\alpha/2}$$

bits per channel use and a quantizer rate arbitrarily close to

$$\log\left(\frac{n_{\ell+1}^{-1} a_\ell^{\alpha/2}}{\Delta^2}\right)$$

bits per encoded sample. Since by (24) mutual information $I(\tilde{x}_w; y_w | \{h_{\tilde{u}, \tilde{v}}\})$ is at least K_6 for every $w \in W(n_{\ell+1})$ during the fraction of time we actually communicate, this implies that there are at most $1/K_6$ encoded message samples for each relay node per $n_{\ell+1}$ total message bits received by the destination nodes $W(n_{\ell+1})$. Thus, the number of bits required at each relay node to quantize the encoded message samples is at most

$$\begin{aligned}
\frac{1}{K_6} &\log\left(\frac{n_{\ell+1}^{-1} a_\ell^{\alpha/2}}{\Delta^2}\right) \\
&= \frac{1}{K_6} \log\left(\frac{1}{\Delta^2} 2^{\ell+1} \gamma^{1+\ell(1-\alpha/2)}(n) n^{\alpha/2-1}\right) \\
&\leq \frac{1}{K_6} \log\left(\frac{1}{\Delta^2} 2^{\ell+1} n^{\alpha/2}\right) \\
&\leq K_7(\ell+1)\log(n)
\end{aligned}$$

bits per $n_{\ell+1}$ total message bits received by the destination nodes, and where we have used $\gamma(n) \leq n$ by (12). $\qquad \square$

VII. PROOF OF THEOREM 1

The proof of Theorem 1 is split into two parts. In Section VII-A, we prove the theorem for fast fading, and in Section VII-B, for slow fading.

A. Fast Fading

In this section, we prove Theorem 1 under fast fading, i.e., $\{\theta_{u,v}[t]\}_t$ is stationary and ergodic in t. We first prove that the assumptions on the power constraint and the interference made in Section VI (see Lemmas 9 and 10) during the analysis of one level of the hierarchical relaying scheme are valid. We then use the results proved there to analyze the behavior of the entire hierarchy, yielding a lower bound on the per-node rate achievable with hierarchical relaying.

We first argue that the constraint $P_\ell(n) \leq n_{\ell+1}^{-1} a_\ell^{\alpha/2}$ needed in Lemmas 9 and 10 is satisfied. Consider the hierarchical relaying scheme as described in Section IV and fix a level $\ell, 0 \leq \ell < L = L(n)$ in this hierarchy. At level ℓ, we have a square of area $a_\ell = n/\gamma^\ell(n)$, with $n_\ell = n/2^\ell \gamma^\ell(n)$ source–destination pairs. Since we are time sharing between $K_2 2^{-\ell} \gamma(n)$ relay squarelets at this level, we have an average power constraint of

$$P_\ell(n) \triangleq K_2 2^{-\ell} \gamma(n)$$

during the time any particular relay squarelet is active. Since $\alpha > 2$ and since $n\gamma^{-L(n)}(n) \to \infty$ as $n \to \infty$, we have, for n large enough (independent of ℓ), that

$$\begin{aligned}
P_\ell(n) &= K_2 2^{-\ell} \gamma(n) \\
&\leq 2^{-\ell} \gamma(n) \left(\frac{n}{\gamma^{L(n)}(n)}\right)^{\alpha/2-1} \\
&\leq 2^{\ell+1} \gamma(n) \left(\frac{n}{\gamma^\ell(n)}\right)^{\alpha/2-1} \\
&= n_{\ell+1}^{-1} a_\ell^{\alpha/2}.
\end{aligned}$$

Therefore, the power constraint in Lemmas 9 and 10 is satisfied.

We continue by analyzing the interference caused by spatial reuse. Recall that the MAC and BC phases at level ℓ induce permutation traffic within the dense squarelets at level $\ell+1$. The permutation traffic within those dense squarelets at level $\ell+1$ is

transmitted in parallel with spatial reuse. We now describe in detail how this spatial reuse is performed. Partition the squarelets of area $a_{\ell+1}$ (i.e., at level $\ell+1$) into four subsets such that in each subset all squarelets are at distance at least $\sqrt{a_{\ell+1}}$ from each other. The traffic that the MAC and BC phases at level ℓ induce in each of the relay squarelets at level $\ell+1$ is transmitted simultaneously within all relay squarelets in the same subset. Consider now one such subset. We show that at any relay squarelet the interference from other relay squarelets in the same subset is stationary and ergodic within each phase, additive (i.e., independent of the signals and channel gains in this relay squarelet), and of bounded power $N_0 - 1$ independent of n.

We first argue that the interference is stationary and ergodic within each phase. Note first that on any level $\ell+1$ in the hierarchy, all relay squarelets are either simultaneously in the MAC phase or simultaneously in the BC phase. Furthermore, all relay squarelets are also synchronized for transmissions within each of these phases (recall that the induced traffic in level $\ell+1$ is uniform and is sent sequentially as permutation traffic). Hence, it suffices to show that the interference generated by either the MAC or the BC induced by some permutation traffic matrix is stationary and ergodic. Since all codebooks for either of these cases are generated as i.i.d. Gaussian multiplied by a Bernoulli process, and in the BC phase beamformed for stationary and ergodic fading, this is indeed the case.

The additivity of the interference follows easily for the MAC phase, since codebooks are generated independently of the channel realization in this case. Moreover, since the channel gains are independent from each other and all codebooks are generated as independent zero mean processes, the interference in the MAC phase is also uncorrelated (over space) within each relay squarelet. For the BC phase, the codebook depends only on the channel gains within each relay squarelet at level $\ell+1$. Since the channel gains within relay squarelets are independent of the channel gains between relay squarelets, this interference is additive as well.

We now bound the interference power. Note that by the randomized time-sharing construction within the MAC and BC phases (see Lemmas 9 and 10), in each relay squarelet, at most $n_{\ell+1}$ nodes transmit at an average power of 1. In the MAC phase, all nodes use independently generated codebooks with power at most 1, and thus, the received interference power from another relay squarelet at distance $i\sqrt{a_{\ell+1}}$ is at most

$$n_{\ell+1} i^{-\alpha} a_{\ell+1}^{-\alpha/2} = i^{-\alpha} 2^{-(\ell+1)} \left(\frac{n}{\gamma^{\ell+1}(n)} \right)^{1-\alpha/2} \leq i^{-\alpha}$$

by (12). In the BC phase, the nodes in each active relay squarelet use beamforming to transmit to nodes within their own squarelet. Since the channel gains within a relay squarelet are independent of the channel gains between relay squarelets, the same calculation as in (19) shows that we can upper bound the received interference power from another relay squarelet at distance $i\sqrt{a_{\ell+1}}$ by

$$n_{\ell+1} i^{-\alpha} a_{\ell+1}^{-\alpha/2} \leq i^{-\alpha}$$

in the BC phase as well.

Now, by the way in which we perform spatial reuse, every active relay squarelet has at most $8i$ active relay squarelets at distance at least $i\sqrt{a_{\ell+1}}$. Hence, the total interference power received at an active relay squarelet is at most

$$\sum_{i=1}^{\infty} 8i 2^{\alpha} i^{-\alpha} \triangleq N_0 - 1 < \infty$$

since $\alpha > 2$. With this, we have shown that the interference term has the properties required for Lemmas 9 and 10 to apply.

We now apply those two lemmas to obtain a lower bound on the rate achievable with hierarchical relaying. Call $\tau_\ell(n)$ the number of channel uses to transmit one bit from each of n_ℓ source nodes to the corresponding destination nodes at level ℓ. Lemma 7 states that for n large enough (independent of ℓ), we relay over each dense squarelet at most $K_3 2^{\ell}$ times. Combining this with Lemma 9, we see that to transmit one bit from each source to its destination at this level we need at most

$$4 K_3 2^{\ell} K_2 2^{-\ell} \gamma(n) \frac{1}{K_4 P_\ell(n)} n_{\ell+1}^{-1} a_\ell^{\alpha/2}$$
$$= \frac{K_3 2^{2\ell+3}}{K_4} n^{\alpha/2-1} \gamma^{1+\ell(1-\alpha/2)}(n)$$

channel uses for the MAC phase. Here, the factor 4 accounts for the spatial reuse, $K_3 2^{\ell}$ accounts for relaying over the same relay squarelets multiple times, $K_2 2^{-\ell} \gamma(n)$ accounts for time sharing between the relay squarelets, and the last term accounts for the time required to communicate over the MAC. Similarly, combining Lemmas 7 and 10, we need at most

$$\frac{K_3 2^{2\ell+3}}{K_6} n^{\alpha/2-1} \gamma^{1+\ell(1-\alpha/2)}(n)$$

channel uses for the BC phase. Moreover, at level $\ell+1$ in the hierarchy, this induces a per-node traffic demand of at most K_5 bits from the MAC phase, and at most $K_7(\ell+1)\log(n)$ from the BC phase. Thus, we obtain the following recursion:

$$\begin{aligned}
\tau_\ell(n) &\leq 8 K_3 \left(\frac{1}{K_4} + \frac{1}{K_6} \right) n^{\alpha/2-1} \gamma(n) (4\gamma^{1-\alpha/2}(n))^{\ell} \\
&\quad + (K_5 + K_7(\ell+1)\log(n)) \tau_{\ell+1}(n) \\
&\leq \tilde{K} n^{\alpha/2-1} \gamma(n) 4^{\ell} + K(\ell+1)\log(n) \tau_{\ell+1}(n) \\
&\leq \tilde{K} n^{\alpha/2-1} \gamma(n) 4^{L} + KL \log(n) \tau_{\ell+1}(n) \qquad (25)
\end{aligned}$$

for positive constants $K, \tilde{K}$ independent of n and ℓ.

We use TDMA at scale a_L with n_L nodes and source–destination pairs. Time sharing between all source–destination pairs, we have (during the time we communicate for each node) an average power constraint of n_L. Since at this level we communicate over a distance of at most $2a_L^{1/2}$, we have

$$\tau_L(n) \leq n_L \log^{-1} \left(1 + \frac{n_L}{2^{\alpha} N_0 a_L^{\alpha/2}} \right). \qquad (26)$$

Since

$$n_L a_L^{-\alpha/2} \leq n_L a_L^{-1} = 2^{-L} \to 0$$

as $n \to \infty$, we can upper bound (26) as

$$
\begin{aligned}
\tau_L(n) &\leq K' a_L^{\alpha/2} \\
&= K' n^{\alpha/2} \gamma^{-L\alpha/2}(n) \\
&\leq K' n^{\alpha/2} \gamma^{-L}(n)
\end{aligned}
\tag{27}
$$

for some constant K'.

Now, using the recursion (25) L times, and combining with (27), we obtain

$$
\begin{aligned}
\tau_0(n) &\leq \tilde{K} n^{\alpha/2-1} \gamma(n) 4^L + KL \log(n) \tau_1(n) \\
&\leq \cdots \\
&\leq \tilde{K} n^{\alpha/2-1} \gamma(n) 4^L \left(\sum_{\ell=0}^{L-1} (KL \log(n))^\ell \right) \\
&\quad + (KL \log(n))^L \tau_L(n) \\
&\leq n^{\alpha/2-1} (KL \log(n))^L (\tilde{K} 4^L \gamma(n) + K' n \gamma^{-L}(n)).
\end{aligned}
\tag{28}
$$

Using the definition of $\gamma(n)$ and $L = L(n)$ in (11), we have for n large enough

$$
\begin{aligned}
(KL(n) \log(n))^{L(n)} &\leq n^{2\log^{-1/2-\delta}(n) \log\log(n)} \\
4^{L(n)} \gamma(n) &\leq n^{2\log^{-1/2-\delta}(n) + \log^{\delta-1/2}(n)} \\
n\gamma^{-L(n)}(n) &\leq n^{\log^{\delta-1/2}(n)}.
\end{aligned}
$$

Since $\delta > 0$, the $n^{\log^{\delta-1/2}(n)}$ term dominates in (28), and we obtain

$$
\tau_0(n) \leq \tilde{b}(n) n^{\alpha/2-1}
$$

where

$$
\tilde{b}(n) \leq n^{O(\log^{\delta-1/2}(n))}
$$

as $n \to \infty$. Therefore

$$
\rho^*(n) \geq \rho^{\mathsf{HR}}(n) = 1/\tau_0(n) \geq b(n) n^{1-\alpha/2}
$$

with

$$
b(n) \geq n^{-O(\log^{\delta-1/2}(n))}
$$

concluding the proof for the fast-fading case.

B. Slow Fading

In this section, we prove Theorem 1 under slow fading, i.e., $\{\theta_{u,v}[t]\}_t$ is constant as a function of t. We sketch the necessary modifications for the scheme described in Section IV to achieve a per-node rate of at least $b(n) n^{1-\alpha/2}$ in the slow-fading case.

Consider level $\ell, 0 \leq \ell < L(n)$ in the hierarchy. Instead of relaying the message of a source–destination pair over one relay squarelet as in the scheme described in Section IV, we relay the message over many dense squarelets that are at least at distance $\sqrt{2a_{\ell+1}}$ from both the source and the destination nodes. We time share between the different relays. The idea here is that the wireless channel between any node and its relay squarelet might be in a bad state due to the slow fading, making communication

over this relay squarelet impossible. Averaged over many relay squarelets, however, we get essentially the same performance as in the fast-fading case.

We first state a (somewhat weaker) version of Lemma 7, appropriate for this setup. Consider again the collection of schedules $\mathcal{S}(n_\ell)$ and $\tilde{\mathcal{S}}(n_\ell)$ satisfying the conditions that no relay squarelet is selected by more than $n_{\ell+1}$ source–destination pairs and that all sources and destinations are at least at distance $\sqrt{2a_{\ell+1}}$ from their relay squarelet (see Section VI-A for the formal definition). The next lemma shows that for each source–destination pair, we can find $K_2 2^{-\ell-1} \gamma(n)$ distinct relay squarelets satisfying the above conditions (the requirement that these relay squarelets are distinct is expressed by the orthogonality condition of the schedules in Lemma 11).

Lemma 11: For every n large enough (independent of ℓ) and every permutation traffic matrix $\lambda(n_\ell) \in \{0,1\}^{n_\ell \times n_\ell}$, there are schedules $\{S^{(i)}(n_\ell)\}_{i=1}^{K_2 2^{-\ell} \gamma^2(n)} \subset \mathcal{S}(n_\ell), \{\tilde{S}^{(i)}(n_\ell)\}_{i=1}^{K_2 2^{-\ell} \gamma^2(n)} \subset \tilde{\mathcal{S}}(n_\ell)$ satisfying

$$
\lambda(n_\ell) = \frac{1}{K_2 2^{-\ell-1} \gamma(n)} \sum_{i=1}^{K_2 2^{-\ell} \gamma^2(n)} S^{(i)}(n_\ell) \tilde{S}^{(i)}(n_\ell)
$$

where $\{S^{(i)}(n_\ell)\}_i, \{\tilde{S}^{(i)}(n_\ell)\}_i$ are collections of orthogonal matrices in the sense that for $i \neq i'$

$$
\begin{aligned}
\sum_{u,k} s_{u,k}^{(i)} s_{u,k}^{(i')} &= 0 \\
\sum_{k,u} \tilde{s}_{k,u}^{(i)} \tilde{s}_{k,u}^{(i')} &= 0.
\end{aligned}
\tag{29}
$$

Proof: The proof is similar to that of Lemma 7. In order to construct $\{S^{(i)}(n_\ell)\}$ and $\{\tilde{S}^{(i)}(n_\ell)\}$, consider the sequential pass over all n source–destination pairs (assume n is large enough for Lemma 7 to hold). As before, for each source–destination pair, there are $K_2 2^{-\ell-1} \gamma(n)$ dense relay squarelets that are at distance at least $\sqrt{2a_{\ell+1}}$. Each pair chooses all of these $K_2 2^{-\ell-1} \gamma(n)$ squarelets, instead of just one as before. Stop one round of this procedure as soon as any of the relay squarelets is chosen by $n_{\ell+1}$ pairs. Since by the end of one round at least one relay squarelet is matched by $n_{\ell+1}$ source–destination pairs, there are at most $n_\ell/n_{\ell+1} = 2\gamma(n)$ such rounds.

Consider now the result of one such round. We construct $K_2 2^{-\ell-1} \gamma(n)$ matrices $S^{(i)}(n_\ell)$ and $\tilde{S}^{(i)}(n_\ell)$, with the ith pair of matrices describing communication over the ith relay squarelets chosen by source–destination pairs matched in this round. Thus, this process produces a total of $2\gamma(n) K_2 2^{-\ell-1} \gamma(n) = K_2 2^{-\ell} \gamma^2(n)$ such matrices. The orthogonality property follows since each source–destination pair relays over the same relay squarelet only once. $\square$

Given a decomposition of the scaled traffic matrix $K_2 2^{-\ell-1} \gamma(n) \lambda(n)$ into $K_2 2^{-\ell} \gamma^2(n)$ matrices, each source–destination pair tries to relay over $K_2 2^{-\ell-1} \gamma(n)$ dense squarelets. We time share between these relay squarelets. Since each source–destination pair relays only a $(K_2 2^{-\ell-1} \gamma(n))^{-1}$ fraction of traffic over any of its relay squarelets, the loss due to this time sharing is now

$$
\frac{K_2 2^{-\ell} \gamma^2(n)}{K_2 2^{-\ell-1} \gamma(n)} = 2\gamma(n)
$$

as opposed to $K_3 2^\ell$ in Lemma 7. In other words, the loss is at most a factor $2\gamma(n)$ more than in Lemma 7. Using the definition of $\gamma(n)$ in (11), we have

$$\gamma(n) \leq n^{-\log \delta^{-1/2}(n)} \leq b^{-1}(n).$$

In other words, this additional loss is small.

Consider now a specific relay squarelet. If a source–destination pair can communicate over this relay squarelet at a rate at least $1/64$th of the rate achievable in the fast-fading case (given by Lemmas 9 and 10), it sends information over this relay. Otherwise, it does not send anything during the period of time it is assigned this relay. We now show that, with probability $1 - o(1)$ as $n \to \infty$, for every source–destination pair on every level of the hierarchy, at least one quarter of its relay squarelets can support this rate. As we only communicate over a quarter of the relay squarelets, this implies that we can achieve at least $1/256$th of the per-node rate for the fast-fading case (see Section VII-A), i.e., that $b(n)n^{1-\alpha/2}$ is achievable with probability $1 - o(1)$ as $n \to \infty$.

Assume that for each source–destination pair (u, w) we have picked $K_2 2^{-\ell-1}\gamma(n)$ dense squarelets over which it can relay; call those relay squarelets $\{A_{u,w,k}\}_{k=1}^{K_2 2^{-\ell-1}\gamma(n)}$. Consider the event $B_{u,w,k}$ that source node u can communicate at the desired rate to destination node w over relay squarelets $A_{u,w,k}$ (assuming, as before, that we can solve the communication problem within this squarelet).

Let $\{B_{u,w,k}^{(i)}\}_{i=1}^4$ be the events that the interference due to matched filtering in the MAC phase, the interference from spatial reuse in the MAC phase, the interference due to beamforming in the BC phase, and the interference from spatial reuse in the BC phase are less than eight times the one for fast fading, respectively. From the proof of Lemmas 9 and 10 and Theorem 1 for the fast-fading case in Section VII-A, we see that

$$\bigcap_{i=1}^4 B_{u,w,k}^{(i)} \subset B_{u,w,k}.$$

Due to spatial reuse, multiple relay squarelets will be active in parallel. Let $\tilde{H}$ denote the set of channel gains between active relay squarelets. Using essentially the same arguments as for the fast-fading case (see Lemmas 9 and 10 and Section VII-A) and from Markov's inequality, we have $\mathbb{P}(B_{u,w,k}^{(i)} \mid \tilde{H}) \geq 7/8$ for all $i \in \{1, \ldots, 4\}$, and hence, $\mathbb{P}(B_{u,w,k} \mid \tilde{H}) \geq 1/2$.

We now argue that the events

$$\left\{ \cap_{i=1}^4 B_{u,w,k}^{(i)} \right\}_{k=1}^{K_2 2^{-\ell-1}\gamma(n)} \tag{30}$$

are independent conditioned on $\tilde{H}$, by showing that these events depend on disjoint sets of channel gains and codebooks. Assuming the codebooks are generated new for each communication round, then they are all independent. Thus, we only have to consider the dependence on the channel gains. Let U_k and W_k be the source and destination nodes communicating over relay squarelet $A_{u,w,k}$ in round k, and let V_k be the nodes in $A_{u,w,k}$. Let $\tilde{U}_k, \tilde{W}_k$ be the source and destination nodes that are communicating at the same time as (u, w) due to spatial reuse. Let $\tilde{V}_k$ be the relay nodes of $\tilde{U}_k$ and $\tilde{W}_k$. Now, $B_{u,w,k}^{(1)}$ and $B_{u,w,k}^{(2)}$ depend

(for fixed $\tilde{H}$) on the channel gains between U_k and V_k. $B_{u,w,k}^{(3)}$ depends on the channel gains between V_k and W_k. $B_{u,w,k}^{(4)}$ depends (again for fixed $\tilde{H}$) on the channel gains between $\tilde{V}_k$ and $\tilde{W}_k$. Since these sets are disjoint for different k by the orthogonality of the schedules [see (29)], conditional independence of the events in (30) follows.

To summarize, conditioned on the channel gains $\tilde{H}$ between active relay squarelets, the random variables $\{\mathbb{1}_{B_{u,w,k}}\}_k$ are independent and have expected value $\mathbb{E}(\mathbb{1}_{B_{u,w,k}} \mid \tilde{H}) \geq 1/2$. The sum

$$\sum_{k=1}^{K_2 2^{-\ell-1}\gamma(n)} \mathbb{1}_{B_{u,w,k}}$$

is the number of relay squarelets over which the source–destination pair (u, w) successfully relays traffic. We now show that with high probability at least one quarter of these relay squarelets allow successful transmission. Applying the Chernoff bound yields that

$$\mathbb{P}\left(\sum_k \mathbb{1}_{B_{u,w,k}} < K_2 2^{-\ell-3}\gamma(n) \mid \tilde{H} \right)$$
$$\leq \mathbb{P}\left(\sum_k \mathbb{1}_{B_{u,w,k}} < K_2 2^{-\ell-2}\gamma(n)\mathbb{P}(B_{u,w,k} \mid \tilde{H}) \right)$$
$$\leq \exp(-2K_2 2^{-\ell}\gamma(n)\mathbb{P}(B_{u,w,k} \mid \tilde{H}))$$
$$\leq \exp(-K 2^{-\ell}\gamma(n))$$

for some constant $K > 0$. Since the right-hand side is the same for all $\tilde{H}$, this implies

$$\mathbb{P}\left(\sum_k \mathbb{1}_{B_{u,w,k}} < K_2 2^{-\ell-3}\gamma(n) \right) \leq \exp(-K 2^{-\ell}\gamma(n)).$$

In each of the $L(n)$ levels of the hierarchy, there are at most n^2 source–destination pairs, and hence, by the union bound with probability at least

$$1 - L(n)n^2 \exp(-K 2^{-L(n)}\gamma(n))$$

for every source–destination pair on every level of the hierarchy at least one quarter of its relay squarelets can support the desired rate. By the choices of $\gamma(n)$ and $L(n)$ in (11), this probability is at least

$$1 - L(n)n^2 \exp\left(-K 2^{-L(n)}\gamma(n) \right)$$
$$\geq 1 - n^3 \exp\left(-K 2^{-L(n)} 2^{\log(n)/2L(n)} \right)$$
$$\geq 1 - \exp\left(\tilde{K} 2^{\log\log(n)} - K 2^{\frac{1}{2}\log^{1/2+\delta}(n) - \log^{1/2-\delta}(n)} \right)$$
$$\geq 1 - \exp(-2^{\Omega(\log^{1/2+\delta}(n))})$$
$$\geq 1 - o(1)$$

as $n \to \infty$, and for some constant $\tilde{K}$. This proves that the same order rate as in the fast-fading case can be achieved with high probability for all levels $0 \leq \ell < L(n)$.

It remains to argue that the same holds for level $\ell = L(n)$. Note that since we assume phase fading only, the received signal

power is only a function of distance and not of the fading realization. Since at level $L(n)$ we use simple TDMA, this implies that we can always achieve the same rate at level $L(n)$ as in the fast-fading case.

Hence, with probability $1 - o(1)$ as $n \to \infty$, we achieve the same order rate at each level $0 \le \ell \le L(n)$ as for fast fading, proving Theorem 1 for the slow-fading case.

VIII. PROOF OF THEOREM 2

Here, we provide a generalization and sharpening of the converse in [8]. Most of the arguments follow [8, Th. 5.2]. We start by proving a lemma upper bounding the multiple-input–multiple-output (MIMO) capacity.

Consider two subsets $S_1, S_2 \subset V(n)$ such that $S_1 \cap S_2 = \emptyset$. Assume we allow the nodes within S_1 and S_2 to cooperate without any restriction. The maximum achievable sum rate between the nodes in S_1 and S_2 is given by the MIMO capacity $C(S_1, S_2)$ between them. The next lemma upper bounds $C(S_1, S_2)$ in terms of the node distances between the two sets and the *normalized channel gains*

$$\tilde{h}_{u,v} \triangleq \frac{h_{u,v}}{\sqrt{\sum_{\tilde{v} \in S_2} r_{u,\tilde{v}}^{-\alpha}}}.$$

Lemma 12: Under either fast or slow fading, for every $\alpha > 2$, $S_1, S_2 \subset V(n)$ with $S_1 \cap S_2 = \emptyset$, we have

$$C(S_1, S_2) \le 4 \left(\max \left\{ 1, \max_{v \in S_2} \sum_{u \in S_1} |\tilde{h}_{u,v}|^2 \right\} \right) \sum_{u \in S_1} \sum_{v \in S_2} r_{u,v}^{-\alpha}.$$

Proof: Let

$$\boldsymbol{H} \triangleq \{h_{u,v}\}_{u \in S_1, v \in S_2}$$
$$\tilde{\boldsymbol{H}} \triangleq \{\tilde{h}_{u,v}\}_{u \in S_1, v \in S_2}$$

be the matrix of (normalized) channel gains between the nodes in S_1 and S_2. Consider first fast fading. Under this assumption, we have

$$C(S_1, S_2) \triangleq \max_{\substack{\boldsymbol{Q}(\boldsymbol{H}) > 0: \\ \mathbb{E}(q_{u,u}) \le 1 \ \forall u \in S_1}} \mathbb{E}(\log \det(\boldsymbol{I} + \boldsymbol{H}^\dagger \boldsymbol{Q}(\boldsymbol{H}) \boldsymbol{H})).$$

Define

$$P_{S_1, S_2} \triangleq \sum_{u \in S_1} \sum_{v \in S_2} r_{u,v}^{-\alpha}$$

as the total received power in S_2 from S_1, and set

$$P_{u, S_2} \triangleq P_{\{u\}, S_2}$$

with slight abuse of notation. Then

$$C(S_1, S_2)$$
$$= \max_{\substack{\boldsymbol{Q}(\boldsymbol{H}) > 0: \\ \mathbb{E}(q_{u,u}) \le P_{u,S_2} \forall u \in S_1}} \mathbb{E}(\log \det(\boldsymbol{I} + \tilde{\boldsymbol{H}}^\dagger \boldsymbol{Q}(\boldsymbol{H}) \tilde{\boldsymbol{H}}))$$
$$\le \max_{\substack{\boldsymbol{Q}(\boldsymbol{H}) > 0: \\ \mathbb{E}(\mathrm{tr}\boldsymbol{Q}(\boldsymbol{H})) \le P_{S_1,S_2}}} \mathbb{E}(\log \det(\boldsymbol{I} + \tilde{\boldsymbol{H}}^\dagger \boldsymbol{Q}(\boldsymbol{H}) \tilde{\boldsymbol{H}})). \quad (31)$$

Define the event

$$B \triangleq \{\|\tilde{\boldsymbol{H}}\|^2 > b\}$$

for some b and where $\|\tilde{\boldsymbol{H}}\|$ denotes the largest singular value of $\tilde{\boldsymbol{H}}$. In words, B is the event that the channel gains between S_1 and S_2 are "good." We argue that, for appropriately chosen b, the event B has probability zero (i.e., the channel cannot be too "good"). By Markov's inequality

$$\mathbb{P}(B) \le b^{-m} \mathbb{E}(\|\tilde{\boldsymbol{H}}\|^{2m}) \quad (32)$$

for any m. We continue by upper bounding $\mathbb{E}(\|\tilde{\boldsymbol{H}}\|^{2m})$. We have

$$\|\tilde{\boldsymbol{H}}\|^{2k} \le \mathrm{tr}((\tilde{\boldsymbol{H}}\tilde{\boldsymbol{H}}^\dagger)^k)$$

for any k, and hence

$$\mathbb{E}(\|\tilde{\boldsymbol{H}}\|^{2m}) \le \mathbb{E}((\mathrm{tr}((\tilde{\boldsymbol{H}}\tilde{\boldsymbol{H}}^\dagger)^k))^{m/k}). \quad (33)$$

Now, for any $k \ge m$, we have by Jensen's inequality

$$\mathbb{E}((\mathrm{tr}((\tilde{\boldsymbol{H}}\tilde{\boldsymbol{H}}^\dagger)^k))^{m/k}) \le (\mathbb{E}\mathrm{tr}((\tilde{\boldsymbol{H}}\tilde{\boldsymbol{H}}^\dagger)^k))^{m/k}. \quad (34)$$

Combining (32)–(34) yields

$$\mathbb{P}(B) \le b^{-m} (\mathbb{E}\mathrm{tr}((\tilde{\boldsymbol{H}}\tilde{\boldsymbol{H}}^\dagger)^k))^{m/k} \quad (35)$$

for any $k \ge m$.

Now, the arguments in [8, Lemma 5.3] show that

$$\mathbb{E}(\mathrm{tr}((\tilde{\boldsymbol{H}}\tilde{\boldsymbol{H}}^\dagger)^k)) \le t_k n \left(\max \left\{ 1, \max_{v \in S_2} \sum_{u \in S_1} |\tilde{h}_{u,v}|^2 \right\} \right)^k$$

where t_k is the kth Catalan number. Combining with (35), this yields

$$\mathbb{P}(B) \le \left(b^{-1} t_k^{1/k} n^{1/k} \left(\max \left\{ 1, \max_{v \in S_2} \sum_{u \in S_1} |\tilde{h}_{u,v}|^2 \right\} \right) \right)^m.$$

Taking the limit as $k \to \infty$ and using that $t_k^{1/k} \to 4$ yields

$$\mathbb{P}(B) \le \left(b^{-1} 4 \left(\max \left\{ 1, \max_{v \in S_2} \sum_{u \in S_1} |\tilde{h}_{u,v}|^2 \right\} \right) \right)^m.$$

Assume

$$b > 4 \left(\max \left\{ 1, \max_{v \in S_2} \sum_{u \in S_1} |\tilde{h}_{u,v}|^2 \right\} \right) \quad (36)$$

then taking the limit as $m \to \infty$ shows that

$$\mathbb{P}(B) = 0.$$

Using this, we can upper bound (31) as

$$C(S_1, S_2) \le \max_{\substack{\boldsymbol{Q}(\boldsymbol{H}) > 0: \\ \mathbb{E}(\mathrm{tr}\boldsymbol{Q}(\boldsymbol{H})) \le P_{S_1,S_2}}} \mathbb{E}(\mathrm{tr}(\tilde{\boldsymbol{H}}^\dagger \boldsymbol{Q}(\boldsymbol{H}) \tilde{\boldsymbol{H}}))$$
$$= \max_{\substack{\boldsymbol{Q}(\boldsymbol{H}) > 0: \\ \mathbb{E}(\mathrm{tr}\boldsymbol{Q}(\boldsymbol{H})) \le P_{S_1,S_2}}} \mathbb{E}(\mathbb{1}_{B^c} \mathrm{tr}(\tilde{\boldsymbol{H}}^\dagger \boldsymbol{Q}(\boldsymbol{H}) \tilde{\boldsymbol{H}}))$$
$$\le \max_{\substack{\boldsymbol{Q}(\boldsymbol{H}) > 0: \\ \mathbb{E}(\mathrm{tr}\boldsymbol{Q}(\boldsymbol{H})) \le P_{S_1,S_2}}} \mathbb{E}(\mathbb{1}_{B^c} \|\tilde{\boldsymbol{H}}\|^2 \mathrm{tr}\boldsymbol{Q}(\boldsymbol{H}))$$
$$\le b P_{S_1, S_2}.$$

Since this is true for all b satisfying (36), we obtain the lemma for the fast-fading case.

Under slow fading

$$C(S_1, S_2) \triangleq \max_{\substack{\boldsymbol{Q} > 0: \\ q_{u,u} \leq P \ \forall u \in S_1}} \log \det(\boldsymbol{I} + \boldsymbol{H}^\dagger \boldsymbol{Q} \boldsymbol{H})$$

and the lemma can be obtained by the same steps. $\square$

We now proceed to the proof of Theorem 2. Consider a vertical cut dividing the network into two parts. By the minimum-separation requirement, an area of size $o(n)$ can contain at most $o(n)$ nodes, and hence, we can find a cut such that each part is of size $\Theta(n)$ and contains $\Theta(n)$ nodes. Call the left part of the cut S. Since there are $\Theta(n)$ nodes in S and in S^c, there are $\Theta(n)$ sources in S with their destination in S^c with probability $1 - o(1)$. For technical reasons, we add a node inside each square in $V(n)$ of the form $[id, (i+1)d] \times [jd, (j+1)d]$ for some $i, j \in \mathbb{N}$, where $d \triangleq \sqrt{2 \log(n)}$. These additional nodes have no traffic demands on their own, and simply help with the transmission. This can clearly only increase achievable rates. Moreover, this increases the number of nodes in V by less than a factor 2. We now show that

$$C(S, S^c) = O(\log^6(n) n^{2-\alpha/2}) \tag{37}$$

and, hence, by the cut-set bound, and since there are $\Theta(n)$ sources in S with their destination in S^c, we have

$$\rho^*(n) = O(\log^6(n) n^{1-\alpha/2}).$$

We prove (37) using Lemma 12. To this end, we need to upper bound

$$\max_{v \in S^c} \sum_{u \in S} |\tilde{h}_{u,v}|^2.$$

The proof of [8, Lemma 5.3] shows that if:
1) there are less than $\log(n)$ nodes inside $[i, i+1] \times [j, j+1]$ for any $i, j \in \{0, \ldots, \sqrt{n} - 1\}$;
2) there is at least one node inside $[id, (i+1)d] \times [jd, (j+1)d]$ for any i, j, where $d \triangleq \sqrt{2 \log n}$;
then

$$\max_{v \in S^c} \sum_{u \in S} |\tilde{h}_{u,v}|^2 \leq K \log^3(n) \tag{38}$$

and for $\alpha \in (2, 3]$

$$\sum_{u \in S} \sum_{v \in S^c} r_{u,v}^{-\alpha} \leq \tilde{K} \log^3(n) n^{2-\alpha/2} \tag{39}$$

for constants $K, \tilde{K}$. For arbitrary node placement with minimum separation, the first requirement is satisfied for n large enough, since only a constant number of nodes can be contained in each area of constant size. By our addition of nodes into $V(n)$ described above, the second condition is also satisfied. Using Lemma 12 with (38) and (39) yields (37), concluding the proof of Theorem 2.

IX. PROOF OF THEOREM 3

Consider a node placement with $n/2$ nodes located uniformly on $[0, \sqrt{n}/4] \times [0, \sqrt{n}]$ and $n/2$ nodes located on $[\sqrt{n}/2, \sqrt{n}] \times [0, \sqrt{n}]$ with minimum separation $r_{\min} = 1/2$. A random traffic matrix $\lambda(n)$ is such that at least $n/4$ communication pairs have

their sources in the left cluster and destinations in the right cluster with probability $1 - o(1)$. Assume we are dealing with such a $\lambda(n)$ in the following.

In this setup, with multihop, at least one hop has to cross the gap between the left and the right cluster. Thus, even without any interference from other nodes, we can obtain at most

$$\rho^{\mathsf{MH}}(n) \leq 4^\alpha n^{-\alpha/2}.$$

Moreover, considering a cut between the two clusters (say, S and S^c), and applying Lemma 12 yields that

$$\rho^*(n) \leq 16 n^{-1} \left(\max \left\{ 1, \max_{v \in S^c} \sum_{u \in S} |\tilde{h}_{u,v}|^2 \right\} \right) \sum_{u \in S} \sum_{v \in S^c} r_{u,v}^{-\alpha}. \tag{40}$$

Now note that for any $u \in S, v \in S^c$, we have

$$\frac{1}{4} \sqrt{n} \leq r_{u,v} \leq 2\sqrt{n}.$$

Hence

$$\sum_{u \in S} |\tilde{h}_{u,v}|^2 = \sum_{u \in S} \frac{r_{u,v}^{-\alpha}}{\sum_{\tilde{v} \in S^c} r_{u,\tilde{v}}^{-\alpha}} \leq 2^{3\alpha}$$

and

$$\sum_{u \in S} \sum_{v \in S^c} r_{u,v}^{-\alpha} \leq 4^{\alpha-1} n^{2-\alpha/2}.$$

Combining this with (40) yields

$$\rho^*(n) \leq 2^{2+5\alpha} n^{1-\alpha/2}$$

for all $\alpha > 2$.

X. PROOF OF THEOREM 4

We construct a cooperative multihop communication scheme and lower bound the per-node rate $\rho^{\mathsf{CMH}}(n)$ it achieves. We use the hierarchical relaying scheme as building block. Assume the node placement $V(n)$ is μ-regular at resolution $d(n)$ for all $n \geq 1$. We show that this implies that we can achieve a per-node rate of at least $d^{3-\alpha}(n) n^{-1/2-\beta(n)}$ as $n \to \infty$. Taking the smallest such $d(n)$ then yields the result.

We consider three cases for the value of $d(n)$ [namely, $d(n) = \Theta(\sqrt{n}), d(n) \geq n^{o(1)}$, and $d(n) \leq n^{o(1)}$]. First, if $d(n) = \Theta(\sqrt{n})$ as $n \to \infty$, then the result follows directly from Theorem 1. Considering a subsequence if necessary, we can therefore assume without loss of generality that $d(n) = o(\sqrt{n})$ in the following.

Second, consider $d(n)$ satisfying

$$d(n) \geq n^{\frac{1}{2+\alpha}} \log^{\delta - 1/2}(n). \tag{41}$$

Divide $A(n)$ into squares of sidelength $d(n)$. Since $d(n) = o(\sqrt{n})$, the number of such squares grows unbounded as $n \to \infty$. We now show that we can use multihop communication with a hop length of $d(n)$ where each hops is implemented by squares cooperatively sending information to a neighboring square. In other words, we perform cooperative communication at local scale $d(n)$ and multihop communication at global scale $\sqrt{n}$.

Since $V(n)$ is μ-regular at resolution $d(n)$, each such square contains at least $\mu d^2(n)$ nodes. Pick the top left most square and construct the square of sidelength $2d(n)$ consisting of it together
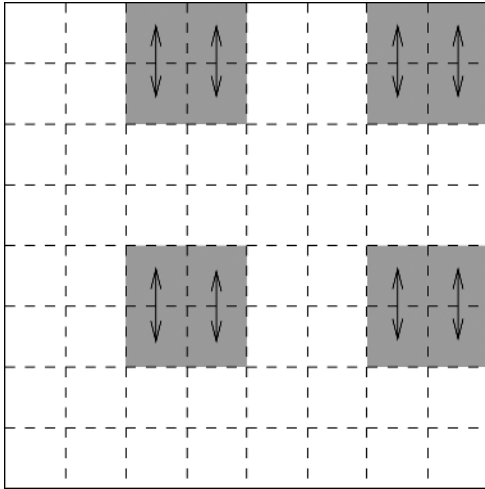
Fig. 6. Sketch of the construction of the cooperative multihop scheme in the proof of Theorem 4. The dashed squares have sidelength $d(n)$. The gray area is one of the four subsets of bigger squares that communicate simultaneously. The arrows indicate the traffic matrix λ_1.

with its three neighbors. Continue in the same fashion, partitioning all of $A(n)$ into squares of sidelength $2d(n)$. Note that each such bigger square contains at least $4\mu d^2(n)$ nodes by the definition of $d(n)$. We assume this worst case in the following. Partition $A(n)$ into four subsets of those bigger squares such that within each such subset each square is at distance at least $2d(n)$ from any other square (see Fig. 6). We time share between those four subsets. Consider in the following one such subset. For every bigger square, we construct two permutation traffic matrices $\lambda_1(4\mu d^2(n))$ and $\lambda_2(4\mu d^2(n))$. In λ_1, the nodes in the top two squares have as destinations the nodes in the bottom two squares and the nodes in the bottom two squares have as destinations the nodes in the top two squares (see Fig. 6). Similarly, λ_2 contains communication pairs between left and right squares. We time share between λ_1 and λ_2.

Communication according to λ_i within bigger squares in the same subset occurs simultaneously. We are going to use hierarchical relaying within each bigger square. This is possible since each such square contains at least $4\mu d^2(n)$ nodes. We have to show that the additional interference from bigger squares in the same subset is such that Theorem 1 still applies. In particular, we need to show that the interference has bounded power, say K. Using the same arguments as in the proof of Theorem 1 in Section VII yields that this is indeed the case (the interference from other bigger squares here behaves the same way as the interference due to spatial reuse from other active relay squarelets there). With this, we are now dealing with a hierarchical relaying scheme with area $4d^2(n)$, $4\mu d^2(n)$ nodes, and additive noise with power $1 + K$. Both the lower number of nodes and the higher noise power will decrease the achievable per-node rate by only some constant factor, and hence, Theorem 1 shows that under fast fading we can achieve a per-node rate of at least

$$b_1(d^2(n))(d^2(n))^{1-\alpha/2} \geq b_1(n)d^{2-\alpha}(n)$$

as $n \to \infty$, where

$$b_1(n) \geq n^{-O(\log^{\delta-1/2}(n))}.$$

Moreover, the same rate is achievable under slow fading with probability $1 - b_2(d^2(n))$, where

$$b_2(n) \leq \exp\left(-2^{\Omega(\log^{1/2+\delta}(n))}\right).$$

The setup is the same for all bigger squares within each of the four subsets.

We now "shift" the way we defined the bigger squares by $d(n)$ to the right and to the bottom. With this, each new bigger square intersects with four bigger squares as defined before. We use the same communication scheme within these new bigger squares and time share between the two ways of defining bigger squares.

Construct now a graph where each vertex corresponds to a square of sidelength $d(n)$ and where two vertices are connected by an edge if they are adjacent in either the same old or new bigger square. This graph is depicted in Fig. 4.

With the above construction, we can communicate along each edge of this graph simultaneously at a per-node rate of

$$\frac{b_1(n)}{16}d^{2-\alpha}(n)$$

in the fast-fading case. In the slow-fading case, this statement holds with probability at least

$$1 - \frac{n}{d^2(n)}b_2(d^2(n))$$
$$= 1 - \frac{n}{d^2(n)}\exp\left(-2^{\Omega(\log^{1/2+\delta}(d^2(n)))}\right)$$
$$\geq 1 - \exp\left(K'2^{\log\log(n)} - 2^{\tilde{K}\log^{1/2+\delta}(d(n))}\right)$$

for constants $K', \tilde{K}$. By assumption (41)

$$\log^{1/2+\delta}(d(n)) \geq \left(\frac{1}{2+\alpha}\log^{1/2+\delta}(n)\right)^{1/2+\delta}$$

and, hence

$$1 - \frac{n}{d^2(n)}b_2(d^2(n)) \geq 1 - o(1)$$

as $n \to \infty$, showing that with high probability, we achieve the same order rate under slow fading as under fast fading.

The communication graph constructed forms a grid with $n/d^2(n)$ nodes. Using that each bigger square can contain at most $K_1 d^2(n)$ nodes by the minimum-separation requirement, standard arguments for routing over grid graphs (see [16]) show that in the fast-fading case we can achieve a per-node rate of

$$\rho^{\mathsf{CMH}}(n) \geq \tilde{b}(n)d^{2-\alpha}(n)\frac{d(n)}{\sqrt{n}} \geq \tilde{b}(n)d^{3-\alpha}(n)n^{-1/2}$$

where

$$\tilde{b}(n) = n^{-O(\log^{\delta-1/2}(n))}.$$

Moreover, the same statement holds in the slow-fading case with probability $1 - o(1)$.

Finally, consider $d(n)$ such that

$$d(n) \leq n^{\frac{1}{2+\alpha}\log^{\delta-1/2}(n)}. \tag{42}$$

Construct the same communication graph as before, but this time we use simple multihop communication between adjacent squares of sidelength $d(n)$. By time sharing between the at most

$K_1 d^2(n)$ nodes in each square, and since we communicate over a distance of at most $3d(n)$, we achieve under either fast or slow fading a per-node rate between the squares of at least

$$K'' d^{-2-\alpha}(n) \geq K'' n^{-\log^{\delta-1/2}(n)}$$

for some constant K'', and where we have used (42). Using the analysis of grid graphs as before, we can achieve a per-node rate of at least

$$\rho^{\mathsf{CMH}}(n) \geq K'' n^{-\log^{\delta-1/2}(n)} \frac{d(n)}{\sqrt{n}}$$
$$\geq \tilde{b}(n) d^{3-\alpha}(n) n^{-1/2}$$

for either the fast- or slow-fading case.

XI. PROOF OF THEOREM 5

Consider $V(n)$ with $n/2$ nodes located uniformly on $[0, (\sqrt{n} - d^*(n))/2] \times [0, \sqrt{n}]$ and $n/2$ nodes located uniformly on $[\sqrt{n}/2, \sqrt{n}] \times [0, \sqrt{n}]$ such that $r_{\min} = 1/2$. This node placement is $1/2$-regular at resolution $d^*(n)$. A random traffic matrix $\lambda(n)$ is such that $\Theta(n)$ communication pairs have their sources in the left cluster and destinations in the right cluster with probability $1 - o(1)$. Assume we are dealing with such a $\lambda(n)$ in the following.

Considering a cut between the two clusters and applying Lemma 12 (slightly adapting the arguments in Section VIII) yields that

$$\rho^*(n) = O(\log^6(n) d^{*3-\alpha}(n) n^{-1/2})$$

for $\alpha > 3$.

XII. DISCUSSION

We briefly discuss several aspects of the proposed hierarchical relaying scheme. Section XII-A comments on the full CSI assumption and Section XII-B on the use of bursty communication. Sections XII-C and XII-D outline how the results obtained here can be extended to the case of dense networks and networks without minimum separation between nodes. Section XII-E compares our hierarchical relaying scheme to the hierarchical cooperation scheme presented in [8].

A. Full CSI Assumption

Throughout our analysis, we have made a full CSI assumption. In other words, we assumed that the phase shifts $\{\theta_{u,v}[t]\}_{u,v}$ are available at time t at all nodes in the network. As this assumption is quite strong, it is worth commenting on. First, we make the full CSI assumption in all the converse results in this paper. This implies that all the converses also hold under weaker assumptions on the CSI, and hence, are valid as well under a wide variety of more realistic assumptions on the availability of side information. Second, all achievability results can be shown to hold under weaker assumptions on the availability of CSI. In fact, in all cases, a 2-bit quantization of the channel state $\{\theta_{u,v}[t]\}_{u,v}$ available at all nodes at time t

is sufficient to obtain the same scaling behavior. This follows by an argument similar to the one used in the analysis of the BC phase in Section VI-C, where it is shown that beamforming using a quantized channel state results only in a constant factor rate loss.

B. Burstiness of Hierarchical Relaying Scheme

The hierarchical relaying scheme presented here is bursty in the sense that nodes communicate at high power during a small fraction of time. This leads to high peak-to-average power ratio, which is undesirable in practice. We chose burstiness in the time domain to simplify the exposition. The same bursty behavior could be achieved in a more practical manner by using code-division multiple access (CDMA) with several orthogonal signatures or by using orthogonal frequency-division multiplexing (OFDM) with many subcarriers. Each approach leads to many parallel channels out of which only few are used with higher power. This avoids the issue of high peak-to-average power ratio in the time domain.

C. Dense Networks

Throughout this paper, we have only considered *extended* networks, i.e., n nodes placed on a square region of area n with a minimum separation of $r_{u,v} \geq r_{\min}$. The results can, however, be recast for *dense* networks, where n nodes are arbitrarily placed on a square region of unit area with a minimum separation of $r_{u,v} \geq r_{\min}/\sqrt{n}$. It suffices to notice that by rescaling power by a factor $n^{-\alpha/2}$ a dense network can essentially be transformed into an extended network with path-loss exponent α (see also [8]). Hence, the same result for dense networks can be obtained from the result for extended networks by considering the limit $\alpha \to 2$. Applying this to Theorem 1 yields a linear per-node rate scaling of the hierarchical relaying scheme.

D. Minimum-Separation Requirement

The minimum-separation requirement $r_{\min} \in (0, 1)$ on the node placement is sufficient but not necessary for Theorem 1 to hold. A weaker sufficient condition is that a constant fraction of squarelets are dense, as shown in Lemma 6 to be a consequence of the minimum-separation requirement. It is straightforward to show that this weaker condition is satisfied with high probability for nodes placed uniformly at random on $[0, \sqrt{n}]^2$. This yields a different proof of Theorem 5.1 in [8].

E. Comparison With [8]

Both the hierarchical relaying scheme presented here and the hierarchical scheme presented in [8] share that they use virtual multiple-antenna communication and a hierarchical architecture to achieve essentially global cooperation in the network. The schemes differ, however, in several key aspects, which we point out here.

First, we note that we obtain a slightly better scaling law. Namely

$$b_1(n) n^{1-\alpha/2} \leq \rho^*(n) \leq b_2(n) n^{1-\alpha/2}$$

with

$$b_1(n) \geq n^{-O(\log^{\delta - 1/2}(n))}$$
$$b_2(n) = O(\log^6(n))$$

for any $\delta \in (0, 1/2)$ obtained here, compared to

$$\tilde{b}_1(n)n^{1-\alpha/2} \leq \rho^*(n) \leq \tilde{b}_2(n)n^{1-\alpha/2}$$

with

$$\tilde{b}_1(n) = \Omega(n^{-\varepsilon})$$
$$\tilde{b}_2(n) = O(n^{\varepsilon})$$

for any $\varepsilon > 0$ in [8]. For the lower bound (i.e., achievability), this is because the hierarchy here is not of fixed depth L as in [8], but rather of depth $L(n) = \log^{1/2-\delta}(n)$ (for some constant $\delta \in (0, 1/2)$), i.e., changing with n. For the upper bound (i.e., converse), this is due to a sharpening of the arguments in [8].

Second, note that the multiuser decoding at the relay squarelets during the MAC phase and the multiuser encoding during the BC phase are very simple in our setup. In fact, using matched-filter receivers and transmit beamforming, we convert the multiuser encoding and decoding problems into several single-user decoding and encoding problems. This differs from the approach in [8], in which joint decoding of a number of users on the order of the network size is performed. Our results thus imply that these simpler transmitter and receiver structures provide the same scaling as the more complicated joint decoding in [8]. We note that the scheme proposed in [8] can be modified to also use matched-filter receivers as suggested here.

Third, and probably most important, the schemes differ in how they achieve the throughput gain from using multiple antennas. In [8], the nodes are located almost regularly with high probability. This allowed the use of a scheme in which a source squarelet directly communicates with a destination squarelet. In other words, the multiple-antenna gain comes from setting up a virtual MIMO channel between the source and the destination. In our setup, the arbitrary location of nodes prevents such an approach. Instead, we use that at least some fixed fraction of squarelets is almost regular (we called them dense squarelets). Source–destination pairs relay their traffic over such a dense squarelet. In other words, the multiple-antenna gain comes from setting up a virtual multiple-antenna MAC and BC. Thus, the hierarchical relaying scheme presented here shows that considerably less structure on the node locations than assumed in [8] suffices to achieve a multiple-antenna gain essentially on the order of the network size. Note also that the additional degree of freedom offered by the choice of relay squarelet for a given source–destination pair makes it possible to extend the result to hold also for slow-fading channels.

XIII. CONCLUSION

We considered the problem of the scaling of achievable rates in arbitrary extended wireless networks. We generalized the hierarchical cooperative communication scheme presented in [8] for a fast-fading channel model and with random node placements. We proposed a different hierarchical cooperative communication scheme, which also works for arbitrary node placement (with a minimum-separation requirement) and for either fast or slow fading.

For small path-loss exponent $\alpha \in (2, 3]$, we showed that our scheme is order optimal and achieves the same rate irrespective of the node placement. In particular, this rate is equal to the one achievable under random node placement. In other words, the regularity of the node placement has no impact on achievable rates for small path-loss exponent.

The situation is, however, quite different for large path-loss exponent $\alpha > 3$. We argued that in this regime the regularity of the node placement directly impacts the scaling of achievable rates. We then presented a cooperative communication scheme that smoothly "interpolates" between multihop and hierarchical cooperative communication depending on the regularity of the node placement. We showed that this scheme is order optimal for all $\alpha > 3$ under adversarial node placement with regularity constraint. This contrasts with the situation for more regular networks (like the ones obtained with high probability through random node placement), in which multihop communication is order optimal for all $\alpha > 3$. Thus, for less regular networks, the use of more complicated cooperative communication schemes can be necessary for optimal operation of the network.

REFERENCES

[1] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.

[2] L. Xie and P. R. Kumar, "A network information theory for wireless communication: Scaling laws and optimal operation," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 748–767, May 2004.

[3] A. Jovičić, P. Viswanath, and S. R. Kulkarni, "Upper bounds on transport capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2555–2565, Nov. 2004.

[4] O. Lévêque and İ. 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.

[5] F. Xue, L. Xie, and P. R. Kumar, "The transport capacity of wireless networks over fading channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 834–847, Mar. 2005.

[6] 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] M. Franceschetti, O. Dousse, D. 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.

[8] A. Özgür, O. Lévêque, and D. Tse, "Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3549–3572, Oct. 2007.

[9] P. Gupta and P. R. Kumar, "Towards an information theory of large networks: An achievable rate region," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1877–1894, Aug. 2003.

[10] L. Xie and P. R. Kumar, "An achievable rate for the multiple-level relay channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1348–1358, Apr. 2005.

[11] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.

[12] 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.

[13] M. Franceschetti, M. D. Migliore, and P. Minero, "The capacity of wireless networks: Information-theoretic and physical limits," in *Proc. Allerton Conf. Commun. Control Comput.*, Sep. 2007, pp. 729–735.

[14] M. Franceschetti, M. D. Migliore, and P. Minero, "The degrees of freedom of wireless networks: Information theoretic and physical limits," in *Proc. Allerton Conf. Commun. Control Comput.*, Sep. 2008, pp. 1578–1581.

[15] S. Ihara, "On the capacity of channels with additive non-Gaussian noise," *Inf. Control*, vol. 37, no. 1, pp. 34–39, Apr. 1978.

[16] S. R. Kulkarni and P. Viswanath, "A deterministic approach to throughput scaling in wireless networks," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1041–1049, Jun. 2004.

Urs Niesen received the M.S. degree from the School of Computer and Communication Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, in 2005. Currently, he is working towards the Ph.D. degree at the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology (MIT), Cambridge.

His research interests are in the area of communication and information theory.

Piyush Gupta received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Bombay, India, in 1993, the M.S. degree in computer science and automation from the Indian Institute of Science, Bangalore, India, in 1996, and the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign, Urbana, in 2000.

From 1993 to 1994, he was a Design Engineer at the Center for Development of Telematics, Bangalore. Since September 2000, he has been a Member of Technical Staff at the Mathematics of Networks and Communications Research Department, Bell Laboratories, Alcatel-Lucent, Murray Hill, NJ. His research interests include wireless networks, network information theory, and learning and adaptive systems.

Devavrat Shah received the B.Tech. degree from IIT-Bombay, India, in 1999 and the Ph.D. degree from Stanford University, Stanford, CA, in 2004.

Currently, he is a Jamieson Career Development Associate Professor at the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology (MIT), Cambridge. His research interests include statistical inference and network algorithms.

Dr. Shah was co-awarded the IEEE INFOCOM best paper award in 2004 and ACM SIGMETRICS/Performance best paper award in 2006. He received 2005 George B. Dantzig best dissertation award from the INFORMS. He is the recipient of the first ACM SIGMETRICS Rising Star Award 2008 for his work on network scheduling algorithms.