On Capacity Scaling in Arbitrary Wireless Networks

Urs Niesen, Piyush Gupta, and Devavrat Shah

Abstract

We consider the problem of characterizing per node throughput scaling in arbitrary extended wireless networks. Recently, Özgür, Lévêque, and Tse (2007) obtained a complete characterization of throughput scaling for random extended networks (i.e., nodes are placed in a square region uniformly at random) under a fast fading channel model. They proposed a hierarchical cooperative communication scheme to establish this result. However, their results (both communication scheme and proof technique) are strongly dependent on the "regularity" induced with high probability by the random node placement.

As a main result of this paper, we propose a more general (and very different) hierarchical cooperative communication scheme that works for arbitrarily placed nodes (with a minimum-separation requirement). Under our scheme, we obtain exactly the same per node throughput scaling as in Özgür et. al., showing that much less regularity is necessary for successful hierarchical cooperation. Our result holds under both fast and slow fading channel model. For small path-loss exponents $\alpha \in (2, 3]$, we show that our scheme is order optimal for all node placements with minimum-separation requirement. As a special case, we recover the results of Özgür et. al. for random networks and for both fast as well as slow fading.

For extended networks with random node placement, the following threshold phenomenon exists: for $\alpha \leq 3$, the hierarchical cooperative scheme achieves the optimal throughput scaling; for $\alpha > 3$, multi-hop communication achieves the optimal throughput scaling. We establish that for arbitrary node placement, due to the lack of "regularity", such a threshold phenomenon does not exist. In other words, there are node placements such that multi-hop communication is not order optimal for $\alpha > 3$. We then present a family of schemes that smoothly "interpolate" between multi-hop and hierarchical cooperative communication, depending upon the "level of regularity" in the node placement. We establish optimality of these schemes under adversarial node placement for $\alpha > 3$.

Finally, we show how these results on permutation traffic (i.e., n source-destination pairs) can be used to obtain an inner bound for the n^2 dimensional capacity region of the wireless network (in contrast to the one dimensional characterization for permutation traffic).

I. Introduction

We consider wireless networks with n nodes placed on $[0, \sqrt{n}]^2$ (usually referred to as *extended networks*), with each node being the source for one of n source-destination pairs and the destination for another pair. Our goal is to determine the scaling of $\rho^*(n)$, the largest achievable per node throughput, as the number of nodes n goes to infinity. This problem was first analyzed in [1], where it was shown that, under random placement of nodes in the region and certain models of communication motivated by current technology, the average per node throughput for random source-destination pairing can scale at most as $O(n^{-1/2})$ as $n \to \infty$, and that a scheme based on multi-hop communication can achieve, up to a poly-log factor, the same order of scaling. It was also shown that under arbitrary placement of nodes the network's *transport capacity*, which is the rate-distance product summed over all source-destination pairs, can scale at most as $\Theta(n)$. In particular, in an arbitrary network with source nodes picking their destinations far away, i.e., on average at a distance of $\Theta(\sqrt{n})$, the average per node throughput can scale at most as $O(n^{-1/2})$.

Since [1], the problem has received a considerable amount of attention. One stream of work [2], [3], [4], [5], [6] has progressively broadened the conditions on the channel model, the communication model, and

U. Niesen and D. Shah are with the Laboratory of Information and Decision Systems, Department of EECS at the Massachusetts Institute of Technology. Email: {uniesen,devavrat}@mit.edu

P. Gupta is with the Complex Systems Analysis and Optimization Department, Bell Labs, Alcatel-Lucent. Email: pgupta@research.bell-labs.com

The work of U. Niesen and D. Shah was supported in parts by DARPA grant (ITMANET) 18870740-37362-C and NSF grant CNS-0546590; the work of P. Gupta was supported in part by NSF Grants CCR-0325673 and CNS-0519535.

the node placements under which multi-hop 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 $\alpha > 4$ and "sufficiently regular" node placement, the best achievable per node throughput for a far away destination scales like $\rho^*(n) = \Theta(n^{-1/2})$ as $n \to \infty$ and that this scaling is achievable with multi-hop communication.

Another stream of work [7], [8], [9], [10], [11] has proposed progressively refined multi-user cooperative schemes, which have been shown to significantly out-perform multi-hop communication in many environments. In an exciting recent work, Özgür et. al. [11] have shown that for a fast fading wireless channel model with full channel state information (CSI), with nodes placed uniformly at random, and with *low* signal attenuation $2 < \alpha \leq 3$, a hierarchical cooperative communication scheme can perform significantly better. More precisely, they show that for $2 < \alpha \leq 3$, the best achievable per node rate for random source-destination pairing scales as $\rho^*(n) = O(n^{1-\alpha/2+\varepsilon})$ and hierarchical cooperative communication achieves a per node throughput of $\Omega(n^{1-\alpha/2-\varepsilon})$ (here, $\varepsilon > 0$ is an arbitrary but fixed constant). That is, hierarchical communication is (essentially) order optimal in the attenuation regime $\alpha \in (2, 3]$. For $\alpha > 3$, the multi-hop communication scheme achieves the optimal throughput scaling of $\Omega(\sqrt{n})$ and this is optimal in the sense that the best possible throughput scaling is $\rho^* = O(n^{1/2+\varepsilon})$ for any fixed constant $\varepsilon > 0$. In summary, for random extended networks with random source-destination pairs, the optimal communication scheme exhibits the following threshold behavior: for $\alpha \in (2, 3]$ the hierarchical cooperation communication scheme is order optimal, while for $\alpha > 3$ the multi-hop communication scheme is order optimal.

A. Our Contributions

Our interest is in understanding the scaling behavior of the capacity region of arbitrary wireless networks. In previous work, there have been two restrictive assumptions: (a) the node placement is uniformly at random, and (b) data needs to be communicated only between n source-destination pairs. We relax both of these restrictions.

To address the first issue, we consider the setup with arbitrary node placement (with minimum-separation constraint) and arbitrary n source-destination pairs obtained through a permutation of the n nodes (i.e. each node is source and destination for exactly one pair, and there are n such source-destination pairs). We present a hierarchical communication scheme (which we call the hierarchical relaying scheme in the following) that achieves, under either fast or slow fading, and for any path loss exponent $\alpha > 2$, a transmission rate per source-destination pair of $\rho^{\text{HR}}(n) \geq n^{1-\alpha/2-\beta(n)}$, where $\beta(n) = O(\log^{-1/3} n)$. Further, for $\alpha \in (2, 3]$, we show that the rate of the best communication scheme is upper bounded as $\rho^*(n) = O(n^{1-\alpha/2+\varepsilon})$ for any fixed $\varepsilon > 0$. Thus, our hierarchical communication scheme is (essentially) order optimal for any such arbitrary network for $\alpha \in (2, 3]$. We note that our result recovers the result of Özgür et. al. [11] for random networks and fast fading as a special case. Moreover, our result also holds under slow fading, proving a conjecture in [11].

For $\alpha > 3$ and under random node placement the multi-hop scheme is optimal, as stated earlier. However, we show that under arbitrary node placement the hierarchical scheme can outperform the multi-hop scheme for certain node placements even for $\alpha > 3$. This suggests that for arbitrary networks there is no simple threshold value of α above which multi-hop communication is order optimal. It also suggests that the order optimality of multi-hop schemes for random networks is strongly dependent on the "regularity" (induced with high probability by the random construction) of the node placement. In other words, for less regular networks we need more complicated cooperative communication schemes to achieve optimal throughput scaling. Towards that end, we present a family of communication schemes that smoothly "interpolate" between hierarchical cooperative communication and multi-hop communication, and in which nodes communicate at scales that vary smoothly from local to global. The amount of "interpolation" between hierarchical and multi-hop scheme 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.

To address the second issue, we consider more general traffic demands. More precisely, we consider general multi-commodity flows. We show how the results derived for permutation traffic (i.e., uniform traffic between n source-destination pairs) can be used to find achievable rates for these more general traffic demands. This in turn leads to an inner bound on the n^2 dimensional capacity region of the wireless network.

B. Organization

The remainder of this paper is organized as follows. Section II describes in detail the communication model. In Section III, we provide formal statements of our results. Sections IV and V describe our new hierarchical scheme for arbitrary wireless networks with its detailed analysis. Sections VI and VII state a converse results to establish (essentially) order optimality of our scheme. We recover results of Özgür et. al. [11] for random node placement as a special case in Section VIII. In Section IX, we present our "interpolation" scheme. In Section X, we derive implications of our results on the characterization of the complete capacity region. Finally, Sections XI and XII contain discussions and concluding remarks.

II. Model

We use the following notations. 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. To simplify notation, we assume, when necessary, that fractions are integers and omit $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ operators.

Consider $A(n) \triangleq [0, \sqrt{n}]^2$ and let $V(n) \subset A(n)$ be a set of $|V(n)| = n$ nodes on $A(n)$. We say that $V(n)$ has *minimum-separation* c if $r_{u,v} \geq c$ for all $u, v \in V(n)$, where $r_{u,v}$ is the Euclidean distance between u and v. We use the same channel model as in [11]. Namely, if $\{x_u[t]\}_{u,t}$ are the (sampled) signals sent by the nodes in $V(n)$, then the (sampled) received signal at node v is

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

for all $v \in V(n)$. Here $\{z_v[t]\}_{v,t}$ are i.i.d. with distribution $\mathcal{N}_{\mathbb{C}}(0, 1)$ (i.e., circularly symmetric complex Gaussian with mean 0 and variance 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 either assume $\{\theta_{u,v}[t]\}_{u,v,t}$ i.i.d. with uniform distribution on $[0, 2\pi)$, which is called *fast fading* in the following, or we assume $\{\theta_{u,v}[t]\}_{u,v}$ i.i.d. with uniform distribution on $[0, 2\pi)$ and constant as a function of t, which is called *slow fading* in the following. In either case, we assume full CSI, i.e., each node knows all $\{h_{u,v}[t]\}_{u,v}$ at time t. We also impose an average power constraint of P 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 some fixed rate $\rho(n)$ to some other node $v \in V(n)$. We call u the *source* and v the *destination* node of this communication pair. The set of all communication pairs can be described by a *traffic matrix* $T(n) \in \{0, 1\}^{n \times n}$. We say that $T(n)$ is *valid* if it is a permutation matrix (i.e., every node is a source for exactly one communication pair and a destination for exactly one communication pair). For a traffic matrix $T(n)$, let $\rho^*(n)$ be the highest achievable per node rate.

III. Main Results

This section presents the formal statement of our results. The results are divided into three parts. In the first part, we present a hierarchical 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 $\alpha \in (2, 3]$, and that it is order optimal under adversarial node placement when $\alpha > 3$. In the second part, we presents a communication scheme that "interpolates" between the hierarchical and the multi-hop scheme 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 $\alpha > 3$. In the third part, we present how these results on permutation traffic can be used to obtain an inner bound on the n^2 dimensional capacity region of the wireless network.

A. Hierarchical Relaying Scheme

Theorem 1. *Under fast fading, for any $\alpha > 2$, for any sequences of node placements $\{V(n)\}_{n\geq 1}$ with minimum separation $c > 0$ and valid traffic matrices $\{T(n)\}_{n\geq 1}$, we have*

$$\rho^*(n) \geq \rho^{HR}(n) \geq n^{1-\alpha/2-\beta(n)},$$

where $\beta(n) = O(\log^{-1/3} n)$ as $n \to \infty$. The same conclusion holds for slow fading with probability $1 - o(1)$.

The proof of this theorem relies on the construction of a sophisticated hierarchical communication scheme and on bounding the rate $\rho^{\mathrm{HR}}(n)$ that it achieves. Theorem 1 is first established for the fast fading case in Section V-D by Theorem 11, and then for the slow fading case in Section V-E by Theorem 13.

We note that Theorem 1 remains valid under somewhat weaker conditions than having minimum separation $c > 0$. Specifically, we show that the result of Özgür et. al. [11] can be recovered through (arguments of) Theorem 1 as the random node placement satisfies these weaker conditions. We discuss this in more detail in Section VIII.

The following theorem establishes optimality of the hierarchical scheme in the range of $\alpha \in (2, 3]$ for arbitrary node placement. The proof of Theorem 2 is presented in Section VI. It is a fairly straightforward adaptation of the arguments of [11, Theorem 5.2].

Theorem 2. *Under either fast or slow fading, for any $\alpha > 2$, for any sequence of node placements $\{V(n)\}_{n\geq 1}$ with minimum separation $c > 0$, and for $\{T(n)\}_{n\geq 1}$ uniformly distributed over the set of all valid traffic matrices, we have for any $\varepsilon > 0$*

$$\rho^*(n) = \begin{cases} O(n^{1-\alpha/2+\varepsilon}) & \text{for } 2 < \alpha \leq 3, \\ O(n^{-1/2+\varepsilon}) & \text{for } \alpha > 3, \end{cases}$$

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

Comparing Theorems 1 and 2, we see that for $2 < \alpha \leq 3$ the proposed communication scheme is order optimal. In the case of randomly distributed $V(n)$, multi-hop communication achieves $\rho^{\mathrm{MH}}(n) = \Omega(n^{-1/2})$ with probability $1 - o(1)$ and hence Theorem 2 shows that multi-hop is order optimal for $\alpha > 3$. For arbitrarily placed nodes, this is, however, not the case.

Indeed, Theorem 3 below shows that for arbitrarily placed nodes the hierarchical relaying scheme is order optimal for *all* $\alpha > 2$ under adversarial node placement with minimum-separation constraint. In fact, comparing Theorem 1 and Theorem 3, we see that there exist node placements for which hierarchical relaying achieves a rate of at least a factor of n higher than multi-hop communication for *all* $\alpha > 2$.

Theorem 3. *Under either fast or slow fading, there exist sequences of node placements $\{V(n)\}_{n\geq 1}$ with minimum separation $c > 0$ such that for $\{T(n)\}_{n\geq 1}$ uniformly distributed over the set of all valid traffic matrices, we have for any $\alpha > 3$ and $\varepsilon > 0$*

$$\rho^*(n) = O(n^{1-\alpha/2-\varepsilon}),$$
$$\rho^{MH}(n) = O(n^{-\alpha/2}),$$

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

B. Cooperative Multi-Hop Scheme

Theorem 3 suggests that it is the level of regularity of the topology that decides what scheme to choose for path loss exponent $\alpha > 3$. More precisely, we show that an appropriate "interpolation" between the hierarchical relaying and the multi-hop communication schemes is required for $\alpha > 3$ in order to achieve the optimal performance as established in the result stated below.

Before we state that result, we need to introduce some notation. Consider again $V(n) \subset A(n) \triangleq [0, \sqrt{n}]^2$ nodes with minimum separation $c > 0$. Divide $A(n)$ into squares of sidelength $h(n) \leq \sqrt{n}$, and fix a (possibly large) constant $K \geq 1$. We say that $V(n)$ *is regular at resolution* $h(n)$ if every such square contains at least $h(n)^2/K$ nodes. Note that every $V(n)$ is trivially regular at resolution $\sqrt{n}$. The following statement, bounding the rate $\rho^{\text{CMH}}(n)$ achievable with cooperative multi-hop communication, follows from Theorems 16 and 17 in Section IX.

Theorem 4. *Under either slow or fast fading, for any $\alpha > 2$, for any sequences of node placements $\{V(n)\}_{n\geq 1}$ with minimum separation $c > 0$ and valid traffic matrices $\{T(n)\}_{n\geq 1}$, we have*

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

as $n \to \infty$, where

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

and with $\beta(n) = O(\log^{-1/3} n)$. Further, there exists sequences of node placements $\{V(n)\}_{n\geq 1}$ with minimum separation $c > 0$ and regular at resolution $h^*(n)$ such that for $\{T(n)\}_{n\geq 1}$ uniformly distributed over the set of all valid traffic matrices, we have for any $\alpha > 3$ and $\varepsilon > 0$,

$$\rho^*(n) = O\big(h^*(n)^{3-\alpha} n^{-1/2+\varepsilon}\big),$$
$$\rho^{MH}(n) = O(h^*(n)^{-\alpha}),$$

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

As an example, assume than

$$\lim_{n\to\infty} \frac{h^*(n)}{\log n} = \eta$$

for some $\eta \geq 0$ (i.e. $h^*(n) = n^{\eta(1+o(1))}$). Then Theorem 4 shows that

$$\liminf_{n\to\infty} \frac{\rho^{\text{CMH}}(n)}{\log n} \geq (3-\alpha)\eta - 1/2,$$

and there exist node placements such that

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

C. Inner Bound on the Capacity Region

So far, we have only consider permutation traffic, i.e., traffic needs be sent at uniform rate between n source-destination pairs. We show how the results presented so far for permutation traffic can be used to find an inner bound on the n^2 dimensional capacity region of the wireless network.

Before proceeding further, we formally define the capacity region. We call $\lambda \in \mathbb{R}_+^{n\times n}$ *achievable* if there exists a communication scheme that allows for simultaneous reliable transmission of data at rate $\lambda_{u,v}$ from node u to v, for all $u, v \in V(n)$. Define the *capacity region*

$$\Lambda(n) \triangleq \text{cl}\{\lambda \in \mathbb{R}_+^{n\times n} : \lambda \text{ is achievable}\},$$

where $\text{cl}(B)$ is the closure of B. The null-rate vector $\mathbf{0}$ is contained in $\Lambda(n)$, and by the standard time sharing argument $\Lambda(n)$ is a convex set. For any $\lambda \in \mathbb{R}_+^{n\times n}$ with $\lambda \neq \mathbf{0}$, define (with slight abuse of notation)

$$\rho^*_\lambda(n) \triangleq \sup\{b > 0 : b\lambda \in \Lambda(n)\}.$$

That is, $\rho^*_\lambda(n)$ is the maximum distance that can be traveled along "ray" λ starting from $\mathbf{0}$ before hitting the boundary of $\Lambda(n)$. Since $\Lambda(n)$ is a closed convex set, $\{\rho^*_\lambda(n) : \lambda \in \mathbb{R}_+^{n\times n}\}$ characterizes $\Lambda(n)$ completely. Note that $\rho^*(n)$ for a given valid traffic matrix $T(n)$ is precisely $\rho^*_T(n)$.

Our next result provides an inner bound on $\Lambda(n)$ using only information on achievable per node rates under permutation traffic (i.e., $\rho_T^*(n)$ for a valid traffic matrix $T(n)$). In the following we let $\mathbf{1} \in \mathbb{R}^{n\times n}$ be the matrix of all ones.

Theorem 5. *Consider any node placement $V(n) \in \mathbb{R}^2$ and any set of valid traffic matrices $\{T_i(n)\}_{i=1}^n$ such that*

$$\mathbf{1} = \sum_{i=1}^{n} T_i(n).$$

Then for any $\lambda \in \mathbb{R}_+^{n\times n}$, and any channel model,

$$\rho_\lambda^*(n) = \Omega\biggl(\frac{\min_i \rho_{T_i}^*(n)}{\log n}\Bigl(\min_{S\subset V(n):|S|\leq n/2} \frac{|S|}{D_\lambda(S)}\Bigr)\biggr),$$

where $D_\lambda(S) = \sum_{u\in S; v\notin S}(\lambda_{u,v} + \lambda_{v,u})$.

Theorems 1 and 4 yield uniform lower bounds on $\rho_{T_i}^*(n)$, which can be used to further lower bound $\min_i \rho_{T_i}^*(n)$. For example, combining Theorem 1 with Theorem 5, and using that we can write $\mathbf{1}$ as the sum of the identity matrix and its n circular shifts, shows that for either the fast or slow fading channel model for any $\alpha > 2$

$$\rho_\lambda^*(n) = \Omega\biggl(n^{1-\alpha/2-\beta(n)}\Bigl(\min_{S\subset V(n):|S|\leq n/2} \frac{|S|}{D_\lambda(S)}\Bigr)\biggr),$$

where the statement holds with probability $1 - o(1)$ for slow fading.

IV. Description of the Hierarchical Relaying Scheme

This section describes the architecture of our hierarchical scheme. Let $A(b) \triangleq [0, \sqrt{b}]^2$ be the square region of area b. The scheme described here assumes that n nodes are placed arbitrarily in $A(n)$ with minimum separation $c > 0$. We want to find some rate, say ρ_0, that can be supported for all n source-destination pairs of a given valid traffic matrix $T(n)$. 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 $a_1 \triangleq n/\gamma(n)$ for any valid traffic matrix $T(n_1)$ of n_1 nodes. Here $\gamma(n)$ 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 valid traffic matrix $T(n_\ell)$ over $n_\ell \triangleq \frac{n}{2^\ell \gamma(n)^\ell}$ nodes in $A(a_\ell)$, with $a_\ell \triangleq n/\gamma(n)^\ell$, use a scheme achieving rate $\rho_{\ell+1}$ over $n_{\ell+1}$ nodes in $A(a_{\ell+1})$ for any valid traffic matrix $T(n_{\ell+1})$. The recursion (or hierarchy) is terminated at some level $L(n)$ to be chosen later.

A. Hierarchical Relaying Scheme: Construction

We describe how the hierarchy is constructed between levels ℓ and $\ell + 1$ for $0 \leq \ell < L(n)$. This has three phases as described below. This construction, at a high level, is depicted in Figure 1.

Phase 1: Setting up Relays. Given n_ℓ nodes in $A(a_\ell)$, divide it into $\gamma(n)$ ($1 \leq \gamma(n) \leq n_\ell$) 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. We show in Lemma 6 that since the nodes in $A(a_\ell)$ have minimum separation $c > 0$, a squarelet can contain at most $O(a_{\ell+1})$ (i.e. $O(a_\ell/\gamma(n))$) nodes. This is shown to imply that there are $\Theta(\gamma(n))$ 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. The choice of relays is done such that each relay squarelet has at most $n_{\ell+1}$ communication pairs that use it as relay (see Lemma 7), and we assume this worst case in the following discussion.

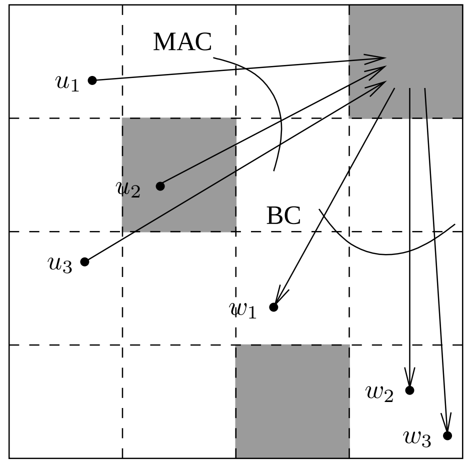

Fig. 1. Sketch of level ℓ of the hierarchical relaying scheme. Groups of source-destination pairs relay their traffic over dense relay squarelets (shaded). We time share between the different relay squarelets. Within all dense relay squarelets the scheme at level $\ell+1$ is used simultaneously to enable joint decoding and encoding at each relay.

Phase 2: Multiple Access. Source nodes that are assigned to the same (dense) relay squarelet in the first phase, send their messages simultaneously to that relay. We time share between the $\Theta(\gamma(n))$ different relay squarelets. If the nodes in the relay squarelet could cooperate, we would be dealing with a multiple access channel (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 construction. This is depicted in Figure 2.

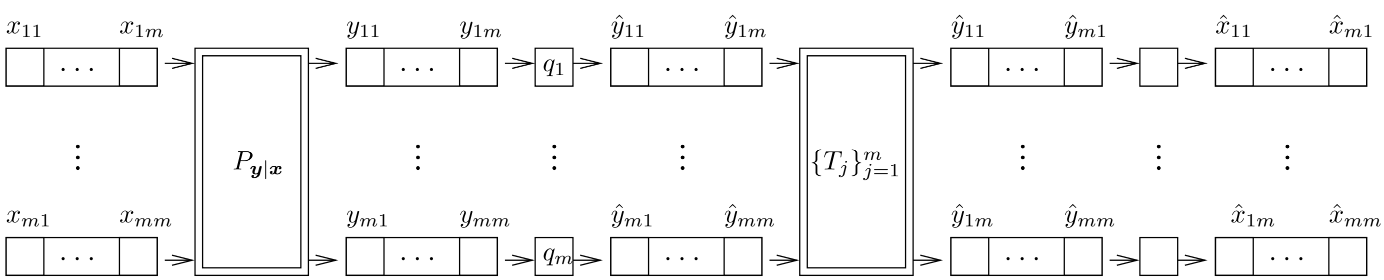

Fig. 2. Description of the MAC phase at level ℓ in the hierarchy with $m \triangleq n_{\ell+1}$. The first system block represents the wireless channel, connecting source nodes $\{u_i\}_{i=1}^{n_{\ell+1}}$ with relay nodes $\{v_i\}_{i=1}^{n_{\ell+1}}$. The second system block are quantizers $\{q_i\}_{i=1}^{n_{\ell+1}}$ used at the relay nodes. The third system block represents using $n_{\ell+1}$ times the communication scheme at level $\ell+1$ (organized as $n_{\ell+1}$ valid traffic matrices $\{T_j(n_{\ell+1})\}_{j=1}^{n_{\ell+1}}$). The fourth system block are matched filters used at the relay nodes.

Suppose there are $n_{\ell+1}$ sources, say $u_1, \dots, u_{n_{\ell+1}}$ and $n_{\ell+1}$ relay nodes, say $v_1, \dots, v_{n_{\ell+1}}$. 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 shall, 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} at time j over $n_{\ell+1}$ time slots. 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 $\{y_{kj}\}_{k=1}^{n_{\ell+1}}$ 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 $\{\hat{y}_{kj}\}_{j=1}^{n_{\ell+1}}$. Now, quantized observation $\hat{y}_{kj}$ is 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 can be organized as $n_{\ell+1}$ valid traffic matrices $\{T_j(n_{\ell+1})\}_{j=1}^{n_{\ell+1}}$ between the $n_{\ell+1}$ relay nodes. Note that these relay nodes are in a square of area $a_{\ell+1}$. Therefore, using $n_{\ell+1}$ times the scheme for transmitting according to a valid traffic matrix for $n_{\ell+1}$ nodes in $A(a_{\ell+1})$, relay v_j can obtain all quantized observations $\{\hat{y}_{kj}\}_{i=1}^{n_{\ell+1}}$. Now v_j uses $n_{\ell+1}$ matched filters on $\{\hat{y}_{kj}\}_{k=1}^{n_{\ell+1}}$ to obtain estimates $\{\hat{x}_{ij}\}_{i=1}^{n_{\ell+1}}$ of $\{x_{ij}\}_{i=1}^{n_{\ell+1}}$. Using these estimates it then decodes the messages corresponding to $\{x_{ij}\}_{i=1}^{n_{\ell+1}}$. The achievable rates, and the sufficiency of quantization at constant rate in this setup, are analyzed in Lemma 9.

Phase 3: Broadcast. 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 broadcast channel (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 construction as for the MAC phase is used. This construction is depicted in Figure 3.

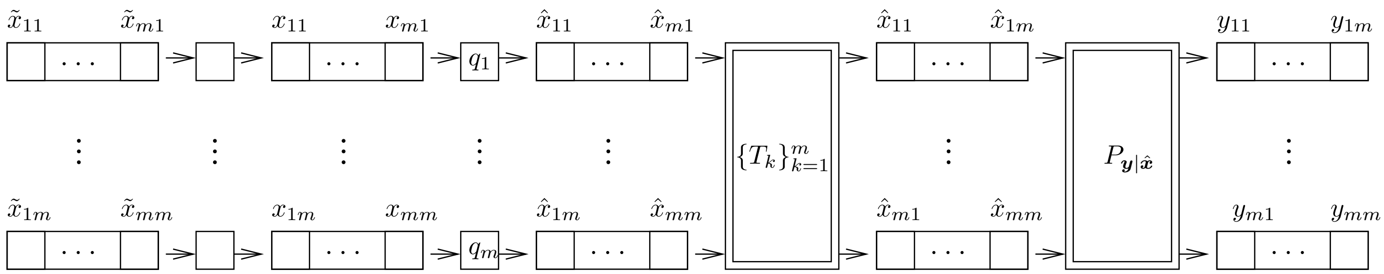

Fig. 3. Description of the BC phase at level ℓ in the hierarchy with $m \triangleq n_{\ell+1}$. The first system block represents transmit beamforming at each of the relay nodes $\{v_i\}_{i=1}^{n_{\ell+1}}$. The second system block are quantizers $\{q_j\}_{j=1}^{n_{\ell+1}}$ used at the relay nodes. The third system block represents using $n_{\ell+1}$ times the communication scheme at level $\ell+1$ (organized as $n_{\ell+1}$ valid traffic matrices $\{T_k(n_{\ell+1})\}_{k=1}^{n_{\ell+1}}$). The fourth system block is the wireless channel, connecting relay nodes $\{v_i\}_{i=1}^{n_{\ell+1}}$ with destination nodes $\{w_i\}_{i=1}^{n_{\ell+1}}$.

Suppose there are $n_{\ell+1}$ relay nodes, say $v_1, \dots, v_{n_{\ell+1}}$ and $n_{\ell+1}$ destination nodes, say $w_1, \dots, w_{n_{\ell+1}}$. 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 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 $\{v_k\}_{k=1}^{n_{\ell+1}}$. Call x_{kj} the resulting channel symbol to be sent from relay node v_k. In order to actually send this channel symbol, relay node v_k needs to obtain x_{kj}. Thus, again all relay nodes need to exchange information. For this, each relay node v_j quantizes its beamformed channel symbols $\{x_{ij}\}_{i=1}^{n_{\ell+1}}$ at an appropriate rate $K\log(n)$ with K independent of n to obtain $\{\hat{x}_{ij}\}_{i=1}^{n_{\ell+1}}$. Now, quantized value $\hat{y}_{kj}$ 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 can be organized as $n_{\ell+1}$ valid traffic matrices $\{T_k(n_{\ell+1})\}_{k=1}^{n_{\ell+1}}$ between the $n_{\ell+1}$ relay nodes. Note that these relay nodes are in a square of area $a_{\ell+1}$. Therefore, using $n_{\ell+1}$ times the scheme for transmitting according to a valid 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 j. Call y_{ij} the received channel output at destination node w_i at time instance j. Using y_{ij}, destination node w_i can now decode part j of the message bits of its source node u_i. The achievable rates, and the sufficiency of quantization at logarithmic rate in this setup, are analyzed in Lemma 10.

Spatial Re-Use. 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 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 is discussed in the proofs.

B. Hierarchical Relaying Scheme: Termination

The scheme terminates at some large enough level $L = L(n)$ (to be chosen later). At this scale, we have n_L nodes in area a_L. A valid 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.

C. Hierarchical Relaying Scheme: Achievable Rates

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

We compute the time utilized for transmission of a single message bit from each source to its destination under a valid traffic matrix $T(n)$. With the above scheme, the data travels through $L(n)$ 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 re-use (described in more detail in Section V-D), each of these regions gets to transmit a constant fraction of time. It can be shown (see again Section V-D) that the addition of interference due to this spatial re-use leads only to a constant loss in achievable throughput. 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 construction, only one of its $\Theta(\gamma(n))$ dense relay squarelets of area $a_{\ell+1}$ is active at any given moment. Hence the time required to send one message bit is a $\gamma(n)$ factor higher than the one needed if we consider just one relay squarelet separately. Consider one such relay squarelet, and assume $n_{\ell+1}$ source nodes communicate each $n_{\ell+1}$ message bits to their respective destination nodes through a MAC and BC phase with the help of the $n_{\ell+1}$ relay nodes in this relay squarelet.

In the MAC phase, each of the $n_{\ell+1}$ sources simultaneously sends one bit to each of the $n_{\ell+1}$ relay nodes. As proved in detail in Lemma 9, the total time for this transmission is composed of the following:

(i) Transmission of $n_{\ell+1}$ message bits from each of the $n_{\ell+1}$ source nodes to those many relay nodes, requiring a total of

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

channel uses. Here we used an average power constraint of $\Theta(2^{-\ell}\gamma(n))$. Since we time share between $\Theta(2^{-\ell}\gamma(n))$ relay squarelets, this satisfies the overall average power constraint. The terms on the left-hand side of (2) can be understood as follows: $a_\ell^{\alpha/2}$ is the power loss since most nodes communicate over a distance of $\Theta(a_\ell)$; $2^{-\ell}\gamma(n)$ is the average transmit power; $n_{\ell+1}$ is the multiplexing gain, since we have that many transmit and receive antennas.

(ii) The $n_{\ell+1}$ bits for all sources generate $O(n_{\ell+1})$ transmissions at level $\ell+1$ of the hierarchy (due to sufficiency of constant rate quantization as proved in Lemma 9). Therefore,

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

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

Combining (2) and (3), accounting for the factor of $2^{-\ell}\gamma(n)$ 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

$$\tau_\ell^{\text{MAC}}(n) = O\big(2^\ell\gamma(n)^{\ell(1-\alpha/2)+1}n^{\alpha/2-1} + \tau_{\ell+1}(n)\big). \tag{4}$$

Next, we compute number of channel uses per message bit received by the destination nodes in the BC phase. Similar to 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 total time computation 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 hierarchy at level $\ell+1$. Lemma 10 shows that each message bit results in $O\big((\ell+1)\log n\big)$ quantized bits. Therefore, $O\big(n_{\ell+1}(\ell+1)\log n\big)$ bits need to be transmitted from each relay node. This requires
$$O\big(n_{\ell+1}(\ell+1)\log(n)\tau_{\ell+1}(n)\big) \tag{5}$$
channel uses.

(ii) From Lemma 10, the transmission of $n_{\ell+1}$ message bits from the relay nodes to each destination node, requires
$$O\left(n_{\ell+1}\frac{a_\ell^{\alpha/2}}{2^{-\ell}\gamma(n)n_{\ell+1}}\right) = O\big(n_{\ell+1}4^\ell\gamma(n)^{\ell(1-\alpha/2)}n^{\alpha/2-1}\big) \tag{6}$$
channel uses. We again used a average power constraint of $\Theta(2^{-\ell}\gamma(n))$, which satisfies the overall average power constraint by time sharing between the relay squarelets. As before $a_\ell^{\alpha/2}$ in the left hand side of (6) can be understood as the power loss for communicating over distance $\Theta(a_\ell)$, $2^{-\ell}\gamma(n)$ as the average transmit power, and $n_{\ell+1}$ as the multiplexing gain.

Combining (5) and (6), accounting for factor of $2^{-\ell}\gamma(n)$ loss due to time division between relay squarelets, we obtain that the transmission time for one message bit from the relays to each destination node in the BC phase at level ℓ is
$$\tau_\ell^{\text{BC}}(n) = O\big(2^\ell\gamma(n)^{\ell(1-\alpha/2)+1}n^{\alpha/2-1} + (\ell+1)\log(n)\tau_{\ell+1}(n)\big). \tag{7}$$
From (4) and (7), we obtain the following recursion
$$\begin{aligned}\tau_\ell(n) &= \tau_\ell^{\text{MAC}}(n) + \tau_\ell^{\text{BC}}(n)\\ &= O\big(2^\ell\gamma(n)^{\ell(1-\alpha/2)+1}n^{\alpha/2-1} + (\ell+1)\log(n)\tau_{\ell+1}(n)\big)\\ &= O\big(2^\ell\gamma(n)n^{\alpha/2-1} + (\ell+1)\log(n)\tau_{\ell+1}(n)\big),\end{aligned} \tag{8}$$
where we have used $\alpha > 2$.

The above recursion holds for all $0 \le \ell < L = L(n)$. At level L, we use TDMA among n_L nodes in region $A(a_L)$ with a valid traffic matrix $T(n_L)$. Each of the n_L pair uses the wireless channel for $1/n_L$ fraction of the time at power $O(n_L)$. Assuming the received power is less than one for all n (so that we operate in the power limited regime), we can achieve a rate $\Omega(a_L^{-\alpha/2})$ between any source-destination pair. Equivalently
$$\tau_L(n) = O(a_L^{\alpha/2}) = O\big(n^{\alpha/2}\gamma(n)^{-L\alpha/2}\big). \tag{9}$$
Combining (8) and (9), we have
$$\begin{aligned}\tau_0(n) &\le O\big(n^{\alpha/2-1}\gamma(n) + \log(n)\tau_1(n)\big)\\ &\le \dots\\ &\le O\Big(n^{\alpha/2-1}\big(L\log(n)\big)^L\big(2^L\gamma(n) + n\gamma(n)^{-L\alpha/2}\big)\Big).\end{aligned} \tag{10}$$
Now, choose
$$\begin{aligned}L \triangleq L(n) &\triangleq \log^{1/3}(n),\\ \gamma(n) &\triangleq n^{2/L(n)\alpha}.\end{aligned}$$
With this
$$n\gamma(n)^{-L(n)\alpha/2} = 1,$$

and (10) yields

$$\tau_0(n) \le n^{\alpha/2-1} \exp\big(O(\log^{2/3} n)\big)$$
$$= n^{\alpha/2-1+\beta(n)},$$

where $\beta(n) = O(\log^{-1/3} n)$. That is, the per node rate of the hierarchical relaying scheme is lower bounded as

$$\rho^{\text{HR}}(n) = 1/\tau_0(n) \ge n^{1-\alpha/2-\beta(n)}.$$

V. Analysis of the Hierarchical Relaying Scheme

Throughout Sections V-A to V-C, we analyze in detail communication at level ℓ, $0 \le \ell < L = L(n)$, 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 $a_\ell \triangleq n/\gamma(n)^\ell$ containing $n_\ell \triangleq \frac{n}{2^\ell \gamma(n)^\ell}$ nodes $V(n_\ell)$. We divide $A(a_\ell)$ into $\gamma(n)$ squarelets of size $a_{\ell+1}$. Recall that a squarelet of size $a_{\ell+1}$ in level ℓ of the hierarchy is called dense if it contains at least $n_{\ell+1}$ nodes. We impose a power constraint of $\widetilde{P}_\ell(n) = \Theta(2^{-\ell}\gamma(n))$ during the time any particular relay squarelet operates. Since we time share between $\Theta(2^{-\ell}\gamma(n))$ relay squarelets, this satisfies the overall average power constraint P.

Since other regions of size a_ℓ are active at the same time as the one under consideration, we have to deal with interference. To this end, 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 the channel gains $\{h_{u,v}[t]\}_{v,t}$, that it is i.i.d. across time t, but may be dependent across nodes u, and that it 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 one in the original model), as well as interference. We show in Section V-D that these assumptions are valid.

We make the following choice of $\gamma(n)$ and $L(n)$:

$$\begin{aligned} L(n) &\triangleq \log^{1/3}(n), \\ \gamma(n) &\triangleq n^{2/L(n)\alpha}. \end{aligned} \tag{11}$$

This satisfies

$$\begin{aligned} \gamma(n) \le \gamma(\tilde{n}) \qquad & \text{if } n \le \tilde{n}, \\ 1 \le \gamma(n)^{L(n)} \le n \qquad & \text{for all } n, \\ 2^{-L(n)}\gamma(n) \to \infty \qquad & \text{as } n \to \infty. \end{aligned} \tag{12}$$

Throughout Sections V-A to V-D, we consider the fast fading channel model. Slow fading is discussed in Section V-E.

A. Phase 1: Setting up Relays

The first lemma states that the minimum-separation requirement $c > 0$ 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 below holds directly. See also Section VIII for further details.

Lemma 6. *For any $V(n_\ell) \subset A(a_\ell)$ with minimum separation $c > 0$, each squarelet 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 $c/2$ around each node. By the minimum-separation requirement, these circles do not intersect. Each node covers an area of $\pi c^2/4$. Increasing the sidelength of each squarelet by c, this provides a total area of

$$\big(\sqrt{a_\ell/\gamma(n)} + c\big)^2 \le \frac{a_\ell}{\gamma(n)}(1+c)^2$$

in which these nodes are packed. Here we have used that by (12), $\gamma(n)^{\ell+1} \leq n$ and hence

$$\gamma(n) \leq n/\gamma(n)^{\ell} = a_\ell.$$

Hence there can be at most $K_1 a_\ell/\gamma(n)$ nodes per squarelet with

$$K_1 \triangleq 4\frac{(1+c)^2}{(\pi c^2)}.$$

We assume that c is small enough such that $K_1 2^\ell \geq 1$ since otherwise no $V(n_\ell)$ with minimum separation c exists.

Since each squarelet contains at most $K_1 a_\ell/\gamma(n)$ nodes, the number of dense squarelets $d(n_\ell)$ must satisfy

$$d(n_\ell)K_1 a_\ell/\gamma(n) + \big(\gamma(n) - d(n_\ell)\big)n_{\ell+1} \geq n_\ell.$$

Thus, using $a_\ell = 2^\ell n_\ell$, $n_{\ell+1} = n_\ell/2\gamma(n)$, and $K_1 2^\ell \geq 1$, we have

$$d(n_\ell) \geq \frac{1-1/2}{K_1 2^\ell - 1/2}\gamma(n) \geq \frac{2^{-\ell}}{2K_1}\gamma(n) \triangleq K_2 2^{-\ell}\gamma(n).$$

□

Choose arbitrary $K_2 2^{-\ell}\gamma(n)$ dense squarelets (as guaranteed by Lemma 6). We can assume without loss of generality that those are the squarelets $\{A_k(a_{\ell+1})\}_{k=1}^{K_2 2^{-\ell}\gamma(n)}$. With slight abuse of notation, call $r_{u,A_k(a_{\ell+1})}$ the distance between node $u \in V(n_\ell)$ and the closest point in $A_k(a_{\ell+1})$. Define the sets

$$\begin{aligned}\mathcal{S}(n_\ell) \triangleq \Big\{ S \in \{0,1\}^{n_\ell \times K_2 2^{-\ell}\gamma(n)} : \ & 0 \leq \textstyle\sum_{u=1}^{n_\ell} S_{u,k} \leq n_{\ell+1} \forall k, \\ & 0 \leq \textstyle\sum_{k=1}^{K_2 2^{-\ell}\gamma(n)} S_{u,k} \leq 1 \forall u, \\ & S_{u,k} = 1 \text{ implies } r_{u,A_k(a_{\ell+1})} \geq \sqrt{2a_{\ell+1}} \Big\}\end{aligned}$$

and

$$\widetilde{\mathcal{S}}(n_\ell) \triangleq \big\{ S \in \{0,1\}^{K_2 2^{-\ell}\gamma(n) \times n_\ell} : S^T \in \mathcal{S} \big\}.$$

Next, we prove that any node placement that satisfies Lemma 6 allows for a decomposition of any valid traffic matrix $T(n_\ell)$ as a convex combination of a constant number of schedules belonging to $\mathcal{S}(n_\ell)$ and $\widetilde{\mathcal{S}}(n_\ell)$. This is reminiscent of the classical Hall's theorem characterizing the existence of a perfect matching in a bipartite graph.

Lemma 7. *Let $K_3 \triangleq 4/K_2$, and $n_0 = n_0(\ell)$ such that $\gamma(n_0) \geq 42K_2^{-1}2^{\ell+1}$. Then for every $n_\ell \geq n_0(\ell)$ and every valid traffic matrix $T(n_\ell) \in \{0,1\}^{n_\ell \times n_\ell}$ there are $\{S^{(i)}(n_\ell)\}_{i=1}^{K_3 2^\ell} \subset \mathcal{S}(n_\ell)$, $\{\widetilde{S}^{(i)}(n_\ell)\}_{i=1}^{K_3 2^\ell} \subset \widetilde{\mathcal{S}}(n_\ell)$ satisfying*

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

Proof. Pick an arbitrary source-destination pair in $T(n_\ell)$, 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 42 squarelets at distance less than $\sqrt{2a_{\ell+1}}$ from either of those two squarelets. Assume $n \geq n_0(\ell)$, then $42 \leq K_2 2^{-\ell-1}\gamma(n)$. 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 the squarelets containing the source and the destination node.

In order to construct a decomposition of $T(n_\ell)$, 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 can not 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 $\widetilde{S}^{(1)}(n_\ell)$. 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. We claim that by repeating this process of generating matrices $S^{(i)}(n_\ell)$ and $\widetilde{S}^{(i)}(n_\ell)$, we can match all source-destination pairs to some dense squarelet with at most $K_3 2^\ell = K_2/4$ such matrices. Indeed, a new pair of matrices is generated only when a source-destination pair can not 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 $K_3 2^\ell = 2^{\ell+2}/K_2$ matrices. □

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

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

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

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

where $S^{(1,k)}(n_{\ell+1})\widetilde{S}^{(1,k)}(n_{\ell+1})$ is the traffic relayed over the dense squarelet $A_k(a_{\ell+1})$. In the worst case, there are exactly $n_{\ell+1}$ communication pairs to be relayed and the relay squarelet $A_k(a_{\ell+1})$ contains exactly $n_{\ell+1}$ nodes. We shall assume this worst case in the following.

We focus on the transmission according to the traffic matrix $S^{(1,1)}(n_{\ell+1})\widetilde{S}^{(1,1)}(n_{\ell+1})$. 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 $S^{(1,1)}(n_{\ell+1})\widetilde{S}^{(1,1)}(n_{\ell+1})$, respectively.

B. Phase 2: Multiple Access

Each source node in $U(n_{\ell+1})$ splits its data into $n_{\ell+1}$ equal length parts. Part j is to be relayed over the j-th node in $A_1(n_{\ell+1})$. Each part is separately encoded at the source and separately decoded at the destination. Consider part j. 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 j-th node in the relay cluster. This node then decodes the message parts of all source nodes. Note that this induces a uniform traffic matrix between the nodes in the relay squarelet, i.e., every node needs to transmit quantized observations to every other node. While this traffic matrix is not valid (since it is not a permutation matrix) it can be written as a sum of $n_{\ell+1}$ valid traffic matrices. A fraction $1/n_{\ell+1}$ of the traffic within the relay squarelet is transmitted according to each of these valid traffic matrices. This setup is depicted in Figure 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 $S^{(1,1)}(n_{\ell+1})$ 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 Figure 4. Here, f is the channel encoder, φ the channel decoder, $\{q_k\}_{k=1}^m$ quantizers. All these have to be chosen. The empty boxes, on the other hand, represent fixed memoryless channels. We call R the rate of (f, φ) and $\{R_k\}_{k=1}^m$ the rates of $\{q_k\}_{k=1}^m$.

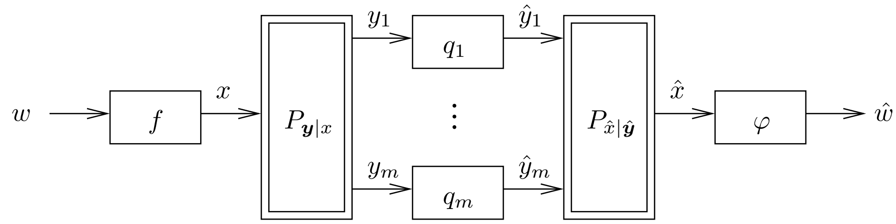

Fig. 4. Sketch of quantized channel model. f and φ are the channel encoder and decoder, respectively; $\{q_k\}_{k=1}^m$ are quantizers. Empty boxes represent memoryless channels.

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 $\left(R, \{R_k\}_{k=1}^m\right)$ is achievable.*

Proof. The proof follows from a simple extension of Theorem 1 in Appendix II of [11]. □

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 $\widetilde{P}_\ell(n) \leq n_{\ell+1}^{-1} a_\ell^{\alpha/2}$, a per source node rate of*

$$\rho_\ell^{MAC}(n) \geq K_4 \widetilde{P}_\ell(n) n_{\ell+1} a_\ell^{-\alpha/2}$$

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

Proof. The source nodes send signals with a power of $n_{\ell+1}^{-1} a_\ell^{\alpha/2}$ for a fraction $\widetilde{P}_\ell(n) n_{\ell+1} a_\ell^{-\alpha/2} \leq 1$ of time and are silent for the remaining time. The time slots during which the nodes send signals are chosen as follows. Generate independently for each relay squarelet k a Bernoulli process $\{T_k[t]\}_{t\in\mathbb{N}}$ with parameter $\widetilde{P}_\ell(n) n_{\ell+1} a_\ell^{-\alpha/2}/(1+\delta) \leq 1$ for some small $\delta > 0$. The nodes in relay squarelet k send signals whenever $T_k[t] = 1$ and remain silent otherwise. Since the blocklength of the codes used is assumed to be large, this satisfies the average power constraint of $\widetilde{P}_\ell(n)$ with high probability for any $\delta > 0$. Since we are interested only in the scaling of capacity, we ignore the additional $1/(1+\delta)$ term in the following to simplify notation. Clearly, we only need to consider the fraction of time during which $T_k[t] = 1$. Let $\boldsymbol{y}$ be the received vector at the relay squarelet, $\hat{\boldsymbol{y}}$ the (componentwise) quantized observations. We use a matched filter at the relay squarelet, i.e.,

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

where column vector $\boldsymbol{h}_u = \{h_{u,v}\}_{v\in V(n_{\ell+1})}$ are the channel gains from node $u \in U(n_{\ell+1})$.

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 $K_4 \widetilde{P}_\ell(n) n_{\ell+1} a_\ell^{-\alpha/2}$. The first channel in Lemma 8 (see Figure 4) 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

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

Let $x_u \sim \mathcal{N}_{\mathbb{C}}(0, \tilde{r}_u^\alpha n_{\ell+1}^{-1} a_\ell^{\alpha/2})$ and $\hat{y}_v = y_v + \tilde{z}_v$ for $\tilde{z}_v \sim \mathcal{N}_{\mathbb{C}}(0, \Delta^2)$ independent of $\boldsymbol{y}$ and for some $\Delta^2 > 0$.

Note that

$$|h_{\tilde{u},v}|^2 = r_{\tilde{u},v}^{-\alpha},$$

and by construction of $\tilde{r}_{\tilde{u}}$ and $S^{(1,1)}(n_{\ell+1})$, we have for $v \in V(n_{\ell+1})$

$$\tilde{r}_{\tilde{u}}\sqrt{2a_\ell} \leq r_{\tilde{u},v} \leq 2\tilde{r}_{\tilde{u}}\sqrt{2a_\ell}. \tag{13}$$

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

$$\begin{aligned}\mathbb{E}\left|\hat{y}_v\right|^2 &= \sum_{u\in U(n_{\ell+1})} \left|h_{u,v}\right|^2 \tilde{r}_u^{\alpha} n_{\ell+1}^{-1} a_\ell^{\alpha/2} + N_0 + \Delta^2 \\ &\leq n_{\ell+1} 2^{-\alpha/2} a_\ell^{-\alpha/2} n_{\ell+1}^{-1} a_\ell^{\alpha/2} + N_0 + \Delta^2 \\ &= 2^{-\alpha/2} + N_0 + \Delta^2.\end{aligned}$$

Hence

$$\begin{aligned}I(y_v;\hat{y}_v) &= h(\hat{y}_v) - h(\hat{y}_v|y_v) \\ &\leq \log\big(2\pi e \mathbb{E}\left|\hat{y}_v\right|^2\big) - \log(2\pi e \Delta^2) \\ &\leq \log\big(2\pi e (2^{-\alpha/2} + N_0 + \Delta^2)\big) - \log(2\pi e\Delta^2) \\ &= \log\Big(1 + \frac{2^{-\alpha/2}+N_0}{\Delta^2}\Big). \end{aligned} \tag{14}$$

We now compute $I(x_u;\hat{x}_u)$. We have

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

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

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

$$\mathbb{E}\bigg(\Big|\textstyle\sum_{\tilde{u}\in U(n_{\ell+1})\setminus\{u\}} \dfrac{\boldsymbol{h}_u^*\boldsymbol{h}_{\tilde{u}}}{\left\|\boldsymbol{h}_u\right\|} x_{\tilde{u}} + \dfrac{\boldsymbol{h}_u^*}{\left\|\boldsymbol{h}_u\right\|}(\boldsymbol{z}+\tilde{\boldsymbol{z}})\Big|^2 \Big| \{\boldsymbol{h}_{\tilde{u}}\}\bigg) = n_{\ell+1}^{-1} a_\ell^{\alpha/2} \displaystyle\sum_{\tilde{u}\in U(n_{\ell+1})\setminus\{u\}} \tilde{r}_{\tilde{u}}^{\alpha} \frac{\left|\boldsymbol{h}_u^*\boldsymbol{h}_{\tilde{u}}\right|^2}{\left\|\boldsymbol{h}_u\right\|^2} + N_0 + \Delta^2,$$

where we have used the assumption that $\{z_v\}_{v\in V(n_{\ell+1})}$ are uncorrelated in the second line.

Using

$$\left\|\boldsymbol{h}_{\tilde{u}}\right\|^2 = \sum_{v\in V(n_{\ell+1})} \left|h_{\tilde{u},v}\right|^2 = \sum_{v\in V(n_{\ell+1})} r_{\tilde{u},v}^{-\alpha}$$

and (13), yields

$$2^{-3\alpha/2} n_{\ell+1} a_\ell^{-\alpha/2} \leq \tilde{r}_{\tilde{u}}^{\alpha} \left\|\boldsymbol{h}_{\tilde{u}}\right\|^2 \leq 2^{-\alpha/2} n_{\ell+1} a_\ell^{-\alpha/2}.$$

Hence we can lower bound the received signal power as

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

Since Gaussian noise is the worst additive noise under a power constraint, we obtain

$$\begin{aligned} I(x_u;\hat{x}_u) &\geq \mathbb{E}\log\left(1 + \frac{2^{-3\alpha/2}}{2^{3\alpha/2}\tilde{r}_u^{\alpha} n_{\ell+1}^{-2} a_\ell^{\alpha} \sum_{\tilde{u}\in U(n_{\ell+1})\setminus\{u\}} \tilde{r}_{\tilde{u}}^{\alpha} \left|\boldsymbol{h}_u^*\boldsymbol{h}_{\tilde{u}}\right|^2 + N_0 + \Delta^2}\right) \\ &\geq \mathbb{P}\bigg(\tilde{r}_u^{\alpha} \sum_{\tilde{u}\in U(n_{\ell+1})\setminus\{u\}} \tilde{r}_{\tilde{u}}^{\alpha} \left|\boldsymbol{h}_u^*\boldsymbol{h}_{\tilde{u}}\right|^2 \leq (B-1)(N_0+\Delta^2) 2^{-3\alpha/2} n_{\ell+1}^2 a_\ell^{-\alpha}\bigg) \\ &\quad \times \log\big(1 + 2^{-3\alpha/2}/B(N_0+\Delta^2)\big), \end{aligned} \tag{15}$$

for any $B \geq 1$, and where the expectation is with respect to $\{\boldsymbol{h}_{\tilde{u}}\}_{\tilde{u}\in U(n_{\ell+1})}$. We have for $u \neq \tilde{u}$,

$$\begin{aligned}\mathbb{E}\left|\boldsymbol{h}_u^*\boldsymbol{h}_{\tilde{u}}\right|^2 &= \mathbb{E}(\boldsymbol{h}_u^*\boldsymbol{h}_{\tilde{u}}\boldsymbol{h}_{\tilde{u}}^*\boldsymbol{h}_u) \\ &= \sum_{v\in V(n_{\ell+1})} \left|h_{u,v}\right|^2 \left|h_{\tilde{u},v}\right|^2 \\ &= \sum_{v\in V(n_{\ell+1})} r_{u,v}^{-\alpha} r_{\tilde{u},v}^{-\alpha}, \end{aligned} \tag{16}$$

and hence

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

Therefore by Markov's inequality

$$\mathbb{P}\left(\sum_{\tilde{u}\in U(n_{\ell+1})\setminus\{u\}} |\boldsymbol{h}_u^*\boldsymbol{h}_{\tilde{u}}|^2 > (B-1)(N_0+\Delta^2)2^{-3\alpha/2} n_{\ell+1}^{-2} a_\ell^\alpha\right) \leq \frac{2^{5\alpha/2}}{(B-1)(N_0+\Delta^2)}.$$

Setting $B \triangleq 2^{5\alpha/2+1}/(N_0+\Delta^2)+1$, we can continue (15) as

$$I(x_u;\hat{x}_u) \geq \frac{1}{2}\log\left(1+2^{-3\alpha/2}/B(N_0+\Delta^2)\right) \triangleq K_4. \tag{17}$$

Using (14) and (17) in Lemma 8, and observing that we only communicate during a fraction

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

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

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

and a quantizer of per relay node rate of arbitrarily close to

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

bits per observation. Since by (17) mutual information $I(x_u;\hat{x}_u)$ 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. □

C. Phase 3: Broadcast

At the end of the MAC phase, each node in the relay squarelet received a part of the message sent by each source node. Each node in the relay cluster 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 induces a uniform traffic matrix between the nodes in the relay squarelet, i.e., every node needs to transmit quantized encoded messages to every other node. While this traffic matrix is not valid (since it is not a permutation matrix) it can be written as a sum of $n_{\ell+1}$ valid traffic matrices. A fraction $1/n_{\ell+1}$ of the traffic within the relay squarelet is transmitted according to each of these valid traffic matrices. This setup is depicted in Figure 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 $\widetilde{S}^{(1,1)}(n_{\ell+1})$ 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* $\widetilde{S}^{(1,1)}(n_{\ell+1})$ in the following.

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

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

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

Proof. Consider a node 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. Let

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

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 channel realization in that time slot. By the law of large numbers, 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 $1/(1+\delta)$ fraction of time with probability approaching one for any $\delta > 0$. Since we have no constraint on the encoding delay in our setup, we can choose δ 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 gain induced by $\{\theta_{v,w}\}_{v,w}$.

As in the MAC phase, the nodes in the relay squarelet send signals at a power $n_{\ell+1}^{-1}a_\ell^{\alpha/2}$ a fraction $\widetilde{P}_\ell(n)n_{\ell+1}a_\ell^{-\alpha/2} \leq 1$ of time and are silent for the remaining time. This is done in the same manner as for the MAC phase, by generating independently for each relay squarelet k a Bernoulli process $\{T_k[t]\}_{t\in\mathbb{N}}$ with parameter $\widetilde{P}_\ell(n)n_{\ell+1}a_\ell^{-\alpha/2}/(1+\delta)$ for some small $\delta > 0$. The nodes in relay squarelet k send signals whenever $T_k[t] = 1$ and remain silent otherwise. As before we ignore the additional $1/(1+\delta)$ term. Again we only need to consider the fraction of time during which $T_k[t] = 1$.

Consider the message part for destination node w. We encode this part independently; call $\tilde{x}_w$ the encoded message part. The receiver then performs transmit beamforming to construct the encoded message for all its destination nodes

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

where row vector $\boldsymbol{h}_w = \{h_{v,w}\}_{v\in V(n)}$ 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 Figure 4) 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 $\hat{x}_v|x_v$. We also need to guarantee that $\hat{x}_v$ satisfies the power constraint at each node v in the relay squarelet. Let $\tilde{x}_w \sim \mathcal{N}_\mathbb{C}(0, K_8 n_{\ell+1}^{-1}a_\ell^{\alpha/2})$ (for some K_8 to be chosen later) and $\hat{x}_v = x_v + \tilde{z}_v$ for $\tilde{z}_v \sim \mathcal{N}_\mathbb{C}(0,\Delta^2)$ independent of $\boldsymbol{x}$ and for some $\Delta^2 > 0$. Then

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

Note that by construction of $\widetilde{S}^{(1,1)}(n_{\ell+1})$, we have for any $w \in W(n_{\ell+1})$

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

and therefore

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

Hence $\hat{x}_v$ has mean zero and variance

$$\begin{aligned}\mathbb{E}\left|\hat{x}_v\right|^2 &= \sum_{w\in W(n_{\ell+1})} \frac{\left|h_{v,w}\right|^2}{\left\|\boldsymbol{h}_w\right\|^2} K_8 n_{\ell+1}^{-1} a_\ell^{\alpha/2} + \Delta^2 \\ &\leq n_{\ell+1}\frac{2^\alpha}{n_{\ell+1}} K_8 n_{\ell+1}^{-1} a_\ell^{\alpha/2} + \Delta^2 \\ &\leq n_{\ell+1}^{-1} a_\ell^{\alpha/2}, \end{aligned} \tag{18}$$

for

$$K_8 \triangleq 2^{-\alpha}(1-\Delta^2),$$

which is positive for $\Delta^2 < 1$. Equation (18) shows that $\hat{x}_v$ satisfies the power constraint of node v in the relay squarelet $V(n_{\ell+1})$. Moreover, we obtain

$$\begin{aligned} I(x_v;\hat{x}_v) &= h(\hat{x}_v) - h(\hat{x}_v|x_v) \\ &\leq \log(2\pi e\mathbb{E}\left|\hat{x}_v\right|^2) - \log(2\pi e\Delta^2) \\ &\leq \log\Big(\frac{n_{\ell+1}^{-1} a_\ell^{\alpha/2}}{\Delta^2}\Big). \end{aligned} \tag{19}$$

It remains to compute $I(\tilde{x}_w; y_w)$. Note that

$$|\boldsymbol{h}_w \hat{\boldsymbol{h}}_w^*|^2 \geq \cos(\pi/4)^2 \left\|\boldsymbol{h}_w\right\|^4,$$

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

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

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

$$I(\tilde{x}_w; y_w) \geq K_6. \tag{20}$$

Using (19) and (20) in Lemma 8, and observing that we only communicate during a fraction

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

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

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

bits per channel use and a quantizer rate arbitrarily close to

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

bits per encoded sample. Since by (20) mutual information $I(\tilde{x}_w;\hat{y}_w)$ 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 per relay node required to quantize the encoded message samples is at most

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

bits per $n_{\ell+1}$ total message bits received by the destination nodes. □

D. Proof of Theorem 1: Fast Fading

The next result implies Theorem 1 under the fast fading assumption, i.e., $\{\theta_{u,v}[t]\}_{u,v,t}$ is i.i.d. with uniform distribution on $[0,2\pi)$.

Theorem 11. *Under fast fading, for any $\alpha > 2$, for any sequences of node placements $\{V(n)\}_{n\geq 1}$ with minimum separation $c > 0$ and valid traffic matrices $\{T(n)\}_{n\geq 1}$, we have*

$$\rho^*(n) \geq \rho^{HR}(n) \geq n^{1-\alpha/2-\beta(n)},$$

where $\beta(n) = O(\log^{-1/3} n)$ as $n \to \infty$.

Proof. Consider the hierarchical relaying scheme as described in Section IV and fix a level ℓ, $0 \leq \ell < L = L(n)$ in this hierarchy. Level ℓ is at a scale of $a_\ell \triangleq n/\gamma(n)^\ell$, with $n_\ell \triangleq n/2^\ell\gamma(n)^\ell$ 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 $\widetilde{P}_\ell(n) \triangleq PK_2 2^{-\ell}\gamma(n)$. Since $\alpha > 2$ and since $\gamma(n)^\ell \leq n$ by (12), we have, for n large enough,

$$\widetilde{P}_\ell(n) = PK_2 2^{-\ell}\gamma(n) \leq 2^{\ell+1}\gamma(n)(n/\gamma(n)^\ell)^{\alpha/2-1} = n_{\ell+1}^{-1}a_\ell^{\alpha/2},$$

and hence the power constraint in Lemmas 9 and 10 is satisfied. Now, partition the squarelets at level ℓ into four subsets such that in each subset all squarelets are at distance at least $\sqrt{a_{\ell+1}}$ from each other. The induced traffic from the MAC and BC phases in each of the relay squarelets at level ℓ is transmitted simultaneously within all relay squarelets in the same subset. Consider one such subset and call the relay squarelets within it *active*. We now show that at any active relay squarelet the interference from other active relay squarelets is memoryless within each phase, additive (i.e., independent of the signals and channel gains in this active relay squarelet), and of bounded power $N_0 - 1$ independent of n.

We first argue that the interference is memoryless within each phase. Note first that on any level $\ell+1$ in the hierarchy, all active relay squarelets are either simultaneously in the MAC phase or simultaneously in the BC phase. Furthermore, all active 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 BC induced by some valid traffic matrix is memoryless. 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 i.i.d. fading, this is clearly 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 an active relay squarelet are independent of the channel gains between two active relay squarelets, this interference is additive as well.

We now bound the interference power. In each active relay squarelet, at most $n_{\ell+1}$ nodes transmit at an average power of P (since the interference is memoryless within each phase). In the MAC phase, all nodes use independently generated codebooks with power at most P, and thus the received interference power from an active relay squarelet at distance $i\sqrt{a_{\ell+1}}$ is at most

$$Pn_{\ell+1}i^{-\alpha}a_{\ell+1}^{-\alpha/2} = Pi^{-\alpha}2^{-(\ell+1)}\Big(\frac{n}{\gamma(n)^{\ell+1}}\Big)^{1-\alpha/2} \leq Pi^{-\alpha},$$

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 (16) shows that we can upper bound the received interference power from an active relay squarelet at distance $i\sqrt{a_{\ell+1}}$ by

$$P2^{\alpha}n_{\ell+1}i^{-\alpha}a_{\ell+1}^{-\alpha/2} \leq P2^{\alpha}i^{-\alpha},$$

again by (12). Now since every active relay squarelet has at most $8i$ active relay squarelets at distance at least $i\sqrt{a_{\ell+1}}$, the total interference power received at an active relay squarelet is at most

$$\sum_{i=1}^{\infty} 8iP2^{\alpha}i^{-\alpha} \triangleq N_0 - 1 < \infty$$

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

Call $\tau_\ell(n)$ the number of channel uses to transmit one bit from each of n_ℓ source to the corresponding destination nodes at level ℓ. Lemma 7 states that for $n \geq n_0(\ell)$ we relay over each dense squarelet at most $K_3 2^\ell$ times, where $n_0(\ell)$ is such that $\gamma(n_0(\ell)) \geq 42K_2^{-1}2^{\ell+1}$. Since $2^{-L(n)}\gamma(n) \to \infty$ as $n \to \infty$, and since $\gamma(n)$ is increasing, both by (12), there exists n_0 independent of ℓ such that for $n \geq n_0$

$$\gamma(n) \geq 42K_2^{-1}2^{L(n)} \geq 42K_2^{-1}2^{\ell+1}.$$

We assume $n \geq n_0$ in the following, such that Lemma 7 applies. 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

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

channel uses for the MAC phase. Here, the factor 4 accounts for the spatial re-use, $K_3 2^\ell$ 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

$$\frac{K_3 4^{\ell+1}}{K_6 P}n^{\alpha/2-1}\gamma(n)^{\ell(1-\alpha/2)+1}$$

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

$$\tau_\ell(n) \leq 4K_3\Big(\frac{1}{PK_4}+\frac{1}{PK_6}\Big)n^{\alpha/2-1}\gamma(n)\big(4\gamma(n)^{1-\alpha/2}\big)^\ell + (K_5+K_7(\ell+1)\log(n))\tau_{\ell+1}(n)$$
$$\leq \widetilde{K}n^{\alpha/2-1}\gamma(n)4^\ell + (K(\ell+1)\log(n))\tau_{\ell+1}(n) \tag{21}$$

for positive constants $K, \widetilde{K}$ independent of n and ℓ, and where we have used $\alpha > 2$.

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 Pn_L^{-1}. This achieves

$$\tau_L(n) \leq \frac{K_8}{P}a_L^{\alpha/2} = \frac{K_8}{P}n^{\alpha/2}\gamma(n)^{-L\alpha/2},$$

for some K_8 and assuming

$$Pn_L^{-1} = P2^L\gamma(n)^L/n \leq n^{\alpha/2}\gamma(n)^{-L\alpha/2}, \tag{22}$$

where (22) ensures that during the time sharing we still operate in the power limited regime. Thus using (21) L times, we obtain

$$\tau_0(n) \leq \widetilde{K}n^{\alpha/2-1}\gamma(n) + (K4\log(n))\tau_1(n)$$
$$\leq \dots$$
$$\leq n^{\alpha/2-1}\big(4KL\log(n)\big)^L\Big(\widetilde{K}L\gamma(n) + \frac{K_8}{P}n\gamma(n)^{-L\alpha/2}\Big). \tag{23}$$

From the definition of $L(n)$ and $\gamma(n)$ in (11),

$$n\gamma(n)^{-L(n)\alpha/2} = 1.$$

We now argue that this choice of $\gamma(n)$ and $L(n)$ satisfies the power requirement (22). We have

$$n^{1+\alpha/2}\gamma(n)^{-L(n)(1+\alpha/2)} = n^{\alpha/2-2/\alpha} \geq P2^{\log^{1/3}(n)} = P2^{L(n)}$$

for n large enough, satisfying (22). We can thus continue (23) as

$$\tau_0(n) \leq n^{\alpha/2-1}\big(4K\log^{4/3}(n)\big)^{\log^{1/3}(n)}\Big(\widetilde{K}\log^{1/3}(n)n^{2/\alpha\log^{1/3}(n)} + \frac{K_8}{P}\Big)$$
$$\leq n^{\alpha/2-1+\beta(n)},$$

where $\beta(n) = O(\log^{-1/3} n)$, as $n \to \infty$. And hence

$$\rho^*(n) \geq \rho^{\text{HR}}(n) = 1/\tau_0(n) \geq n^{1-\alpha/2-\beta(n)}.$$

□

E. Proof of Theorem 1: Slow Fading

In this section, we outline an extension of the hierarchical relaying scheme described in Sections V-A-V-D. Specifically, we show that the above stated scheme works also for slow fading,

We start by recalling a large deviation result, similar to the Chernoff bound, that will be used several times in the following.

Lemma 12. *Let $\{B_i\}$ be a Bernoulli process with parameter p. Then for every $\delta > 0$ there exists $K > 0$ such that for all $n \geq 1$*

$$\mathbb{P}\Big(\sum_{i=1}^n B_i \leq (1-\delta)pn\Big) \leq \exp(-Kpn).$$

Proof. See, for example, [11, Lemma 4.1]. □

The next theorem implies Theorem 1 under the slow fading assumption, i.e., for $\{\theta_{u,v}[t]\}_{u,v}$ i.i.d. with uniform distribution on $[0, 2\pi)$ and constant as a function of t.

Theorem 13. *Under slow fading, for any $\alpha > 2$, for any sequences of node placements $\{V(n)\}_{n\geq 1}$ with minimum separation $c > 0$ and valid traffic matrices $\{T(n)\}_{n\geq 1}$, we have*

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

with probability $1 - o(1)$ as $n \to \infty$, and where $\beta(n) = O(\log^{-1/3} n)$.

Proof. We sketch the necessary modifications of the scheme described in Section IV that provides a per node rate of at least $n^{1-\alpha/2-\beta(n)}$ as $n \to \infty$ in the slow fading case.

Consider level ℓ, $0 \leq \ell < L = L(n)$ 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 node. 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.

Lemma 14. *Let $n_0 = n_0(\ell)$ such that $\gamma(n_0) \geq 42K_2^{-1}2^{\ell+1}$. Then for every $n_\ell \geq n_0(\ell)$ and every valid traffic matrix $T(n_\ell) \in \{0,1\}^{n_\ell \times n_\ell}$ there are $\{S^{(i)}(n_\ell)\}_{i=1}^{K_2 2^{-\ell}\gamma(n)^2} \subset \mathcal{S}(n_\ell)$, $\{\widetilde{S}^{(i)}(n_\ell)\}_{i=1}^{K_2 2^{-\ell}\gamma(n)^2} \subset \widetilde{\mathcal{S}}(n_\ell)$ satisfying*

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

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

$$\begin{aligned} \sum_{jk} S^{(i)}(n_\ell)_{jk} S^{(i')}(n_\ell)_{jk} = 0, \\ \sum_{kj} \widetilde{S}^{(i)}(n_\ell)_{kj} \widetilde{S}^{(i')}(n_\ell)_{kj} = 0. \end{aligned} \tag{24}$$

Proof. The proof is similar to that of Lemma 7. In order to construct the sequence of $S^{(i)}(n_\ell)$ and $\widetilde{S}^{(i)}(n_\ell)$ over i, consider the sequential pass over all $n_\ell \geq n_0$ source-destination pairs. 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 this procedure as soon as any of the relay squarelets is chosen by $n_{\ell+1}$ pairs. Since 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 $\widetilde{S}^{(i)}(n_\ell)$, with the i-th pair of matrices describing communication over the i-th relay squarelets chosen by source-destination pairs matched in this round. Thus, this process produces a total of $2\gamma(n)K_2 2^{-\ell-1}\gamma(n) = K_2 2^{-\ell}\gamma(n)^2$ such matrices. The orthogonality property follows since each source-destination pair relays over the same relay squarelet only once. □

Given a decomposition of the scaled traffic matrix $K_2 2^{-\ell-1}\gamma(n)T(n)$ into $K_2 2^{-\ell}\gamma(n)^2$ matrices, each source-destination pair tries to relay over $K_2\gamma(n)/2$ dense squarelets. We time share between these relay squarelets. Comparing this to Lemma 7, we see that the loss due to this time sharing is now

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

as opposed to $K_4 2^\ell$ there. In other words, the loss is at most a factor $2\gamma(n)$ more than in Lemma 7. Using the definition of $L(n)$ and $\gamma(n)$ in (11), we have $\gamma(n) \leq n^{\beta(n)}$. 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 a fraction $1/64$ 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 half of its relay squarelets can support this rate. As we only communicate over half the relay squarelets, this implies that we can achieve at least $1/128$ of the per node rate given by Theorem 11, i.e., that $n^{1-\alpha/2-\beta(n)}$ is achievable with probability $1-o(1)$ as $n \to \infty$.

Assume now we have for each source-destination pair (u,w) 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 $C_{u,w,k}$ 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 8 times the one for fast fading, respectively. From the proof of Lemmas 9, 10, and of Theorem 1 in SectionV-D, we see that

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

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

We now argue that the events

$$\Big\{ \cap_{i=1}^4 B_{u,w,k}^{(i)} \Big\}_{k=1}^{K_2 2^{-\ell-1}\gamma(n)} \tag{25}$$

are independent conditioned on $\widetilde{H}$. We show that these events depend on disjoint sets of channel gains and codebooks. Assuming the codebooks are generated new for each communication round, these are all independent, it follows that the events (25) are as well. 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 $\widetilde{U}_k$, $\widetilde{W}_k$ be the source and destination nodes that are communicating at the same time as (u,w) due to spatial reuse. Let $\widetilde{V}_k$ be the relay nodes of $\widetilde{U}_k$ and $\widetilde{W}_k$. Now, $B_{u,w,k}^{(1)}$ and $B_{u,w,k}^{(2)}$ depend (for fixed $\widetilde{H}$) 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 $\widetilde{H}$) on the channel gains between $\widetilde{V}_k$ and $\widetilde{W}_k$. Since these sets are disjoint for different k by (24), conditional independence of the events in (25) follows.

With this, Lemma 12 yields that

$$\mathbb{P}\Big(\textstyle\sum_{k=1}^{K_2 2^{-\ell-1}\gamma(n)} \mathbb{1}_{C_{u,w,k}} < K_2 2^{-\ell-2}\gamma(n) \big| \widetilde{H}\Big) \leq \exp\big(-K 2^{-\ell}\gamma(n)\big)$$

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

$$\mathbb{P}\Big(\textstyle\sum_{k=1}^{K_2 2^{-\ell-1}\gamma(n)} \mathbb{1}_{C_{u,w,k}} < K_2 2^{-\ell-2}\gamma(n)\Big) \leq \exp\big(-K 2^{-\ell}\gamma(n)\big).$$

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

$$1 - L(n)n^2 \exp\big(-K2^{-\ell}\gamma(n)\big),$$

for every source-destination pair on every level of the hierarchy at least half of its relay squarelets can support the desired rate. By the choice of $\gamma(n)$ and $L(n)$ in (11), this probability is at least $1 - o(1)$ as $n \to \infty$, proving the theorem. □

VI. Proof of Theorem 2

Here, we briefly sketch how the arguments of [11, Theorem 5.2] can be adapted to prove Theorem 2. Theorem 5.2 in [11] is the same statement as Theorem 2 here, but for random node placement and fast fading. Thus we only have to show that the arguments in [11] hold also under arbitrary node placement with minimum separation and for slow fading.

Theorem 5.2 in [11] is proven by considering a cut dividing the network into two parts, each of size $\Theta(n)$ and containing $\Theta(n)$ nodes. By the minimum-separation requirement, an area of size $o(n)$ can contain at most $o(n)$ nodes, and hence such a cut exists also in our case. The proof of Theorem 5.2 in [11] proceeds then by upper bounding the sum rate achievable over this cut, showing that it can not be too big. The crucial observation is that this proof depends only on two properties satisfied with high probability by nodes placed uniformly at random on $[0, \sqrt{n}]^2$ (see [11, Lemma 5.1]):

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 $[ib, (i+1)b] \times [jb, (j+1)b]$ for any $i, j \in \{0, \ldots, \sqrt{n}/b - 1\}$, where $b \triangleq \sqrt{2 \log n}$.

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. For the second requirement, note that we can simply add a node inside each $[ib, (i+1)b] \times [jb, (j+1)b]$ that is empty. This increases the number of nodes by at most a factor of two, and can be done such that the new node placement still satisfies the same minimum-separation requirement. Adding these nodes can only increase the sum rate across the cut, and hence further upper bounds the achievable per node rate. Moreover, since the sum rate is at most polynomial in n, the factor of two increases the sum rate by at most a constant factor, thus yielding the same scaling.

To see that [11, Theorem 5.2] remains valid for slow fading, note that Markov's inequality implies that the probability that the sum rate across the cut is K times larger than the one achievable for fast fading is at most $1/K$.

VII. Proof of Theorem 3

Consider $V(n)$ with $n/2$ nodes located on $[0, \sqrt{n}/4] \times [0, \sqrt{n}]$ and $n/2$ nodes located on $[3\sqrt{n}/4, \sqrt{n}] \times [0, \sqrt{n}]$ (assume $c > 0$ is small enough such that this is possible). A random $T(n)$ 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 $T(n)$ in the following.

In this setup, with multi-hop 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 $\rho^{\mathrm{MH}}(n) = O(n^{-\alpha/2})$ as $n \to \infty$.

Moreover, considering a cut between the two clusters and slightly adapting the argument of [11, Lemma 5.2] yields that $\rho^*(n) = O\big(n^{1-\alpha/2+\varepsilon}\big)$ as $n \to \infty$.

VIII. Minimum-Distance Requirement

As already mentioned, the minimum-separation requirement $c > 0$ 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. The next results shows how this observation can be used to find a lower bound on $\rho^*(n)$ for nodes placed uniformly at random on $[0,\sqrt{n}]^2$.

Proposition 15. *For each $n \geq 1$, let $V(n)$ be n nodes placed independently uniformly at random on $[0,\sqrt{n}]^2$. Under either fast or slow fading, for any $\alpha > 2$, for any sequence of valid traffic matrices $\{T(n)\}_{n\geq 1}$, we have*

$$\rho^*(n) \geq n^{1-\alpha/2-\beta(n)}$$

with probability $1-o(1)$ as $n \to \infty$, and where $\beta(n) = O(\log^{-1/3} n)$.

Proof. As pointed out before Lemma 6, we only need to show that in the case of random node placements there exist enough dense squarelets on each level in the hierarchy. We show that, in fact, all squarelets at each level in the hierarchy are dense with probability $1-o(1)$ as $n \to \infty$.

Let B_i be the event that node i falls into a particular squarelet at level $L(n)$. $\{B_i\}_{i=1}^{\gamma(n)^{L(n)}}$ are i.i.d. and $\mathbb{P}(B_i) = 1/\gamma(n)^{L(n)}$. Hence by Lemma 12, all squarelets at level $L(n)$ contain at least $n/2\gamma(n)^{L(n)}$ nodes with probability larger than

$$1 - \gamma(n)^{L(n)} \exp\big(-Kn/\gamma(n)^{L(n)}\big),$$

for some constant K. By the definition in (11),

$$\gamma(n) \triangleq n^{2/L(n)\alpha}$$

and hence this probability is larger than

$$1 - n^{2/\alpha} \exp\big(-Kn^{1-2/\alpha}\big).$$

This converges to one as $n \to \infty$ since $\alpha > 2$.

Since every squarelet at level $L(n)$ contains at least $n/2\gamma(n)^{L(n)}$ nodes with probability $1-o(1)$ as $n \to \infty$, this implies that all squarelets at every level $\ell \in \{1,\dots,L(n)\}$ are dense with probability $1-o(1)$ as $n \to \infty$. The result follows now from Theorem 1. ☐

Proposition 15 yields a different proof of Theorem [11, Theorem 5.1]. The result is, however, stronger since it not only applies to the fast fading case treated in [11], but also to the slow fading case, proving the conjecture in [11] (at least for extended networks).

IX. Cooperative Multi-Hop Scheme

Theorem 1 shows that for $\alpha > 2$, under either slow or fast fading a per node rate of $\rho^{\text{HR}}(n) \geq n^{1-\alpha/2-\beta(n)}$ is achievable as $n \to \infty$. Theorem 2 shows that for $2 < \alpha \leq 3$ we have $\rho^*(n) \leq O(n^{1-\alpha/2+\varepsilon})$, and for $\alpha > 3$ we have $\rho^*(n) \leq O(n^{-1/2+\varepsilon})$. However, as Theorem 3 shows, this upper bound is not necessarily tight for arbitrary node placement with $\alpha > 3$. This is in contrast to the situation for regular or random node placement, where the upper bound is indeed tight, as multi-hop communication achieves a per node rate of $\rho^{\text{MH}}(n) = \Theta(\sqrt{n})$ for $\alpha > 2$. In other words, for optimal multi-hop communication, a high degree of local regularity is required, whereas for the hierarchical relaying scheme only global regularity (namely, existence of dense squarelets) is needed.

This observation raises the following question. Assume $\alpha > 3$ and nodes located with not enough regularity to allow efficient multi-hop communication. Can we achieve a better per node rate than the $n^{1-\alpha/2-\beta(n)}$ guaranteed by hierarchical relaying? The goal of this section is to show that this is indeed possible, with an improvement depending on the regularity of the node placement. The scheme achieving this “interpolates” between multi-hop communication (which requires only communication between nodes at distance $\Theta(1)$, but offers only a multiplexing gain of one) and hierarchical relaying (which requires communication between nodes at distance $\Theta(\sqrt{n})$, but offers a multiplexing gain of essentially order $\Theta(n)$). We call this the *cooperative multi-hop scheme* in the following.

Recall that $V(n)$ is regular at resolution $h(n)$ if every square $[ih(n),(i+1)h(n)]\times[jh(n),(j+1)h(n)]$ contains at least $h(n)^2/K_9$ nodes for some constant $K_9 \geq 1$. The following result, building on the hierarchical relaying scheme of Theorem 1, gives a lower bound on the per node rate achievable with cooperative multi-hop communication. While this result is valid for all $\alpha > 2$, it is of interest only for $\alpha > 3$, since Theorem 1 and Theorem 2 show that hierarchical relaying is optimal for $2 < \alpha \leq 3$. Theorem 16 implies the achievability part of Theorem 4.

Theorem 16. *Under slow or fast fading, for any $\alpha > 2$, for any sequences of node placements $\{V(n)\}_{n\geq 1}$ with minimum separation $c > 0$ and valid traffic matrices $\{T(n)\}_{n\geq 1}$, we have*

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

as $n \to \infty$, where

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

and with $\beta(n) = O(\log^{-1/3} n)$.

Proof. We outline how a communication scheme can be constructed that achieves the claimed per node rate. We use the hierarchical relay scheme as building block.

Assume $V(n)$ is regular at resolution $h(n)$ for all $n \geq 1$. We show that this implies that we can achieve a per node rate of at least $h(n)^{3-\alpha} n^{-1/2-\beta(n)}$ as $n \to \infty$. Taking the smallest such $h(n)$ then yields the result. If $h(n) = \Theta(\sqrt{n})$ as $n \to \infty$ then the result follows directly from Theorem 1. We therefore assume without loss of generality that $h(n) = o(\sqrt{n})$ as $n \to \infty$.

Divide $A(n)$ into squares of sidelength $h(n)$ as described above. We now show that we can use multi-hop communication where each of these squares cooperatively sends information to one of its neighboring squares. Since $V(n)$ is regular at resolution $h(n)$, each such square contains at least $h(n)^2/K_9$ nodes. Pick the top left most square and construct the square of sidelength $2h(n)$ consisting of it together with its 3 neighbors. Continue in the same fashion, partitioning all of $A(n)$ into squares of sidelength $2h(n)$. Note that each such bigger square contains at least $4h(n)^2/K_9$ nodes by the definition of $h(n)$. We assume this worst case in the following. Partition $A(n)$ into 4 subsets of those bigger squares such that within each such subset each square is at distance at least $2h(n)$ from any other square (see Figure 5). We time share between those 4 subsets. Consider in the following one such subset. For every bigger square, we construct two traffic matrices $T_1(4h(n)^2/K_9)$ and $T_2(4h(n)^2/K_9)$. In T_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 Figure 5). Similarly, T_2 contains communication pairs between left and right squares. We time share between T_1 and T_2.

Communication according to T_i within bigger squares in the same subset occur simultaneously. We are going to use hierarchical relaying within each bigger square. This is possible since each such square contains at least $4h(n)^2/K_9$ 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 proofs of Theorem 11 and 13 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 $4h(n)^2$, $4h(n)^2/K_9$ 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 either slow or fast fading we can achieve a per node rate of at least

$$(h(n)^2)^{1-\alpha/2-\beta(n)} = h(n)^{2-\alpha-2\beta(n)},$$

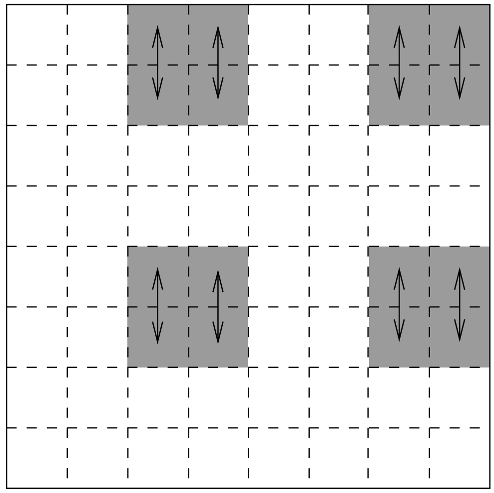

Fig. 5. Sketch of the construction in the proof of Theorem 16. The dashed squares have sidelength $h(n)$. The gray area is one of the 4 subsets of bigger squares that communicate simultaneously. The arrows indicate the traffic matrix T_1.

as $n \to \infty$. The setup is the same for all bigger squares within each of the 4 subsets.

We now "shift" the way we defined the bigger squares by $h(n)$ to the right and to the bottom. With this, each new bigger square intersects with 4 old bigger squares. 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 $h(n)$ 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 Figure 6.

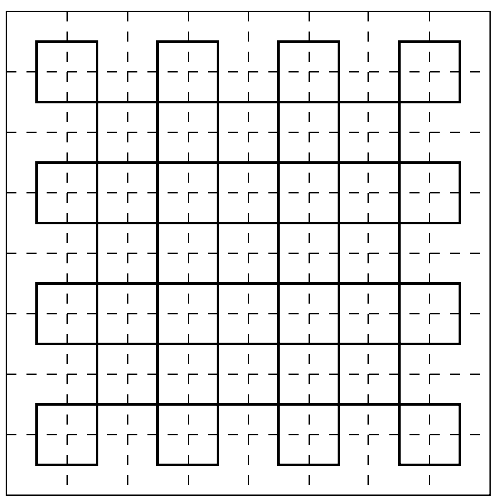

Fig. 6. Communication graph (in bold) in the proof of Theorem 16. The dashed squares have sidelength $h(n)$.

With the above construction, we can communicate along each edge of this graph simultaneously at a per node rate of $h(n)^{2-\alpha-2\beta(n)}/16$ as $n \to \infty$. This is now a regular graph with diameter at most $2\sqrt{n}/h(n)$. Using that each bigger square can contain at most $K_1 h(n)^2$ nodes by the minimum-separation requirement, standard arguments show that we can achieve a per node rate of

$$h(n)^{2-\alpha-\beta(n)} h(n)/\sqrt{n} \geq h(n)^{3-\alpha} n^{-1/2-\beta(n)}$$

as $n \to \infty$, where we have redefined $\beta(n)$ appropriately. □

Theorem 16 shows that if $\{V(n)\}_{n\geq 1}$ are regular at resolution $h^*(n)$ then a per node rate of at least $\rho^{\mathrm{CMH}}(n) \geq h^*(n)^{3-\alpha} n^{-1/2-\beta(n)}$ is achievable under either fast or slow fading. The next theorem shows that Theorem 16 is tight under adversarial node placement with a constraint on the regularity. Moreover, it shows that, at least for $h^*(n) > n^{1/6}$, hierarchical cooperation (either hierarchical relaying for $2 < \alpha \leq 3$ or hierarchical multi-hop for $\alpha > 3$) can be strictly better than classical multi-hop communication. Theorem 17 implies the converse part of Theorem 4.

Theorem 17. *Under either fast or slow fading, there exist sequences of node placements $\{V(n)\}_{n\geq 1}$ with minimum separation $c > 0$ and regular at resolution $h^*(n)$, $1 \leq h^*(n) \leq \sqrt{n}$, such that for $\{T(n)\}_{n\geq 1}$ uniformly distributed over the set of all valid traffic matrices, we have for any $\alpha > 3$ and $\varepsilon > 0$*

$$\rho^*(n) = O\big(h^*(n)^{3-\alpha} n^{-1/2+\varepsilon}\big),$$
$$\rho^{MH}(n) = O\big(h^*(n)^{-\alpha}\big),$$

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

Proof. Consider $[0, \sqrt{n}]^2$ and divide it into squares of sidelength $h^*(n)$, yielding a total of $n/h^*(n)^2$ such squares. Choose a quarter of these squares such that each chosen square is at least at distance $h^*(n)$ from any other chosen square (see Figure 7). Put $4h^*(n)^2$ nodes in each of these chosen squares. We assume that $c > 0$ is small enough such that this is possible. Clearly, $V(n)$ is regular at resolution $2h^*(n)$.

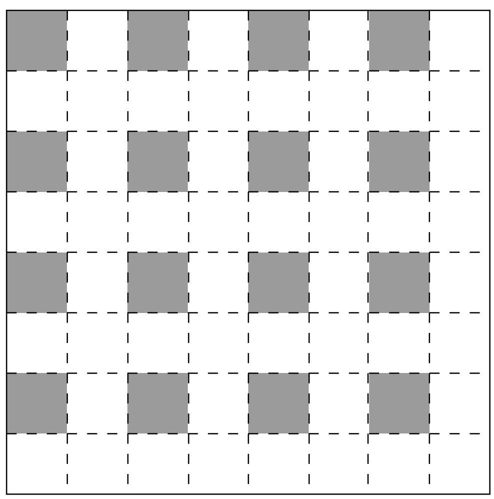

Fig. 7. Construction of node placements in the proof of Theorem 17. Each square has a sidelength of $h^*(n)$. There are $n/4h^*(n)^2$ shaded squares, each of them containing $4h^*(n)^2$ nodes.

We now show that for $\{T(n)\}_{n\geq 1}$ chosen uniformly at random over the set of valid traffic matrices, we also have $\rho^*(n) = O\big(h^*(n)^{3-\alpha} n^{-1/2+\varepsilon}\big)$ for any $\varepsilon > 0$. Divide $V(n)$ vertically into two equal size pieces. With probability $1 - o(1)$, there are $\Theta(n)$ communication pairs crossing from left to right over this cut. Slightly adapting the argument of [11, Lemma 5.2] yields that for any $\varepsilon > 0$, $\rho^*(n) = O\big(h^*(n)^{3-\alpha} n^{-1/2+\varepsilon}\big)$ for $\alpha > 3$.

On the other hand, classical multi-hop is *not* order optimal in this example for any α and for $h^*(n)$ growing fast enough. This is due to the large "gaps" in the node placement $V(n)$, which require cooperative communication to be crossed. Indeed, since with probability $1 - o(1)$ there are $\Theta(n)$ communication pairs having the source and destination node in different squares of sidelength $h^*(n)$, and since for each such pair at least one hop has to cross the "gap" between two squares, we have, ignoring interference from other nodes transmitting simultaneously, $\rho^{\mathrm{MH}}(n) = O(h^*(n)^{-\alpha})$. □

X. Proof of Theorem 5

In this section, we show how results on communication according to permutation traffic matrices can be used to find inner bounds on the n^2 dimensional capacity region Λ of a wireless network. In the remainder of this section, we consider a wireless network with n nodes placed arbitrarily on $\mathbb{R}^2$. As we shall see, the arguments in this section are independent of the specific choice of channel model.

Recall the definition of the capacity region $\Lambda(n)$ as the closure of the set of achievable traffic matrices $\lambda \in \mathbb{R}^{n\times n}$, and $\rho^*_\lambda(n)$ as the largest multiple of λ that is achievable. For example, $\rho^*(n)$ for a valid traffic matrix $T(n)$ is precisely $\rho^*_T(n)$.

In order to prove Theorem 5, we use a result of London, Linial and Rabinovich [12] about approximate multi-commodity flows. In order to state their result, we need some notation. Given an undirected capacitated graph $G = (V, E)$ of n nodes with edge capacities $C(u, v) \geq 0$ for $(u, v) \in E$. Consider m source-destination pairs $\{(u_i, w_i)\}_{i=1}^m$ that require sending data at rate proportional to $d_i > 0$ from u_i to w_i. Define

$$D(S) \triangleq \sum_{i:u_i\in S, w_i\in S^c} d_i + \sum_{i:u_i\in S^c, w_i\in S} d_i,$$

and let

$$C(S) \triangleq \sum_{(u,v)\in E: u\in S, v\in S^c} C(u, v).$$

Let $\phi \geq 0$ be the largest constant so that data can be routed in the network from u_i to w_i at rate ϕd_i for all $i \in \{1, \dots m\}$. Then [12, Theorem 4.1] states that

$$\Omega\biggl(\frac{1}{\log m}\Bigl(\min_{S\subset V} \frac{C(S)}{D(S)}\Bigr)\biggr) = \phi = O\biggl(\min_{S\subset V} \frac{C(S)}{D(S)}\biggr). \tag{26}$$

It is worth noting here that the upper bound in (26) follows immediately. The difficulty lies in establishing the lower bound within $O(\log m)$ factor.

We are now ready to prove Theorem 5. Any $\lambda \in \mathbb{R})_+^{n\times n}$ defines n^2 source-destination pairs that want to transmit data at rate proportional to $\lambda_{u,v}$. In order to apply result (26), we need to define an appropriate undirected capacitated graph G. Let G be the complete graph on n nodes with

$$C(u, v) = \frac{\min_i \rho^*_{T_i}(n)}{n}.$$

Since $C(u, v) = C(v, u)$ this is a valid set of edge capacities for G. Time sharing between $\{T_i(n)\}_{i=1}^n$, we see that data can be transmitted over the wireless channel at rate $C(u, v)$ simultaneously between each $u, v \in V(n)$.

Applying the lower bound in (26) yields thus

$$\rho^*_\lambda(n) = \Omega\biggl(\frac{1}{\log(n^2)}\Bigl(\min_{S\subset V(n)} \frac{C(S)}{D_\lambda(S)}\Bigr)\biggr). \tag{27}$$

Now note that $C(S) = C(S^c)$ and hence we can restrict the minimization over $S \subset V(n) : |S| \leq n/2$. For such an S, we have

$$C(S) = \frac{\min_i \rho^*_{T_i}(n)}{n} |S| |S^c| \geq \frac{\min_i \rho^*_{T_i}(n)}{2} |S|.$$

Combining this with (27)

$$\rho^*_\lambda(n) = \Omega\biggl(\frac{\min_i \rho^*_{T_i}(n)}{\log n}\Bigl(\min_{S\subset V(n):|S|\leq n/2} \frac{|S|}{D_\lambda(S)}\Bigr)\biggr),$$

concluding the proof.

XI. Discussion

A. 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 CDMA with several orthogonal signatures or by using OFDM with many sub-carriers. 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.

B. Comparison with [11]

Both, the hierarchical relaying scheme presented here and the hierarchical scheme presented in [11], share that they use virtual multiple-antenna communication [13], [14] 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 $\rho^*(n) \geq n^{1-\alpha/2-\beta(n)}$ with $\beta(n) = O(\log^{-1/3} n)$ compared to $\rho^*(n) = \Omega(n^{1-\alpha/2-\varepsilon})$ for any $\varepsilon > 0$. This is because the hierarchy here is not of fixed depth L as in [11], but rather of depth $L(n) \triangleq \log^{1/3}(n)$, changing with n.

Second, note that the multi-user decoding at the relay squarelets during the MAC phase and the multi-user encoding during the BC phase are very simple in our setup. In fact, using matched filter receivers and transmit beamforming, we convert the problems into single-user decoding and encoding problems. This differs from the approach in [11], 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 [11].

Third, and probably most important, the schemes differ in how they achieve the multiplexing gain from using multiple antennas. In [11], 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 multiplexing gain comes from setting up a virtual MIMO channel between the source and the destination. In our setup, the arbitrary location of nodes makes it impossible to use such an approach. Instead, we used 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 multiplexing 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 [11] suffices to achieve a multiplexing 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 (unlike the scheme in in [11], which was only shown to work for the fast fading channel model).

XII. Conclusions

We considered the problem of the scaling of achievable per node throughput in arbitrary extended wireless networks. We generalized the hierarchical cooperative communication scheme presented in [11] for a fast fading channel model and with random node placements. We provided 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.

We showed that our scheme is always order optimal for small path-loss exponent ($2 < \alpha \leq 3$). Moreover, and unlike the case for random node placements, for certain node placements this scheme can be optimal for all $\alpha > 2$ – in particular, it can be strictly better than multi-hop communication for every $\alpha > 2$. This contrasts with the situation for more "regular" networks (like the ones obtained with high probability through random node placement), in which multi-hop 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.

For extended networks with random node placement, the value $\alpha = 3$ is a threshold below which global communication (i.e., hierarchical cooperative communication) is necessary and above which local communication (i.e., multi-hop communication) is sufficient. We argued that this threshold effect does not necessarily occur for arbitrary node placements. We also presented a family of extensions to our hierarchical cooperative scheme that smoothly "interpolates" between multi-hop and hierarchical cooperative communication, and showed that this scheme can again be order optimal for all $\alpha > 2$.

Finally, we argued how the results derived for permutation traffic can be used to obtain an inner bound on the entire n^2 dimensional capacity region $\Lambda(n)$ of the wireless network.

XIII. Acknowledgments

The authors would like to thank Olivier Lévêque and Greg Wornell for helpful discussions.

References

[1] P. Gupta and P. R. Kumar. The capacity of wireless networks. *IEEE Transactions on Information Theory*, 46(2):388–404, March 2000.
[2] L. Xie and P. R. Kumar. A network information theory for wireless communication: Scaling laws and optimal operation. *IEEE Transactions on Information Theory*, 50(5):748–767, May 2004.
[3] A. Jovicic, P. Viswanath, and S. R. Kulkarni. Upper bounds on transport capacity of wireless networks. *IEEE Transactions on Information Theory*, 50(11):2555– 2565, November 2004.
[4] F. Xue, L. Xie, and P. R. Kumar. The transport capacity of wireless networks over fading channels. *IEEE Transactions on Information Theory*, 51(3):834–847, March 2005.
[5] 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 Transactions on Information Theory*, 52(6):2313–2328, June 2006.
[6] M. Franceschetti, O. Dousse, D. Tse, and P. Thiran. Closing the gap in the capacity of wireless networks via percolation theory. *IEEE Transactions on Information Theory*, 53(3):1009–1018, March 2007.
[7] P. Gupta and P. R. Kumar. Towards an information theory of large networks: An achievable rate region. *IEEE Transactions on Information Theory*, 49(8):1877–1894, August 2003.
[8] L. Xie and P. R. Kumar. An achievable rate for the multiple-level relay channel. *IEEE Transactions on Information Theory*, 51(4):1348–1358, April 2005.
[9] G. Kramer, M. Gastpar, and P. Gupta. Cooperative strategies and capacity theorems for relay networks. *IEEE Transactions on Information Theory*, 51(9):3037–3063, September 2005.
[10] S. Aeron and V. Saligrama. Wireless ad hoc networks: Strategies and scaling laws for the fixed SNR regime. *IEEE Transactions on Information Theory*, 53(6):2044–2059, June 2007.
[11] A. Özgür, O. Lévêque, and D. Tse. Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks. *IEEE Transactions on Information Theory*, 53(10):3549–3572, October 2007.
[12] E. Linial, N. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, June 1995.
[13] G. J. Foschini. Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas. *AT&T Bell Labs. Technical Journal*, 1(2):41–59, 1996.
[14] I. E. Telatar. Capacity of multi-antenna gaussian channels. *European Transactions on Telecommunications*, 10(6):585–595, November 1999.