

The Balanced Unicast and Multicast Capacity Regions of Large Wireless Networks

Urs Niesen, Piyush Gupta, and Devavrat Shah

Abstract—We consider the question of determining the scaling of the n^2 -dimensional balanced unicast and the $n2^n$ -dimensional balanced multicast capacity regions of a wireless network with n nodes placed uniformly at random in a square region of area n and communicating over Gaussian fading channels. We identify this scaling of both the balanced unicast and multicast capacity regions in terms of $\Theta(n)$, out of 2^n total possible, cuts. These cuts only depend on the geometry of the locations of the source nodes and their destination nodes and the traffic demands between them, and thus can be readily evaluated. Our results are constructive and provide optimal (in the scaling sense) communication schemes.

Index Terms—Capacity region, capacity scaling, multicast, multicommodity flow, wireless networks.

I. INTRODUCTION

CHARACTERIZING the capacity region of wireless networks is a long standing open problem in information theory. The exact capacity region is, in fact, not known for even simple networks like a three node relay channel or a four node interference channel. In this paper, we consider the question of approximately determining the unicast and multicast capacity regions of wireless networks by identifying their scaling in terms of the number of nodes in the network.

A. Related Work

In the last decade, exciting progress has been made towards approximating the capacity region of wireless networks. We shall mention a small subset of work related to this paper.

We first consider unicast traffic. The unicast capacity region of a wireless network with n nodes is the set of all simultaneously achievable rates between all possible n^2 source–destination pairs. Since finding this unicast capacity region of a wireless

network exactly seems to be intractable, Gupta and Kumar proposed a simpler but insightful question in [1]. First, instead of asking for the entire n^2 -dimensional unicast capacity region of a wireless network with n nodes, attention was restricted to the scenario where each node is source exactly once and chooses its destination uniformly at random from among all the other nodes. All these n source–destination pairs communicate at the same rate, and the interest is in finding the maximal achievable such rate. Second, instead of insisting on finding this maximal rate exactly, they focused on its asymptotic behavior as the number of nodes n grows to infinity.

This setup has indeed turned out to be more amenable to analysis. In [1], it was shown that under random placement of nodes in a given region and under certain models of communication motivated by current technology (called *protocol channel model* in the following), the per-node rate for random source–destination pairing with uniform traffic can scale at most as $O(n^{-1/2})$ and this can be achieved (within poly-logarithmic factor in n) by a simple scheme based on multihop communication. Many works since then have broadened the channel and communication models under which similar results can be proved (see, for example, [2]–[13]). In particular, under the *Gaussian fading channel model* with a power-loss of $r^{-\alpha}$ for signals sent over a distance of r , it was shown in [12] that in extended wireless networks (i.e., n nodes are located in a region of area $\Theta(n)$) the largest uniformly achievable per-node rate under random source–destination pairing scales essentially like $\Theta(n^{1-\min\{3,\alpha\}/2})$.

Analyzing such random source–destination pairing with uniform traffic yields information about the n^2 -dimensional unicast capacity region along one dimension. Hence, the results in [1] and in [12] mentioned above provide a complete characterization of the scaling of this one-dimensional slice of the capacity region for the protocol and Gaussian fading channel models, respectively. It is, therefore, natural to ask if the scaling of the entire n^2 -dimensional unicast capacity region can be characterized. To this end, we describe two related approaches taken in recent works.

One approach, taken by Madan *et al.* [14], builds upon the celebrated works of Leighton and Rao [15] and Linial *et al.* [16] on the approximate characterization of the unicast capacity region of capacitated wireline networks. For such wireline networks, the scaling of the unicast capacity region is determined (within a $\log(n)$ factor) by the minimum weighted cut of the network graph. As shown in [14], this naturally extends to wireless networks under the protocol channel model, providing an approximation of the unicast capacity region in this case.

Another approach, first introduced by Gupta and Kumar [1], utilizes geometric properties of the wireless network. Specifically, the notion of the *transport capacity* of a network, which

Manuscript received July 17, 2008; revised June 24, 2009. Current version published April 21, 2010. The work of U. Niesen and D. Shah was supported in part by DARPA grant (ITMANET) 18870740-37362-C and in part by AFOSR grant (complex networks) MIT subaward 00006517. The work of P. Gupta was supported in part by NSF under grant CNS-0519535 and in part by AFOSR under grant FA9550-09-1-0317. The material in this paper was presented in part at the Allerton Conference on Communication, Control, and Computing, Monticello, IL, September 2008, and in part at the IEEE INFOCOM Conference, Rio de Janeiro, Brazil, April 2009.

U. Niesen was with the Laboratory for Information and Decision Systems at the Massachusetts Institute of Technology, Cambridge, MA 02139 USA. He is now with the Mathematics of Networks and Communications Research Department, Alcatel-Lucent, Bell Laboratories, Murray Hill, NJ 07974 USA (e-mail: urs.niesen@alcatel-lucent.com)

P. Gupta is with the Mathematics of Networks and Communications Research Department, Bell Labs, Alcatel-Lucent, Bell Laboratories, Murray Hill, NJ 07974 USA (e-mail: pgupta@research.bell-labs.com).

D. Shah is with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: devavrat@mit.edu).

Communicated by G. Kramer, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2010.2043979

is the rate-distance product summed over all source-destination pairs, was introduced in [1]. It was shown that in an extended wireless network with n nodes and under the protocol channel model, the transport capacity can scale at most as $\Theta(n)$. This bound on the transport capacity provides a hyper-plane which has the capacity region and origin on the same side. Through a repeated application of this transport capacity bound at different scales [17], [18] obtained an implicit characterization of the unicast capacity region under the protocol channel model.

For the Gaussian fading channel model, asymptotic upper bounds for the transport capacity were obtained in [2] and [3], and for more general distance weighted sum rates in [19].

So far, we have only considered unicast traffic. We now turn to multicast traffic. The multicast capacity region of a wireless network with n nodes is the set of all simultaneously achievable rates between all possible $n2^n$ source-multicast-group pairs. Instead of considering this multicast capacity region directly, various authors have analyzed the scaling of restricted traffic patterns under a protocol channel model assumption (see [20]–[24], among others). For example, in [20], Li *et al.* obtained a scaling characterization under a protocol channel model and random node placement for multicast traffic when each node chooses a certain number of its destinations uniformly at random. Independently, in [21], Shakkottai *et al.* considered a similar setup and also obtained the precise scaling when sources and their multicast destinations are chosen at random. Both these results assume a protocol channel model and are hence not information-theoretic. Furthermore, they provide information about the scaling of the $n2^n$ -dimensional multicast capacity region only along one particular dimension.

B. Our Contributions

Despite the long list of results, the question of approximately characterizing the unicast capacity region under the Gaussian fading channel model remains far from being resolved. In fact, for Gaussian fading channels, the only traffic pattern that is well understood is random source-destination pairing with uniform rate. This is limiting in several aspects. First, by choosing for each source a destination at random, most source-destination pairs will be at a distance of the diameter of the network with high probability, i.e., at distance $\Theta(\sqrt{n})$ for an extended network. However, in many wireless networks, some degree of locality of traffic can be expected. Second, all source-destination pairs are assumed to be communicating at uniform rate. Again, in many settings we would expect nodes to be generating traffic at widely varying rates. Third, each node is source exactly once, and destination on average once. However, in many scenarios, the same source node (e.g., a server) might transmit data to many different destination nodes, or the same destination node might request data from many different source nodes. All these heterogeneities in the traffic demands can result in different scaling behavior of the performance of the wireless network than what is obtained for random source-destination pairing with uniform rate.

As is pointed out in the last section, even less is known about the multicast capacity region under Gaussian fading. In fact, the only available results are for the protocol channel model, and even there only for special traffic patterns resulting from

randomly choosing sources and their multicast groups and assuming uniform rate. To the best of our knowledge, no information-theoretic results (i.e., assuming Gaussian fading channels) are available even for special traffic patterns.

We address these issues by analyzing the scaling of a broad class of traffic, termed *balanced traffic* in the following, in a wireless network of n randomly placed nodes under a Gaussian fading channel model. The notion of balanced traffic is a natural generalization of symmetric traffic, in which the data to be transmitted from a node u to a node v is equal to the amount of data to be transmitted from v to u . We analyze the scaling of the set of achievable balanced unicast traffic (the *balanced unicast capacity region*) and achievable balanced multicast traffic (the *balanced multicast capacity region*). The balanced unicast capacity region provides information about $n^2 - n$ of the n^2 dimensions of the unicast capacity region; the balanced multicast capacity region provides information about $n2^n - n$ of the $n2^n$ dimensions of the multicast capacity region.

As a first set of results of this paper, we present an approximate characterization of the balanced unicast and multicast capacity regions. We show that both regions can be approximated by a polytope with less than $2n$ faces, each corresponding to a distinct cut (i.e., a subset of nodes) in the wireless network. This polyhedral characterization provides a succinct approximate description of the balanced unicast and multicast capacity regions even for large values of n . Moreover, it shows that only $2n$ out of 2^n possible cuts in the wireless network are asymptotically relevant and reveals the geometric structure of these relevant cuts.

Second, we establish the approximate equivalence of the wireless network and a wireline tree graph, in the sense that balanced traffic can be transmitted reliably over the wireless network if and only if approximately the same traffic can be routed over the tree graph. This equivalence is the key component in the derivation of the approximation result for the balanced unicast and multicast capacity regions and provides insight into the structure of large wireless networks.

Third, we propose a novel three-layer communication architecture that achieves (in the scaling sense) the entire balanced unicast and multicast capacity regions. The top layer of this scheme treats the wireless network as the aforementioned tree graph and routes messages between sources and their destinations—dealing with heterogeneous traffic demands. The middle layer of this scheme provides this tree abstraction to the top layer by appropriately distributing and concentrating traffic over the wireless network—choosing the level of cooperation in the network. The bottom layer implements this distribution and concentration of messages in the wireless network—dealing with interference and noise. The approximate optimality of this three-layer architecture implies that a separation based approach, in which routing is performed independently of the physical layer, is order-optimal. In other words, techniques such as network coding can provide at most a small (in the scaling sense) multiplicative gain for transmission of balanced unicast or multicast traffic in wireless networks.

C. Organization

The remainder of this paper is organized as follows. Section II introduces the network model and notation. Section III presents

our main results. We illustrate these results in Section IV by analyzing various example scenarios with heterogeneous unicast and multicast traffic patterns. Section V provides a high level description of the proposed communication schemes. Sections VI–VIII contain proofs. Finally, Sections IX and X contain discussions and concluding remarks.

II. MODELS AND NOTATION

In this section, we discuss network and traffic models, and we introduce some notational conventions.

A. Network Model

Consider the square region

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

and let $V(n) \subset A(n)$ be a set of $|V(n)| = n$ nodes on $A(n)$. Each such node represents a wireless device, and the n nodes together form a wireless network. This setting with n nodes on a square of area n is referred to as an *extended network*. Throughout this paper, we consider this extended network setting. However, all results carry over for *dense networks*, where n nodes are placed on a square of unit area (see Section IX-E for the details).

We use the same channel model as in [12]. Namely, the received signal at node v and time t is

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

for all $v \in V(n), t \in \mathbb{N}$, where the $\{x_u[t]\}_{u,t}$ are the signals sent by the nodes in $V(n)$. We impose an average power constraint of 1 on the signal $\{x_u[t]\}_t$ for every node $u \in V(n)$. The additive noise terms $\{z_v[t]\}_{v,t}$ are independent and identically distributed (i.i.d.) circularly symmetric complex Gaussian random variables with mean 0 and variance 1, and

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

for *path-loss exponent* $\alpha > 2$, and where $r_{u,v}$ is the Euclidean distance between u and v . As a function of $u, v \in V(n)$, we assume that $\{\theta_{u,v}[t]\}_{u,v}$ are i.i.d.¹ with uniform distribution on $[0, 2\pi)$. As a function of t , we either assume that $\{\theta_{u,v}[t]\}_t$ is stationary and ergodic, which is called *fast fading* in the following, or we assume $\{\theta_{u,v}[t]\}_t$ is constant, which is called *slow fading* in the following. In either case, we assume full channel state information (CSI) is available at all nodes, i.e., each node knows all $\{h_{u,v}[t]\}_{u,v}$ at time t . This full CSI assumption is rather strong, and so is worth commenting on. All the converse results presented are proved under the full CSI assumption and

¹It is worth pointing out that recent results [25] suggest that under certain assumptions on scattering elements, for $\alpha \in (2, 3)$ and very large values of n , the i.i.d. phase assumption does not accurately reflect the physical behavior of the wireless channel. However, in follow-up work [26], the authors show that under different assumptions on the scatterers, this assumption is still justified in the $\alpha \in (2, 3)$ regime even for very large values of n . This indicates that the issue of channel modeling for large networks in the low path-loss regime is somewhat delicate and requires further investigation.

are, hence, also valid under more realistic assumptions on the availability of CSI. Moreover, it can be shown that for achievability only 2-bit quantized CSI is necessary for path-loss exponent $\alpha \in (2, 3]$ and no CSI is necessary for $\alpha > 3$ to achieve the same scaling behavior.

B. Traffic Model

A *unicast traffic matrix* $\lambda^{\text{UC}} \in \mathbb{R}_+^{n \times n}$ associates with each pair $u, w \in V(n)$ the rate $\lambda_{u,w}^{\text{UC}}$ at which node u wants to communicate to node w . We assume that messages for distinct source–destination pairs (u, w) are independent. However, we allow the same node u to be source for multiple destinations, and the same node w to be destination for multiple sources. In other words, we consider general unicast traffic. The *unicast capacity region* $\Lambda^{\text{UC}}(n) \subset \mathbb{R}_+^{n \times n}$ of the wireless network is the collection of achievable unicast traffic matrices, i.e., $\lambda^{\text{UC}} \in \Lambda^{\text{UC}}(n)$ if and only if every source–destination pair $(u, w) \in V(n) \times V(n)$ can reliably communicate independent messages at rate $\lambda_{u,w}^{\text{UC}}$.

A *multicast traffic matrix* $\lambda^{\text{MC}} \in \mathbb{R}_+^{n \times 2^n}$ associates with each pair $u \in V(n), W \subset V(n)$ the rate $\lambda_{u,W}^{\text{MC}}$ at which node u wants to multicast a message to the nodes in W . In other words, all nodes in W want to obtain the same message from u . We assume that messages for distinct source–multicast-group pairs (u, W) are independent. However, we allow the same node u to be source for several multicast-groups, and the same set W of nodes to be multicast destination for multiple sources. In other words, we consider general multicast traffic. The *multicast capacity region* $\Lambda^{\text{MC}}(n) \subset \mathbb{R}_+^{n \times 2^n}$ is the collection of achievable multicast traffic matrices, i.e., $\lambda^{\text{MC}} \in \Lambda^{\text{MC}}(n)$ if and only if every source–multicast-group pair (u, W) can reliably communicate independent messages at rate $\lambda_{u,W}^{\text{MC}}$.

The following example illustrates the concept of unicast and multicast traffic matrices.

Example 1: Assume $n = 4$, and label the nodes as $\{u_i\}_{i=1}^4 = V(n)$. Assume further node u_1 needs to transmit a message $m_{1,2}$ to node u_2 at rate 1 bit per channel use, and an independent message $m_{1,3}$ to node u_3 at rate 2 bits per channel use. Node u_2 needs to transmit a message $m_{2,3}$ to node u_3 at rate 4 bits per channel use. All the messages $m_{1,2}, m_{1,3}, m_{2,3}$ are independent. This traffic pattern can be described by a unicast traffic matrix $\lambda^{\text{UC}} \in \mathbb{R}_+^{4 \times 4}$ with $\lambda_{u_1, u_2}^{\text{UC}} = 1, \lambda_{u_1, u_3}^{\text{UC}} = 2, \lambda_{u_2, u_3}^{\text{UC}} = 4$, and $\lambda_{u,w}^{\text{UC}} = 0$ otherwise. Note that in this example node u_1 is source for two (independent) messages, and node u_3 is destination for two (again independent) messages. Node u_4 in this example is neither source nor destination for any message and can be understood as a helper node.

Assume now that node u_1 needs to transmit the same message $m_{1,\{2,3,4\}}$ to all nodes u_2, u_3, u_4 at a rate of 1 bit per channel use, and an independent message $m_{1,\{2\}}$ to only node 2 at rate 2 bits per channel use. Node 2 needs to transmit a message $m_{2,\{1,3\}}$ to both u_1, u_3 at rate 4 bits per channel use. All the messages $m_{1,\{2,3,4\}}, m_{1,\{2\}}, m_{2,\{1,3\}}$ are independent. This traffic pattern can be described by a multicast traffic matrix $\lambda^{\text{MC}} \in \mathbb{R}_+^{4 \times 16}$ with $\lambda_{u_1, \{u_2, u_3, u_4\}}^{\text{MC}} = 1, \lambda_{u_1, \{u_2\}}^{\text{MC}} = 2, \lambda_{u_2, \{u_1, u_3\}}^{\text{MC}} = 4$, and $\lambda_{u,W}^{\text{MC}} = 0$ otherwise. Note that in this

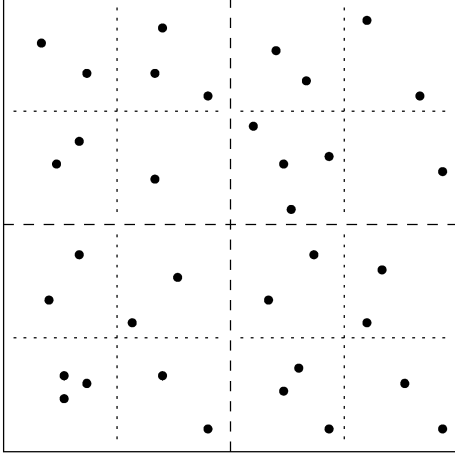


Fig. 1. Square-grids with $0 \leq \ell \leq 2$. The grid at level $\ell = 0$ is the area $A(n)$ itself. The grid at level $\ell = 1$ is indicated by the dashed lines. The grid at level $\ell = 2$ by the dashed and the dotted lines. Assume for the sake of example that the subsquares are numbered from left to right and then from bottom to top (the precise order of numbering is immaterial). Then, $V_{0,1}(n)$ are all the nodes $V(n)$, $V_{1,1}(n)$ are the nine nodes in the lower left corner (separated by dashed lines), and $V_{2,1}(n)$ are the three nodes in the lower left corner (separated by dotted lines).

example node u_1 is source for two (independent) multicast messages, and node 2 and 3 are destinations for more than one message. The message $m_{1,\{2,3,4\}}$ is destined for all the nodes in the network and can, hence, be understood as a broadcast message. The message $m_{1,\{2\}}$ is only destined for one node and can, hence, be understood as a private message. \diamond

In the following, we will be interested in *balanced* traffic matrices that satisfy certain symmetry properties. Consider a symmetric unicast traffic matrix λ^{UC} satisfying $\lambda_{u,w}^{\text{UC}} = \lambda_{w,u}^{\text{UC}}$ for all node pairs $u, w \in V(n)$. The notion of a balanced traffic matrix generalizes this idea of symmetric traffic.

Before we provide a precise definition of balanced traffic, we need to introduce some notation. Partition $A(n)$ into several square-grids. The ℓ th square-grid divides $A(n)$ into 4^ℓ squares, each of sidelength $2^{-\ell}\sqrt{n}$, denoted by $\{A_{\ell,i}(n)\}_{i=1}^{4^\ell}$. Let $V_{\ell,i}(n) \subset V(n)$ be the nodes in $A_{\ell,i}(n)$ (see Fig. 1). The square grids in levels $\ell \in \{1, \dots, L(n)\}$ with²

$$L(n) \triangleq \frac{1}{2} \log(n)(1 - \log^{-1/2}(n))$$

will be of particular importance. Note that $L(n)$ is chosen such that

$$4^{-L(n)}n = n^{\log^{-1/2}(n)}$$

and, hence

$$\lim_{n \rightarrow \infty} |A_{L(n),i}(n)| = \lim_{n \rightarrow \infty} 4^{-L(n)}n = \infty.$$

while at the same time

$$|A_{L(n),i}(n)| = 4^{-L(n)}n \leq n^{o(1)}$$

as $n \rightarrow \infty$. In other words, the area of the region $A_{L(n),i}(n)$ at level $\ell = L(n)$ grows to infinity as $n \rightarrow \infty$, but much slower than n .

²All logarithms are with respect to base 2.

A unicast traffic matrix λ^{UC} is γ -balanced if

$$\sum_{u \notin V_{\ell,i}(n)} \sum_{w \in V_{\ell,i}(n)} \lambda_{u,w}^{\text{UC}} \leq \gamma \sum_{u \in V_{\ell,i}(n)} \sum_{w \notin V_{\ell,i}(n)} \lambda_{u,w}^{\text{UC}} \quad (1)$$

for all $\ell \in \{1, \dots, L(n)\}$ and $i \in \{1, \dots, 4^\ell\}$. In other words, for a balanced unicast traffic matrix the amount of traffic to the nodes $V_{\ell,i}(n)$ is not much larger than the amount of traffic from them. In particular, all symmetric traffic matrices, i.e., satisfying $\lambda_{u,w}^{\text{UC}} = \lambda_{w,u}^{\text{UC}}$, are 1-balanced. Denote by $\mathcal{B}^{\text{UC}}(n) \subset \mathbb{R}_+^{n \times n}$ the collection of all $\gamma(n)$ -balanced unicast traffic matrices for some fixed $\gamma(n) = n^{o(1)}$. In the following, we refer to traffic matrices $\lambda^{\text{UC}} \in \mathcal{B}^{\text{UC}}(n)$ simply as balanced traffic matrices. The *balanced unicast capacity region* $\Lambda^{\text{BUC}}(n) \subset \mathbb{R}_+^{n \times n}$ of the wireless network is the collection of balanced unicast traffic matrices that are achievable, i.e.,

$$\Lambda^{\text{BUC}}(n) \triangleq \Lambda^{\text{UC}}(n) \cap \mathcal{B}^{\text{UC}}(n).$$

Note that (1) imposes at most n linear inequality constraints, and, hence, $\Lambda^{\text{UC}}(n)$ and $\Lambda^{\text{BUC}}(n)$ coincide along at least $n^2 - n$ of n^2 total dimensions.

A multicast traffic matrix λ^{MC} is γ -balanced if

$$\sum_{u \notin V_{\ell,i}(n)} \sum_{\substack{W \subset V(n): \\ W \cap V_{\ell,i}(n) \neq \emptyset}} \lambda_{u,W}^{\text{MC}} \leq \gamma \sum_{u \in V_{\ell,i}(n)} \sum_{\substack{W \subset V(n): \\ W \setminus V_{\ell,i}(n) \neq \emptyset}} \lambda_{u,W}^{\text{MC}} \quad (2)$$

for all $\ell \in \{1, \dots, L(n)\}$, $i \in \{1, \dots, 4^\ell\}$. Thus, for γ -balanced multicast traffic, the amount of traffic to the nodes $V_{\ell,i}(n)$ is not much larger than the amount of traffic from them. This is the natural generalization of the notion of γ -balanced unicast traffic to the multicast case. Denote by $\mathcal{B}^{\text{MC}}(n) \subset \mathbb{R}_+^{n \times 2^n}$ the collection of all $\gamma(n)$ -balanced multicast traffic matrices for some fixed $\gamma(n) = n^{o(1)}$. As before, we will refer to a multicast traffic matrices $\lambda^{\text{MC}} \in \mathcal{B}^{\text{MC}}(n)$ simply as balanced multicast traffic matrices. The *balanced multicast capacity region* $\Lambda^{\text{BMC}}(n) \subset \mathbb{R}_+^{n \times 2^n}$ of the wireless network is the collection of balanced multicast traffic matrices that are achievable, i.e.,

$$\Lambda^{\text{BMC}}(n) \triangleq \Lambda^{\text{MC}}(n) \cap \mathcal{B}^{\text{MC}}(n).$$

Equation (2) imposes at most n linear inequality constraints, and, hence, $\Lambda^{\text{MC}}(n)$ and $\Lambda^{\text{BMC}}(n)$ coincide along at least $n2^n - n$ of $n2^n$ total dimensions.

C. Notational Conventions

Throughout, $\{K_i\}_i$, K , \tilde{K} , \dots , indicate strictly positive finite constants independent of n and ℓ . To simplify notation, we assume, when necessary, that large real numbers are integers and omit $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ operators. For the same reason, we also suppress dependence on n within proofs whenever this dependence is clear from the context.

III. MAIN RESULTS

In this section, we present the main results of this paper. In Section III-A, we provide an approximate (i.e., scaling) characterization of the entire balanced unicast capacity region

$\Lambda^{\text{BUC}}(n)$ of the wireless network, and in Section III-B, we provide a scaling characterization of the entire balanced multicast capacity region $\Lambda^{\text{BMC}}(n)$. In Section III-C, we discuss implications of these results on the behavior of the unicast and multicast capacity regions for large values of n . In Section III-D, we consider computational aspects.

A. Balanced Unicast Capacity Region

Here, we present a scaling characterization of the complete balanced unicast capacity region $\Lambda^{\text{BUC}}(n)$.

Define

$$\hat{\Lambda}^{\text{UC}}(n) \triangleq \left\{ \lambda^{\text{UC}} \in \mathbb{R}_+^{n \times n} : \right. \\ \left. \sum_{u \in V_{\ell,i}(n)} \sum_{w \notin V_{\ell,i}(n)} \lambda_{u,w}^{\text{UC}} \leq (4^{-\ell} n)^{2-\min\{3,\alpha\}/2} \right. \\ \left. \forall \ell \in \{1, \dots, L(n)\}, i \in \{1, \dots, 4^\ell\} \right. \\ \left. \sum_{w \neq u} (\lambda_{u,w}^{\text{UC}} + \lambda_{w,u}^{\text{UC}}) \leq 1 \forall u \in V(n) \right\} \quad (3)$$

and set

$$\hat{\Lambda}^{\text{BUC}}(n) \triangleq \hat{\Lambda}^{\text{UC}}(n) \cap \mathcal{B}^{\text{UC}}(n).$$

$\hat{\Lambda}^{\text{BUC}}(n)$ is the collection of all balanced unicast traffic matrices λ^{UC} such that for various cuts $S \subset V(n)$ in the network, the total traffic demand (in either one or both directions)

$$\sum_{u \in S} \sum_{w \notin S} \lambda_{u,w}^{\text{UC}} \\ \sum_{u \in S} \sum_{w \notin S} (\lambda_{u,w}^{\text{UC}} + \lambda_{w,u}^{\text{UC}})$$

across the cut S is not too big. Note that the number of cuts S we need to consider is actually quite small. In fact, there are at most n cuts of the form $S = V_{\ell,i}(n)$ for $\ell \in \{1, \dots, L(n)\}$, and there are n cuts of the form $S = \{u\}$ for $u \in V(n)$. Hence, $\hat{\Lambda}^{\text{BUC}}(n)$ is described by at most $2n$ cuts.

The next theorem shows that $\hat{\Lambda}^{\text{BUC}}(n)$ is approximately (in the scaling sense) equal to the balanced unicast capacity region $\Lambda^{\text{BUC}}(n)$ of the wireless network.

Theorem 1: Under either fast or slow fading, for any $\alpha > 2$, there exist

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

such that

$$b_1(n) \hat{\Lambda}^{\text{BUC}}(n) \subset \Lambda^{\text{BUC}}(n) \subset b_2(n) \hat{\Lambda}^{\text{BUC}}(n)$$

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

We point out that Theorem 2 holds only with probability $1 - o(1)$ for different reasons for the fast and slow fading cases. Under fast fading, the theorem holds only for node placements that are “regular enough”. The node placement itself is random, and we show that the required regularity property is satisfied with high probability as $n \rightarrow \infty$. Under slow fading, the theorem holds under the same regularity requirements on the node

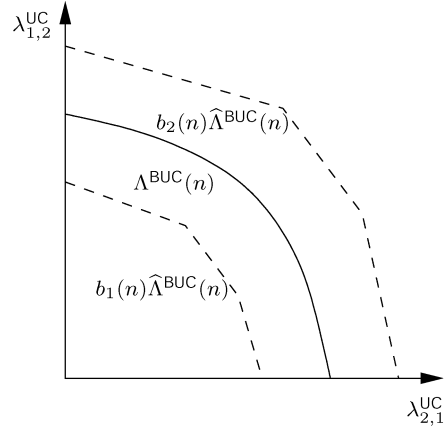


Fig. 2. Set $\hat{\Lambda}^{\text{BUC}}(n)$ approximates the balanced unicast capacity region $\Lambda^{\text{BUC}}(n)$ of the wireless network in the sense that $b_1(n)\hat{\Lambda}^{\text{BUC}}(n)$ (with $b_1(n) \geq n^{-o(1)}$) provides an inner bound to $\Lambda^{\text{BUC}}(n)$ and $b_2(n)\hat{\Lambda}^{\text{BUC}}(n)$ (with $b_2(n) = O(\log^6(n))$) provides an outer bound to $\Lambda^{\text{BUC}}(n)$. The figure shows two dimensions (namely $\lambda_{1,2}^{\text{UC}}$ and $\lambda_{2,1}^{\text{UC}}$) of the n^2 -dimensional set $\Lambda^{\text{BUC}}(n)$.

placement, but now it also only holds with high probability for the realization of the fading $\{\theta_{u,v}\}_{u,v}$.

Theorem 2 provides a tight scaling characterization of the entire balanced unicast capacity region $\Lambda^{\text{BUC}}(n)$ of the wireless network as depicted in Fig. 2. The approximation is within a factor $n^{\pm o(1)}$. This factor can be further sharpened as is discussed in detail in Section IX-B.

We point out that for large values of path-loss exponent ($\alpha > 5$) the restriction to balanced traffic can be removed, yielding a tight scaling characterization of the entire n^2 -dimensional unicast capacity region $\Lambda^{\text{UC}}(n)$. See Section IX-D for the details. For $\alpha \in (2, 5]$, bounds on achievable rates for traffic that is not balanced are discussed in Section IX-C.

B. Balanced Multicast Capacity Region

We now present an approximate characterization of the complete balanced multicast capacity region $\Lambda^{\text{BMC}}(n)$.

Define

$$\hat{\Lambda}^{\text{MC}}(n) \triangleq \left\{ \lambda^{\text{MC}} \in \mathbb{R}_+^{n \times 2^n} : \right. \\ \left. \sum_{u \in V_{\ell,i}(n)} \sum_{\substack{W \subset V(n): \\ W \setminus V_{\ell,i}(n) \neq \emptyset}} \lambda_{u,W}^{\text{MC}} \leq (4^{-\ell} n)^{2-\min\{3,\alpha\}/2} \right. \\ \left. \forall \ell \in \{1, \dots, L(n)\}, i \in \{1, \dots, 4^\ell\} \right. \\ \left. \sum_{\substack{W \subset V(n): \\ W \setminus \{u\} \neq \emptyset}} \lambda_{u,W}^{\text{MC}} + \sum_{\substack{\tilde{u} \neq u \\ W \subset V(n): \\ u \in W}} \lambda_{\tilde{u},W}^{\text{MC}} \leq 1 \right. \\ \left. \forall u \in V(n) \right\} \quad (4)$$

and set

$$\hat{\Lambda}^{\text{BMC}}(n) \triangleq \hat{\Lambda}^{\text{MC}}(n) \cap \mathcal{B}^{\text{MC}}(n).$$

The definition of $\hat{\Lambda}^{\text{BMC}}(n)$ is similar to the definition of $\hat{\Lambda}^{\text{BUC}}(n)$ in (3). $\hat{\Lambda}^{\text{BMC}}(n)$ is the collection of all balanced multicast traffic

matrices λ^{MC} such that for various cuts $S \subset V(n)$ in the network, the total traffic demand (in either one or both directions)

$$\sum_{u \in S} \sum_{\substack{W \subset V(n): \\ W \setminus S \neq \emptyset}} \lambda_{u,W}^{\text{MC}} \\ \sum_{u \in S} \sum_{\substack{W \subset V(n): \\ W \setminus S \neq \emptyset}} \lambda_{u,W}^{\text{MC}} + \sum_{u \notin S} \sum_{\substack{W \subset V(n): \\ W \cap S \neq \emptyset}} \lambda_{u,W}^{\text{MC}}$$

across the cut S is not too big. Note that, unlike in the definition of $\hat{\Lambda}^{\text{BUC}}(n)$, we count $\lambda_{u,W}^{\text{MC}}$ as crossing the cut S if $u \in S$ and $W \setminus S \neq \emptyset$, i.e., if there is at least one node w in the multicast destination group W that lies outside S . The number of such cuts S we need to consider is at most $2n$, as in the unicast case.

The next theorem shows that $\hat{\Lambda}^{\text{BMC}}(n)$ is approximately (in the scaling sense) equal to the balanced multicast capacity region $\Lambda^{\text{BMC}}(n)$ of the wireless network.

Theorem 2: Under either fast or slow fading, for any $\alpha > 2$, there exist

$$b_3(n) \geq n^{-o(1)} \\ b_4(n) = O(\log^6(n))$$

such that

$$b_3(n)\hat{\Lambda}^{\text{BMC}}(n) \subset \Lambda^{\text{BMC}}(n) \subset b_4(n)\hat{\Lambda}^{\text{BMC}}(n)$$

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

As with Theorem 1, Theorem 2 holds only with probability $1 - o(1)$ for different reasons for the fast and slow fading cases. Theorem 2 implies that the quantity $\hat{\Lambda}^{\text{BMC}}(n)$ determines the scaling of the balanced multicast capacity region $\Lambda^{\text{BMC}}(n)$. The approximation is up to a factor $n^{\pm o(1)}$ as in the unicast case, and can again be sharpened (see the discussion in Section IX-B). As in the unicast case, for $\alpha > 5$ the restriction of balanced traffic can be dropped resulting in a scaling characterization of the entire n^{2^n} -dimensional multicast capacity region $\Lambda^{\text{MC}}(n)$. The details can be found in Section IX-D. Similarly, we can obtain bounds on achievable rates for traffic that is not balanced, as is discussed in Section IX-C.

C. Implications of Theorems 1 and 2

Theorems 1 and 2 can be applied in two ways. First, the theorems can be used to analyze the asymptotic achievability of a sequence of traffic matrices. Consider the unicast case, and let $\{\lambda^{\text{UC}}(n)\}_{n \geq 1}$ be a sequence of balanced unicast traffic matrices with $\lambda^{\text{UC}}(n) \in \mathbb{R}_+^{n \times n}$. Define

$$\rho_{\lambda^{\text{UC}}(n)}^* \triangleq \sup\{\rho : \rho\lambda^{\text{UC}}(n) \in \Lambda^{\text{BUC}}(n)\} \\ \hat{\rho}_{\lambda^{\text{UC}}(n)}^* \triangleq \sup\{\hat{\rho} : \hat{\rho}\lambda^{\text{UC}}(n) \in \hat{\Lambda}^{\text{BUC}}(n)\}$$

i.e., $\rho_{\lambda^{\text{UC}}(n)}^*$ is the largest multiplier ρ such that the scaled traffic matrix $\rho\lambda^{\text{UC}}(n)$ is contained in $\Lambda^{\text{BUC}}(n)$ (and similar for $\hat{\rho}_{\lambda^{\text{UC}}(n)}^*$ with respect to $\hat{\Lambda}^{\text{BUC}}(n)$). Then Theorem 1 provides

asymptotic information about the achievability of $\{\lambda^{\text{UC}}(n)\}_{n \geq 1}$ in the sense that³

$$\lim_{n \rightarrow \infty} \frac{\log(\rho_{\lambda^{\text{UC}}(n)}^*)}{\log(n)} = \lim_{n \rightarrow \infty} \frac{\log(\hat{\rho}_{\lambda^{\text{UC}}(n)}^*)}{\log(n)}.$$

Theorem 2 can be used similarly to analyze sequences of balanced multicast traffic matrices. Several applications of this approach are explored in Section IV.

Second, Theorems 1 and 2 provide information about the shape of the balanced unicast and multicast capacity regions $\Lambda^{\text{BUC}}(n)$ and $\Lambda^{\text{BMC}}(n)$. Consider again the unicast case. We now argue that even though the approximation $\hat{\Lambda}^{\text{BUC}}(n)$ of $\Lambda^{\text{BUC}}(n)$ is only up to $n^{\pm o(1)}$ scaling, its shape is largely preserved.

To illustrate this point, consider a rectangle

$$R(n) \triangleq [0, r_1(n)] \times [0, r_2(n)]$$

and let

$$\hat{R}(n) \triangleq [0, \hat{r}_1(n)] \times [0, \hat{r}_2(n)]$$

where

$$\hat{r}_i \triangleq b_i(n)r_i(n)$$

for some $b_i(n) = n^{\pm o(1)}$, be its approximation. The shape of $R(n)$ is then determined by the ratio between $r_1(n)$ and $r_2(n)$. For example, assume $r_1(n) = n^\beta r_2(n)$. Then

$$\frac{\hat{r}_1(n)}{\hat{r}_2(n)} = n^{\beta \pm o(1)} = n^{\pm o(1)} \frac{r_1(n)}{r_2(n)}$$

i.e.,

$$\lim_{n \rightarrow \infty} \frac{\log(r_1(n)/r_2(n))}{\log(n)} = \beta = \lim_{n \rightarrow \infty} \frac{\log(\hat{r}_1(n)/\hat{r}_2(n))}{\log(n)}$$

and, hence, the approximation $\hat{R}(n)$ preserves the exponent of the ratio of sidelengths of $R(n)$. In other words, if the two sidelengths $r_1(n)$ and $r_2(n)$ differ on exponential scale (i.e., by a factor n^β for $\beta \neq 0$) then this shape information is preserved by the approximation $\hat{R}(n)$.

Let us now return to the balanced unicast capacity region $\Lambda^{\text{BUC}}(n)$ and its approximation $\hat{\Lambda}^{\text{BUC}}(n)$. We consider several boundary points of $\Lambda^{\text{BUC}}(n)$ and show that their behavior varies at scale n^β for various values of β . From the discussion in the previous paragraph, this implies that a significant part of the shape of $\Lambda^{\text{BUC}}(n)$ is preserved by its approximation $\hat{\Lambda}^{\text{BUC}}(n)$. First, let $\lambda^{\text{UC}} \triangleq \rho(n)\mathbf{1}$ for some scalar $\rho(n)$ depending only on n , and where $\mathbf{1}$ is the $n \times n$ matrix of all ones. If $\lambda^{\text{UC}} \in \Lambda^{\text{BUC}}(n)$ then the largest achievable value of $\rho(n)$ is $\rho^*(n) \leq n^{-\min\{3, \alpha\}/2 + o(1)}$ (by applying Theorem 1). Second, let λ^{UC} such that $\lambda_{u^*, w^*}^{\text{UC}} = \lambda_{w^*, u^*}^{\text{UC}} = \rho(n)$ for only one source–destination pair (u^*, w^*) with $u^* \neq w^*$ and $\lambda_{u, w}^{\text{UC}} = 0$, otherwise. Then $\rho^*(n)$, the largest achievable value of $\rho(n)$, satisfies

³We assume here that the limits exist; otherwise, the same statement holds for \limsup and \liminf .

$\rho^*(n) \geq n^{-o(1)}$. Hence, the boundary points of $\Lambda^{\text{BUC}}(n)$ vary at least from $n^{-\min\{3,\alpha\}/2+o(1)}$ to $n^{-o(1)}$, and this variation on exponential scale is preserved by $\hat{\Lambda}^{\text{BUC}}(n)$.

Again, a similar analysis is possible also for the multicast capacity region, showing that the approximate balanced multicast capacity region $\hat{\Lambda}^{\text{BMC}}(n)$ preserves the shape of the balanced multicast capacity region $\Lambda^{\text{BMC}}(n)$ on exponential scale.

D. Computational Aspects

Since we are interested in large wireless networks, computational aspects are of importance. In this section, we show that the approximate characterizations $\hat{\Lambda}^{\text{BUC}}(n)$ and $\hat{\Lambda}^{\text{BMC}}(n)$ in Theorems 1 and 2 provide a computationally efficient approximate description of the balanced unicast and multicast capacity regions $\Lambda^{\text{BUC}}(n)$ and $\Lambda^{\text{BMC}}(n)$, respectively.

Consider first the unicast case. Note that $\Lambda^{\text{BUC}}(n)$ is a n^2 -dimensional set, and, hence, its shape could be rather complicated. In particular, in the special cases where the capacity region is known, its description is often in terms of cut-set bounds. Since there are 2^n possible subsets of n nodes, there are 2^n possible cut-set bounds to be considered. In other words, the description complexity of $\Lambda^{\text{BUC}}(n)$ is likely to be growing exponentially in n . On the other hand, as was pointed out in Section III-A, the description of $\hat{\Lambda}^{\text{BUC}}(n)$ is in terms of only $2n$ cuts. This implies that $\hat{\Lambda}^{\text{BUC}}(n)$ can be computed efficiently (i.e., in polynomial time in n). Hence, even though the description complexity of $\Lambda^{\text{BUC}}(n)$ is likely to be of order $\Theta(2^n)$, the description complexity of its approximation $\hat{\Lambda}^{\text{BUC}}(n)$ is only of order $\Theta(n)$ —an exponential reduction. In particular, this implies that membership $\lambda^{\text{UC}} \in \hat{\Lambda}^{\text{BUC}}(n)$ (and, hence, by Theorem 1, also the approximate achievability of the balanced unicast traffic matrix λ^{UC}) can be computed in polynomial time in the network size n . More precisely, evaluating each of the $\Theta(n)$ cuts takes at most $\Theta(n^2)$ operations, yielding a $\Theta(n^3)$ -time algorithm for approximate testing of membership in $\Lambda^{\text{BUC}}(n)$.

Consider now the multicast case. $\Lambda^{\text{BMC}}(n)$ is a $n2^n$ -dimensional set, i.e., the number of dimensions is exponentially large in n . Nevertheless, its approximation $\hat{\Lambda}^{\text{BMC}}(n)$ can (as in the unicast case) be computed by evaluating at most $2n$ cuts. This yields a very compact approximate representation of the balanced multicast capacity region $\Lambda^{\text{BMC}}(n)$ (i.e., we represent a region of exponential size in n as an intersection of only linearly many halfspaces—one halfspace corresponding to each cut). Moreover, it implies that membership $\lambda^{\text{MC}} \in \hat{\Lambda}^{\text{BMC}}(n)$ can be computed efficiently. More precisely, evaluating each of the $\Theta(n)$ cuts takes at most $|\{(u, W) : \lambda_{u,W}^{\text{MC}} > 0\}|$ operations. Thus, membership $\lambda^{\text{MC}} \in \hat{\Lambda}^{\text{BMC}}(n)$ (and, hence, by Theorem 2, also the approximate achievability of the balanced multicast traffic matrix λ^{MC}) can be tested in at most $\Theta(n)$ times more operations than required to just read the problem parameters. In other words, we have a linear time (in the length of the input) algorithm for testing membership of a multicast traffic matrix λ^{MC} in $\hat{\Lambda}^{\text{BMC}}(n)$ and, hence, for approximate testing of membership in $\Lambda^{\text{BMC}}(n)$. However, this algorithm is not necessarily polynomial time in n , since just reading the input $\lambda^{\text{MC}} \in \mathbb{R}_+^{n \times 2^n}$ itself might take exponential time in n .

IV. EXAMPLE SCENARIOS

We next illustrate the above results by determining achievable rates in a few specific wireless network scenarios with nonuniform traffic patterns.

Example 2: Multiple classes of source–destination pairs

There are K classes of source–destination pairs for some fixed K . Each source node in class i generates traffic at the same rate $\rho_i(n)$ for a destination node that is chosen randomly within distance $\Theta(n^{\beta_i/2})$, for some fixed $\beta_i \in [0, 1]$. Each node randomly picks the class it belongs to. The resulting traffic matrix is balanced (with $\gamma(n) = n^{o(1)}$) with high probability, and applying Theorem 1 shows that $\rho_i^*(n)$, the largest achievable value of $\rho_i(n)$, satisfies

$$\rho_i^*(n) = n^{\beta_i(1-\bar{\alpha}/2) \pm o(1)}$$

with probability $1 - o(1)$ for all i , and where

$$\bar{\alpha} \triangleq \min\{3, \alpha\}. \quad (5)$$

Hence, for a fixed number of classes K , source nodes in each class can obtain rates as a function of only the source–destination separation in that class.

Set $\tilde{n}_i \triangleq n^{\beta_i}$, and note that \tilde{n}_i is on the order of the expected number of nodes that are closer to a source than its destination. Then

$$\rho_i^*(n) = n^{\pm o(1)} \tilde{n}_i^{1-\bar{\alpha}/2}.$$

Now $\tilde{n}_i^{1-\bar{\alpha}/2}$ is precisely the per-node rate that is achievable for an extended network with \tilde{n}_i nodes under random source–destination pairing [12]. In other words, the local traffic pattern here allows us to obtain a rate that is as good as the one achievable under random source–destination pairing for a much smaller network. \diamond

Example 3: Traffic variation with source–destination separation

Assume each node is source for exactly one destination, chosen uniformly at random from among all the other nodes (as in the traditional setting). However, instead of all sources generating traffic at the same rate, source node u generates traffic at a rate that is a function of its separation from destination w , i.e., the traffic matrix is given by $\lambda_{u,w}^{\text{UC}} = \psi(r_{u,w})$ for some function ψ . In particular, let us consider

$$\psi(r) \triangleq \rho(n) \times \begin{cases} r^\beta, & \text{if } r \geq 1 \\ 1, & \text{else} \end{cases}$$

for some fixed $\beta \in \mathbb{R}$ and some $\rho(n)$ depending only on n . The traditional setting corresponds to $\beta = 0$, in which case all n source–destination pairs communicate at uniform rate.

While such traffic is not balanced for small values of β , the results in Section IX-C, extending Theorem 1 to general traffic that is not balanced, can be used to establish the scaling of $\rho^*(n)$, the largest achievable value of $\rho(n)$, as

$$\rho^*(n) = \begin{cases} n^{1-(\bar{\alpha}+\beta)/2 \pm o(1)}, & \text{if } \beta \geq 2 - \bar{\alpha} \\ n^{\pm o(1)}, & \text{else} \end{cases}$$

with probability $1 - o(1)$. For $\beta = 0$, and noting that $2 - \bar{\alpha} \leq 0$, this recovers the results from [12] for random source–destination pairing with uniform rate. \diamond

Example 4: Sources with multiple destinations

All the example scenarios so far are concerned with traffic in which each node is source exactly once. Here, we consider more general traffic patterns. There are K classes of source nodes, for some fixed K . Each source node in class i has $\Theta(n^{\beta_i})$ destination nodes for some fixed $\beta_i \in [0, 1]$ and generates independent traffic at the same rate $\rho_i(n)$ for each of them (i.e., we still consider unicast traffic). Each of these destination nodes is chosen uniformly at random among the $n - 1$ other nodes. Every node randomly picks the class it belongs to. Noting that the resulting traffic matrix is balanced with high probability, Theorem 1 provides the following scaling of the rates achievable by different classes:

$$\rho_i^*(n) = n^{1 - \beta_i - \bar{\alpha}/2 \pm o(1)}$$

with probability $1 - o(1)$ for all i . In other words, for each source node time sharing between all K classes and then (within each class) between all its $\Theta(n^{\beta_i})$ destination nodes is order-optimal in this scenario. However, different sources are operating simultaneously. \diamond

Example 5: Broadcast

Consider a scenario with every node u in the network broadcasting an independent message to all other nodes at rate $\rho(n)\lambda_u$. In other words, we have a multicast traffic matrix of the form

$$\lambda_{u,W}^{\text{MC}} = \begin{cases} \rho(n)\lambda_u, & \text{if } W = V(n) \\ 0, & \text{else} \end{cases}$$

for some $\rho(n) > 0$. Applying the generalization in Section IX-C of Theorem 2 yields that $\rho^*(n)$, the largest achievable $\rho(n)$, satisfies

$$\rho^*(n) = n^{\pm o(1)} \frac{1}{\sum_{u \in V(n)} \lambda_u}$$

as $n \rightarrow \infty$. \diamond

V. COMMUNICATION SCHEMES

In this section, we provide a high-level description of the communication schemes used to prove achievability (i.e., the inner bound) in Theorems 1 and 2. In Section V-A, we present a communication scheme for general unicast traffic, in Section V-B, we show how this scheme can be adapted for general multicast traffic. Both schemes use as a building block a communication scheme introduced in prior work for a particular class of traffic, called *uniform permutation traffic*. In such uniform permutation traffic, each node in the network is source and destination exactly once, and all these source–destination pairs communicate at equal rate. For $\alpha \in (2, 3]$, the order-optimal scheme for such uniform permutation traffic (called *hierarchical relaying scheme* in the following) enables global cooperation in the network. For $\alpha > 3$, the order-optimal

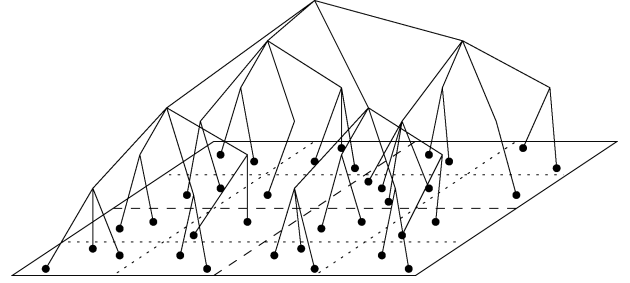


Fig. 3. Construction of the tree graph G . We consider the same nodes as in Fig. 1 with $L(n) = 2$. The leaves of G are the nodes $V(n)$ of the wireless network. They are always at level $\ell = L(n) + 1$ (i.e., 3 in this example). At level $0 \leq \ell \leq L(n)$ in G , there are 4^ℓ nodes. The tree structure is the one induced by the grid decomposition $\{V_{\ell,i}(n)\}_{\ell,i}$ as shown in Fig. 1. Level 0 contains the root node of G .

scheme is multihop routing. We recall these two schemes for uniform permutation traffic in Section V-C.

A. Communication Scheme for Unicast Traffic

In this section, we present a scheme to transmit general unicast traffic. This scheme has a tree structure that makes it convenient to work with. This tree structure is crucial in proving the compact approximation of the balanced unicast capacity region $\Lambda^{\text{BUC}}(n)$ in Theorem 1.

The communication scheme consists of three layers: A top or routing layer, a middle or cooperation layer, and a bottom or physical layer. The routing layer of this scheme treats the wireless network as a tree graph G and routes messages between sources and their destinations—dealing with heterogeneous traffic demands. The cooperation layer of this scheme provides this tree abstraction G to the top layer by appropriately distributing and concentrating traffic over the wireless network—choosing the level of cooperation in the network. The physical layer implements this distribution and concentration of messages in the wireless network—dealing with interference and noise.

Seen from the routing layer, the network consists of a noiseless capacitated graph G . This graph is a tree, whose leaf nodes are the nodes $V(n)$ in the wireless network. The internal nodes of G represent larger clusters of nodes (i.e., subsets of $V(n)$) in the wireless network. More precisely, each internal node in G represents a set $V_{\ell,i}(n)$ for $\ell \in \{1, \dots, L(n)\}$ and $i \in \{1, \dots, 4^\ell\}$. Consider two sets $V_{\ell,i}(n), V_{\ell+1,j}(n)$ and let ν, μ be the corresponding internal nodes in G . Then ν and μ are connected by an edge in G if $V_{\ell+1,j}(n) \subset V_{\ell,i}(n)$. Similarly, for $V_{L(n),i}(n)$ and corresponding internal node ν in G , a leaf node u in G is connected by an edge to ν if $u \in V_{L(n),i}(n)$ (recall that the leaf nodes of G are the nodes $V(n)$ in the wireless network). This construction is shown in Fig. 3. In the routing layer, messages are sent from each source to its destination by routing them over G . To send information along an edge of G , the routing layer calls upon the cooperation layer.

The cooperation layer implements the tree abstraction G . This is done by ensuring that whenever a message is located at a node in G , it is evenly distributed over the corresponding cluster in the wireless network, i.e., every node in the cluster has access to a distinct part of equal length of the message.

To send information from a child node to its parent in G (i.e., towards the root node of G), the message at the cluster in $V(n)$ represented by the child node is distributed evenly among all nodes in the bigger cluster in $V(n)$ represented by the parent node. More precisely, let ν be a child node of μ in G , and let $V_{\ell+1,i}(n), V_{\ell,j}(n)$ be the corresponding subsets of $V(n)$. Consider the cooperation layer being called by the routing layer to send a message from ν to its parent μ over G . In the wireless network, we assume each node in $V_{\ell+1,i}(n)$ has access to a distinct $1/|V_{\ell+1,i}(n)|$ fraction of the message to be sent. Each node in $V_{\ell+1,i}(n)$ splits its message part into four distinct parts of equal length. It keeps one part for itself and sends the other three parts to three nodes in $V_{\ell,j}(n) \setminus V_{\ell+1,i}(n)$. After each node in $V_{\ell+1,i}(n)$ has sent its message parts, each node in $V_{\ell,j}(n)$ now has access to a distinct $1/|V_{\ell,j}(n)|$ fraction of the message. To send information from a parent node to a child node in G (i.e., away from the root node of G), the message at the cluster in $V(n)$ represented by the parent node is concentrated on the cluster in $V(n)$ represented by the child node. More precisely, consider the same nodes ν and μ in G corresponding to $V_{\ell+1,i}(n)$ and $V_{\ell,j}(n)$ in $V(n)$. Consider the cooperation layer being called by the routing layer to send a message from μ to its child ν . In the wireless network, we assume each node in $V_{\ell,j}(n)$ has access to a distinct $1/|V_{\ell,j}(n)|$ fraction of the message to be sent. Each node in $V_{\ell,j}(n)$ sends its message part to a node in $V_{\ell+1,i}(n)$. After each node in $V_{\ell,j}(n)$ has sent its message part, each node in $V_{\ell+1,i}(n)$ now has access to a distinct $1/|V_{\ell+1,i}(n)|$ fraction of the message. To implement this distribution and concentration of messages, the cooperation layer calls upon the physical layer.

The physical layer performs the distribution and concentration of messages. Note that the traffic induced by the cooperation layer in the physical layer is very regular, and closely resembles a uniform permutation traffic (in which each node in the wireless network is source and destination once and all these source–destination pairs want to communicate at equal rate). Hence, we can use either cooperative communication (for $\alpha \in (2, 3]$) or multihop communication (for $\alpha > 3$) for the transmission of this traffic. See Section V-C for a detailed description of these two schemes. It is this operation in the physical layer that determines the edge capacities of the graph G as seen from the routing layer.

The operation of this three-layer architecture is illustrated in the following example.

Example 6: Consider a single source–destination pair (u, w) . The corresponding operation of the three-layer architecture is depicted in Fig. 4.

In the routing layer, the message is routed over the tree graph G between u and w (indicated in black in the figure). The middle plane in the figure shows the induced behavior from using the second edge along this path (indicated in solid black in the figure) in the cooperation layer. The bottom plane in the figure shows (part of) the corresponding actions induced in the physical layer. Let us now consider the specific operations of the three layers for the single message between u and w . Since G is a tree, there is a unique path between u and w , and the routing layer sends the message over the edges along this path. Consider

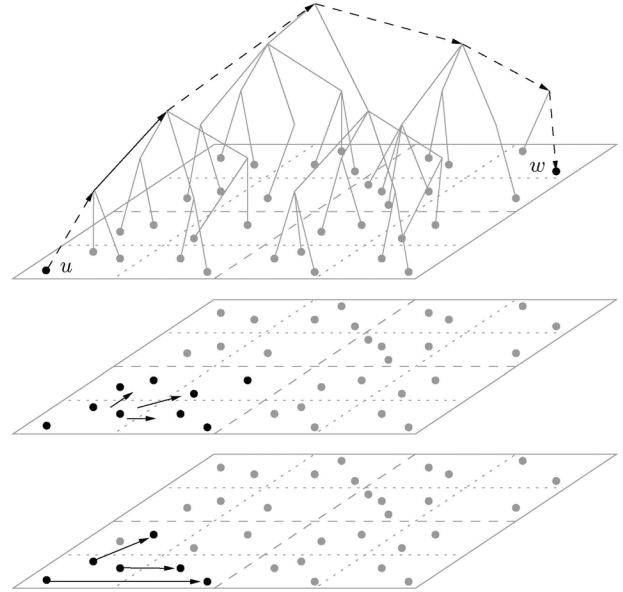


Fig. 4. Example operation of the three-layer architecture under unicast traffic. The three layers depicted are (from top to bottom in the figure) the routing layer, the cooperation layer, and the physical layer.

now the first such edge. Using this edge in the routing layer induces the following actions in the cooperation layer. The node u , having access to the entire message, splits that message into 3 distinct parts of equal length. It keeps one part, and sends the other two parts to the two other nodes in $V_{2,1}(n)$ (i.e., lower left square at level $\ell = 2$ in the hierarchy). In other words, after the message has traversed the edge between u and its parent node in the routing layer, all nodes in $V_{2,1}(n)$ in the cooperation layer have access to a distinct $1/3$ fraction of the original message. The edges in the routing layer leading up the tree (i.e., towards the root node) are implemented in the cooperation layer in a similar fashion by further distributing the message over the wireless network. By the time the message reaches the root node of G in the routing layer, the cooperation layer has distributed the message over the entire network and every node in $V(n)$ has access to a distinct $1/n$ fraction of the original message. Communication down the tree in the routing layer is implemented in the cooperation layer by concentrating messages over smaller regions in the wireless network. To physically perform this distribution and concentration of messages, the cooperation layer calls upon the physical layer, which uses either hierarchical relaying or multihop communication. \diamond

B. Communication Scheme for Multicast Traffic

Here, we show that the same communication scheme presented in the last section for general unicast traffic can also be used to transmit general multicast traffic. Again it is the tree structure of the scheme that is critically exploited in the proof of Theorem 2 providing an approximation for the balanced multicast capacity region $\Lambda^{\text{BMC}}(n)$.

We will use the same three-layer architecture as for unicast traffic presented in Section V-A. To accommodate multicast traffic, we only modify the operation of the top or routing layer; the lower layers operate as before.

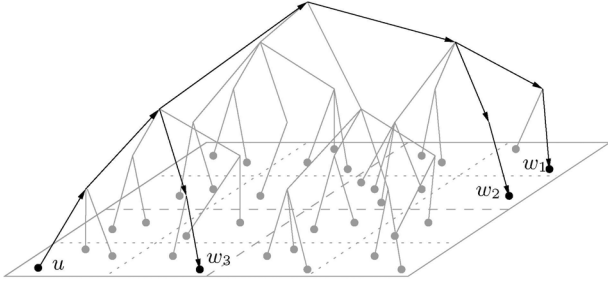


Fig. 5. Example operation of the routing layer in the three-layer architecture under multicast traffic.

We now outline how the routing layer needs to be adapted for the multicast case. Consider a multicast message that needs to be transmitted from a source node $u \in V(n)$ to its set of intended destinations $W \subset V(n)$. In the routing layer, we want to route this message from u to W over G . Since G is a tree, the routing part is simple. In fact, between u and every $w \in W$ there exists a unique path in G . Consider the union of all those paths. It is easy to see that this union is a subtree of G . Indeed, it is the smallest subtree of G that covers $\{u\} \cup W$. Traffic is optimally routed over G from u to W by sending it along the edges of this subtree.

The next example illustrates the operation of the routing layer under multicast traffic.

Example 7: Consider one source node u and the corresponding multicast group $W \triangleq \{w_1, w_2, w_3\}$ as shown in Fig. 5.

In the routing layer, we find the smallest subgraph $G(\{u\} \cup W)$ covering $\{u\} \cup W$ (indicated by black lines in Fig. 5). Messages are sent from the source to its destinations by routing them along this subgraph. In other words, $G(\{u\} \cup W)$ is the multicast tree along which the message is sent from u to W . The cooperation layer and physical layer operate in the same way as for unicast traffic (see Fig. 4 for an example). \diamond

C. Communication Schemes for Uniform Permutation Traffic

Here, we recall communication schemes for uniform permutation traffic on $A(n)$, i.e., each node is source and destination exactly once and all these n pairs communicate at uniform rate. As pointed out in Sections V-A and V-B, these communication schemes are used as building blocks in the communication architecture for general unicast and multicast traffic.

The structure of the optimal communication scheme depends drastically on the path-loss exponent α . For $\alpha \in (2, 3]$ (small path-loss exponent), cooperative communication on a global scale is necessary to achieve optimal performance. For $\alpha > 3$ (large path-loss exponent), local communication between neighboring nodes is sufficient, and traffic is routed in a multihop fashion from the source to the destination. We will refer to the order-optimal scheme for $\alpha \in (2, 3]$ as *hierarchical relaying scheme*, and to the order optimal scheme for $\alpha > 3$ as *multihop scheme*. For a uniform permutation traffic on $V(n)$, hierarchical relaying achieves a per-node rate of $n^{1-\alpha/2-o(1)}$; multihop communication achieves a per-node rate of $n^{-1/2-o(1)}$. By choosing the appropriate scheme (hierarchical relaying for $\alpha \in (2, 3]$, multihop for $\alpha > 3$), we can

thus achieve a per-node rate of $n^{1-\min\{3,\alpha\}/2-o(1)}$. We provide a short description of the hierarchical relaying scheme in the following. The details can be found in [13].

Consider n nodes placed independently and uniformly at random on $A(n)$. Divide $A(n)$ into

$$\frac{1}{n^{1+\log^{1/3}(n)}}$$

squarelets of equal size. Call a squarelet *dense*, if it contains a number of nodes proportional to its area. For each source–destination pair, choose such a dense squarelet as a *relay*, over which it will transmit information (see Fig. 6).

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

To actually enable this kind of cooperation at the relay squarelet, local communication within the relay squarelets is necessary. It can be shown that this local communication problem is actually the same as the original problem, but at a smaller scale. Indeed, we are now considering a square of size

$$n^{1-\frac{1}{1+\log^{1/3}(n)}}$$

with equal number of nodes (at least order wise). Hence, we can use the same scheme recursively to solve this subproblem. We terminate the recursion after $\log^{1/3}(n)$ iterations, at which point we use simple time-division multiple access (TDMA) to bootstrap the scheme.

Observe that at the final level of the scheme, we have divided $A(n)$ into

$$\left(\frac{1}{n^{1+\log^{1/3}(n)}} \right)^{\log^{1/3}(n)} = n^{\frac{1}{1+\log^{-1/3}(n)}}$$

squarelets. A sufficient condition for the scheme to succeed is that all these squarelets are dense (i.e., contain a number of nodes proportional to their area). However, much weaker conditions are sufficient, as well; see [13].

For any permutation traffic, the per-node rate achievable with this scheme is at least $n^{1-\alpha/2-o(1)}$ for any $\alpha > 2$ and under fast fading. Under slow fading the same per-node rate is achievable for all permutation traffic with probability at least

$$1 - \exp\left(-2^{\Omega(\log^{2/3}(n))}\right).$$

Moreover, when $\alpha \in (2, 3]$ and for uniform permutation traffic with a constant fraction of source–destination pairs at distance $\Theta(\sqrt{n})$ (as is the case with high probability if the permutation traffic is chosen at random), this is asymptotically the best uniformly achievable per-node rate.

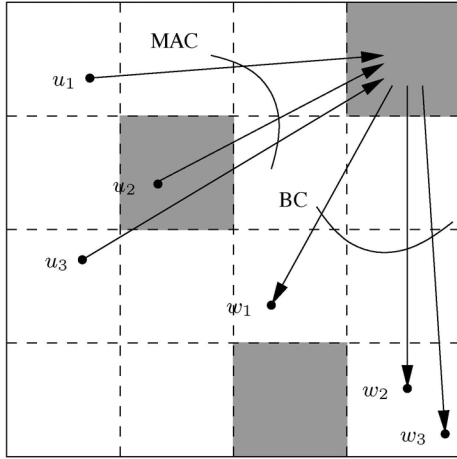


Fig. 6. Sketch of one level of the hierarchical relaying scheme. Here, $\{(u_i, w_i)\}_{i=1}^3$ are three source-destination pairs. Groups of source-destination pairs relay their traffic over dense squarelets (shaded), which contain a number of nodes proportional to their area. We time share between the different relay squarelets. Within each relay squarelet the scheme is used recursively to enable joint decoding and encoding at the relay.

VI. AUXILIARY LEMMAS

In this section, we provide auxiliary results, which will be used several times in the following. These results are grouped into three parts. In Section VI-A, we describe regularity properties exhibited with high probability by the random node placement. In Section VI-B, we provide auxiliary upper bounds on the performance of any scheme in terms of cut-set bounds. Finally, in Section VI-C, we describe auxiliary results on the performance of hierarchical relaying and multihop communication as described in Section V-C.

A. Regularity Lemmas

Here, we prove several regularity properties that are satisfied with high probability by a random node placement. Formally, define $\mathcal{V}(n)$ to be the collection of all node placements $V(n)$ that satisfy the following conditions:

$$\begin{aligned} r_{u,v} &> n^{-1} && \text{for all } u, v \in V(n), u \neq v \\ |V_{\ell,i}(n)| &\leq \log(n) && \text{for } \ell = \frac{1}{2} \log(n) \\ |V_{\ell,i}(n)| &\geq 1, && \text{for } \ell = \frac{1}{2} \log\left(\frac{n}{2 \log(n)}\right) \\ |V_{\ell,i}(n)| &\in [4^{-\ell-1}n, 4^{-\ell+1}n] && \text{for all } \ell \in \{1, \dots, L'(n)\} \end{aligned}$$

where

$$L'(n) \triangleq \frac{1}{2} \log(n) \left(1 - \frac{1}{2} \log^{-5/6}(n)\right)$$

and in each case $i \in \{1, \dots, 4^\ell\}$. The first condition is that the minimum distance between node pairs is not too small. The second condition is that all squares of area 1 contain at most $\log(n)$ nodes. The third condition is that all squares of area $2 \log(n)$ contain at least one node. The fourth condition is that

all squares up to level $\frac{1}{2} \log(n)(1 - \frac{1}{2} \log^{-5/6}(n))$ contain a number of nodes proportional to their area. Note that, since

$$\begin{aligned} L(n) &= \frac{1}{2} \log(n)(1 - \log^{-1/2}(n)) \\ &\leq \frac{1}{2} \log(n) \left(1 - \frac{1}{2} \log^{-5/6}(n)\right) \end{aligned}$$

this holds in particular for nodes up to level $L(n)$. The goal of this section is to prove that

$$\mathbb{P}(V(n) \in \mathcal{V}(n)) = 1 - o(1)$$

as $n \rightarrow \infty$.

The first lemma shows that the minimum distance in a random node placement is at least n^{-1} with high probability.

Lemma 3:

$$\mathbb{P}\left(\min_{u \in V(n), v \in V(n) \setminus \{u\}} r_{u,v} > n^{-1}\right) = 1 - o(1)$$

as $n \rightarrow \infty$.

Proof: For $u, v \in V$, let

$$B_{u,v} \triangleq \{r_{u,v} \leq r\}$$

for some r (depending only on n). Fix a node $u \in V$, then for $v \neq u$

$$\mathbb{P}(B_{u,v}|u) \leq \frac{r^2 \pi}{n}$$

(the inequality being due to boundary effects). Moreover, the events $\{B_{u,v}\}_{v \in V \setminus \{u\}}$ are independent conditioned on u , and thus

$$\begin{aligned} \mathbb{P}\left(\bigcap_{v \in V \setminus \{u\}} B_{u,v}^c | u\right) &= \prod_{v \in V \setminus \{u\}} \mathbb{P}(B_{u,v}^c | u) \\ &\geq \left(1 - \frac{r^2 \pi}{n}\right)^n. \end{aligned}$$

From this

$$\begin{aligned} &\mathbb{P}\left(\min_{u \in V, v \in V \setminus \{u\}} r_{u,v} \leq r\right) \\ &= \mathbb{P}\left(\bigcup_{u \in V, v \in V \setminus \{u\}} B_{u,v}\right) \\ &\leq \sum_{u \in V} \mathbb{P}\left(\bigcup_{v \in V \setminus \{u\}} B_{u,v}\right) \\ &= \sum_{u \in V} (1 - \mathbb{P}\left(\bigcap_{v \in V \setminus \{u\}} B_{u,v}^c\right)) \\ &= \sum_{u \in V} (1 - \mathbb{E}\left(\mathbb{P}\left(\bigcap_{v \in V \setminus \{u\}} B_{u,v}^c | u\right)\right)) \\ &\leq \sum_{u \in V} \left(1 - \left(1 - \frac{r^2 \pi}{n}\right)^n\right) \\ &= n \left(1 - \left(1 - \frac{r^2 \pi}{n}\right)^n\right). \end{aligned}$$

Assuming $r < \sqrt{n/\pi}$, we have

$$n \left(1 - \left(1 - \frac{r^2 \pi}{n}\right)^n\right) \leq nr^2 \pi$$

and, hence

$$\mathbb{P} \left(\min_{u \in V, v \in V \setminus \{u\}} r_{u,v} \leq r \right) \leq nr^2\pi$$

which converges to zero for $r = n^{-1}$. \square

The next lemma asserts that if $\tilde{L}(n)$ is not too large then all squares $\{V_{\ell,i}(n)\}_{\ell,i}$ for $\ell \in \{1, \dots, \tilde{L}(n)\}$ and $i \in \{1, \dots, 4^\ell\}$ in the grid decomposition of $V(n)$ contain a number of nodes that is proportional to their area.

Lemma 4: If $\tilde{L}(n)$ satisfies

$$\lim_{n \rightarrow \infty} \frac{\tilde{L}(n)}{4^{-\tilde{L}(n)}n} = 0$$

then

$$\mathbb{P} \left(\bigcap_{\ell=1}^{\tilde{L}(n)} \bigcap_{i=1}^{4^\ell} \{|V_{\ell,i}(n)| \in [4^{-\ell-1}n, 4^{-\ell+1}n]\} \right) = 1 - o(1)$$

as $n \rightarrow \infty$. In particular, this holds for

$$\tilde{L}(n) = \frac{1}{2} \log(n) \left(1 - \frac{1}{2} \log^{-5/6}(n) \right)$$

and for $\tilde{L}(n) = L(n)$.

Proof: Let B_u be the event that node u lies in $A_{\ell,i}$ for fixed ℓ, i . Note that

$$\sum_{u \in V} \mathbb{1}_{B_u} = |V_{\ell,i}|$$

by definition, and that

$$\mathbb{P}(B_u) = 4^{-\ell}.$$

Hence, using the Chernoff bound

$$\begin{aligned} & \mathbb{P}(|V_{\ell,i}| \notin [4^{-\ell-1}n, 4^{-\ell+1}n]) \\ &= \mathbb{P} \left(\sum_{u \in V} \mathbb{1}_{B_u} \notin [4^{-\ell-1}n, 4^{-\ell+1}n] \right) \\ &\leq \exp(-K4^{-\ell}n) \end{aligned}$$

for some positive constant K , and we obtain, for $\ell = \tilde{L}(n)$

$$\begin{aligned} & \mathbb{P} \left(\bigcap_{i=1}^{4^{\tilde{L}(n)}} \{|V_{\tilde{L}(n),i}| \in [4^{-\tilde{L}(n)-1}n, 4^{-\tilde{L}(n)+1}n]\} \right) \\ &\geq 1 - \sum_{i=1}^{4^{\tilde{L}(n)}} \mathbb{P} \left(|V_{\tilde{L}(n),i}| \notin [4^{-\tilde{L}(n)-1}n, 4^{-\tilde{L}(n)+1}n] \right) \\ &\geq 1 - 4^{\tilde{L}(n)} \exp(-K4^{-\tilde{L}(n)}n) \\ &\geq 1 - \exp(\tilde{K}\tilde{L}(n) - K4^{-\tilde{L}(n)}n) \end{aligned} \quad (6)$$

for some positive constant \tilde{K} . By assumption

$$\lim_{n \rightarrow \infty} \frac{\tilde{L}(n)}{4^{-\tilde{L}(n)}n} = 0$$

and, hence

$$\mathbb{P} \left(\bigcap_{i=1}^{4^{\tilde{L}(n)}} \{|V_{\tilde{L}(n),i}| \in [4^{-\tilde{L}(n)-1}n, 4^{-\tilde{L}(n)+1}n]\} \right) \geq 1 - o(1)$$

as $n \rightarrow \infty$. Since the $\{A_{\ell,i}\}_{\ell,i}$ are nested as a function of ℓ , we have

$$\begin{aligned} & \bigcap_{\ell=1}^{\tilde{L}(n)} \bigcap_{i=1}^{4^\ell} \{|V_{\ell,i}| \in [4^{-\ell-1}n, 4^{-\ell+1}n]\} \\ &= \bigcap_{i=1}^{4^{\tilde{L}(n)}} \{|V_{\tilde{L}(n),i}| \in [4^{-\tilde{L}(n)-1}n, 4^{-\tilde{L}(n)+1}n]\} \end{aligned}$$

which, combined with (6), proves the first part of the lemma.

For the second part, note that for

$$\tilde{L}(n) = \frac{1}{2} \log(n) \left(1 - \frac{1}{2} \log^{-5/6}(n) \right)$$

we have

$$\begin{aligned} \frac{\tilde{L}(n)}{4^{-\tilde{L}(n)}n} &= \frac{\frac{1}{2} \log(n) \left(1 - \frac{1}{2} \log^{-5/6}(n) \right)}{2^{\frac{1}{2} \log^{1/6}(n)}} \\ &\leq \frac{\log(n)}{2^{\frac{1}{2} \log^{1/6}(n)}} \\ &= 2^{\log \log(n) - \frac{1}{2} \log^{1/6}(n)} \rightarrow 0 \end{aligned}$$

and, hence, the lemma is valid in this case. The same holds for $\tilde{L}(n) = L(n)$ since

$$L(n) \leq \frac{1}{2} \log(n) \left(1 - \frac{1}{2} \log^{-5/6}(n) \right).$$

\square

We are now ready to prove that a random node placement $V(n)$ is in $\mathcal{V}(n)$ with high probability as $n \rightarrow \infty$ (i.e., is fairly “regular” with high probability).

Lemma 5:

$$\mathbb{P}(V(n) \in \mathcal{V}(n)) = 1 - o(1)$$

as $n \rightarrow \infty$.

Proof: The first condition

$$r_{u,v} > n^{-1} \quad \text{for all } u, v \in V, u \neq v$$

holds with probability $1 - o(1)$ by Lemma 3. The second and third conditions

$$\begin{aligned} |V_{\ell,i}| &\leq \log(n), \quad \text{for } \ell = \frac{1}{2} \log(n) \\ |V_{\ell,i}| &\geq 1, \quad \text{for } \ell = \frac{1}{2} \log \left(\frac{n}{2 \log(n)} \right) \end{aligned}$$

are shown in [12, Lemma 5.1] to hold with probability $1 - o(1)$. The fourth condition

$$|V_{\ell,i}| \in [4^{-\ell-1}n, 4^{-\ell+1}n] \quad \text{for all } \ell \in \{1, \dots, L'(n)\}$$

holds with probability $1 - o(1)$ by Lemma 4. Together, this proves the result. \square

B. Converse Lemmas

Here, we prove several auxiliary converse results. The first lemma bounds the maximal achievable sum rate for every individual node (i.e., the total traffic for which a fixed node is either source or destination).

Lemma 6: Under either fast or slow fading, for any $\alpha > 2$, there exists $b(n) = O(\log(n))$ such that for all $V(n) \in \mathcal{V}(n)$, $\lambda^{\text{UC}} \in \Lambda^{\text{UC}}(n)$, $u \in V(n)$

$$\sum_{w \in V(n) \setminus \{u\}} \lambda_{u,w}^{\text{UC}} \leq b(n) \quad (7)$$

$$\sum_{w \in V(n) \setminus \{u\}} \lambda_{w,u}^{\text{UC}} \leq b(n). \quad (8)$$

Proof: The argument follows the one in [12, Theorem 3.1]. Denote by $C(S_1, S_2)$ the multiple-input multiple-output (MIMO) capacity between nodes in S_1 and nodes in S_2 , for $S_1, S_2 \subset V$. Consider first (7). By the cut-set bound [27, Theorem 14.10.1]

$$\sum_{w \neq u} \lambda_{u,w}^{\text{UC}} \leq C(\{u\}, \{u\}^c).$$

$C(\{u\}, \{u\}^c)$ is the capacity between u and the nodes in $\{u\}^c$, i.e.,

$$\begin{aligned} C(\{u\}, \{u\}^c) &= \log \left(1 + \sum_{v \neq u} |h_{u,v}|^2 \right) \\ &\leq \log(1 + (n-1)n^\alpha) \\ &\leq K \log(n) \end{aligned}$$

with

$$K \triangleq 2 + \alpha$$

and where for the first inequality we have used that since $V \in \mathcal{V}$, we have $r_{u,v} \geq n^{-1}$ for all $v \neq u$.

Similarly, for (8)

$$\sum_{w \neq u} \lambda_{w,u}^{\text{UC}} \leq C(\{u\}^c, \{u\})$$

and

$$\begin{aligned} C(\{u\}^c, \{u\}) &\leq \log \left(1 + (n-1) \sum_{v \neq u} |h_{v,u}|^2 \right) \\ &\leq \log(1 + (n-1)^2 n^\alpha) \\ &\leq K \log(n). \end{aligned}$$

\square

The next lemma bounds the maximal achievable sum rate across the boundary out of the subsquares $V_{\ell,i}(n)$ for $\ell \in \{1, \dots, L(n)\}$, and $i \in \{1, \dots, 4^\ell\}$.

Lemma 7: Under either fast or slow fading, for any $\alpha > 2$, there exists $b(n) = O(\log^6(n))$ such that for all $V(n) \in \mathcal{V}(n)$, $\lambda^{\text{UC}} \in \Lambda^{\text{UC}}(n)$, $\ell \in \{1, \dots, L(n)\}$, and $i \in \{1, \dots, 4^\ell\}$, we have

$$\sum_{u \in V_{\ell,i}(n)} \sum_{w \notin V_{\ell,i}(n)} \lambda_{u,w}^{\text{UC}} \leq b(n)(4^{-\ell}n)^{2-\min\{3,\alpha\}/2}.$$

Proof: As before, denote by $C(S_1, S_2)$ the MIMO capacity between nodes in S_1 and nodes in S_2 . By the cut-set bound [27, Theorem 14.10.1]

$$\sum_{u \in V_{\ell,i}} \sum_{w \notin V_{\ell,i}} \lambda_{u,w}^{\text{UC}} \leq C(V_{\ell,i}, V_{\ell,i}^c). \quad (9)$$

Let

$$\mathbf{H}_{S_1, S_2} \triangleq [h_{u,v}]_{u \in S_1, v \in S_2}$$

be the matrix of channel gains between the nodes in S_1 and S_2 . Under fast fading

$$\begin{aligned} C(S_1, S_2) &\triangleq \max_{\mathbf{Q}(\mathbf{H}) \geq 0: \mathbb{E}(q_{u,u}) \leq 1 \forall u \in S_1} \mathbb{E} \left(\log \det \left(\mathbf{I} + \mathbf{H}_{S_1, S_2}^\dagger \mathbf{Q}(\mathbf{H}) \mathbf{H}_{S_1, S_2} \right) \right) \end{aligned}$$

and under slow fading

$$C(S_1, S_2) \triangleq \max_{\mathbf{Q} \geq 0: q_{u,u} \leq 1 \forall u \in S_1} \log \det \left(\mathbf{I} + \mathbf{H}_{S_1, S_2}^\dagger \mathbf{Q} \mathbf{H}_{S_1, S_2} \right).$$

Denote by $\partial(V_{\ell,i}^c)$ the nodes in $V_{\ell,i}^c$ that are within distance one of the boundary between $A_{\ell,i}^c$ and $A_{\ell,i}$. Using the generalized Hadamard inequality yields that under either fast or slow fading

$$C(V_{\ell,i}, V_{\ell,i}^c) \leq C(V_{\ell,i}, \partial(V_{\ell,i}^c)) + C(V_{\ell,i}, V_{\ell,i}^c \setminus \partial(V_{\ell,i}^c)). \quad (10)$$

We start by analyzing the first term in the sum in (10). Applying Hadamard's inequality again yields

$$C(V_{\ell,i}, \partial(V_{\ell,i}^c)) \leq \sum_{v \in \partial(V_{\ell,i}^c)} C(V_{\ell,i}, \{v\}).$$

Since $V \in \mathcal{V}$, we have

$$|\partial(V_{\ell,i}^c)| \leq 5 \log(n)(4^{-\ell}n)^{1/2}.$$

By the same analysis as in Lemma 6, we obtain

$$C(V_{\ell,i}, \{v\}) \leq C(\{v\}^c, \{v\}) \leq \frac{K}{5} \log(n)$$

for some constant K (independent of v). Therefore

$$\begin{aligned} C(V_{\ell,i}, \partial(V_{\ell,i}^c)) &\leq 5 \log(n)(4^{-\ell}n)^{1/2} \frac{K}{5} \log(n) \\ &= K \log^2(n)(4^{-\ell}n)^{1/2}. \end{aligned} \quad (11)$$

We now analyze the second term in the sum in (10). The arguments of [13, Lemma 12] (building on [12, Theorem 5.2]) show that under either fast or slow fading there exists $\tilde{K} > 0$ such that for any $V \in \mathcal{V}, \ell \in \{0, \dots, L(n)\}$

$$C(V_{\ell,i}, V_{\ell,i}^c \setminus \partial(V_{\ell,i}^c)) \leq \tilde{K} \log^3(n) \sum_{u \in V_{\ell,i}} \sum_{v \in V_{\ell,i}^c \setminus \partial(V_{\ell,i}^c)} r_{u,v}^{-\alpha}. \quad (12)$$

Moreover, using the same arguments as in [12, Theorem 5.2] shows that there exists a constant $K' > 0$ such that for adjacent squares (i.e., sharing a side) $A_{\ell,i}, A_{\ell,j}$

$$\sum_{u \in V_{\ell,i}} \sum_{v \in V_{\ell,j} \setminus \partial(V_{\ell,i}^c)} r_{u,v}^{-\alpha} \leq K' \log^3(n) (4^{-\ell} n)^{2-\min\{3,\alpha\}/2}. \quad (13)$$

Consider now two diagonal squares (i.e., sharing a corner point) $A_{\ell,i}, A_{\ell,j}$. Using a similar argument and suitably redefining K' shows that (13) holds for diagonal squares as well.

Using this, we now compute the summation in (12). Consider “rings” of squares around $A_{\ell,i}$. The first such “ring” contains the (at most) 8 squares neighboring $A_{\ell,i}$. The next “ring” contains at most 16 squares. In general, “ring” k contains at most $8k$ squares. Let

$$\{A_{\ell,j}\}_{j \in I_k}$$

be the squares in “ring” k . Then

$$\begin{aligned} & \sum_{u \in V_{\ell,i}} \sum_{v \in V_{\ell,i}^c \setminus \partial(V_{\ell,i}^c)} r_{u,v}^{-\alpha} \\ &= \sum_{k \geq 1} \sum_{j \in I_k} \sum_{u \in V_{\ell,i}} \sum_{v \in V_{\ell,j} \setminus \partial(V_{\ell,i}^c)} r_{u,v}^{-\alpha}. \end{aligned} \quad (14)$$

By (13)

$$\sum_{j \in I_1} \sum_{u \in V_{\ell,i}} \sum_{v \in V_{\ell,j} \setminus \partial(V_{\ell,i}^c)} r_{u,v}^{-\alpha} \leq 8K' \log^3(n) (4^{-\ell} n)^{2-\min\{3,\alpha\}/2}. \quad (15)$$

Now note that for $k > 1$ and $j \in I_k$, nodes $u \in V_{\ell,i}$ and $v \in V_{\ell,j}$ are at least at distance $r_{u,v} \geq (k-1)(2^{-\ell} \sqrt{n})$. Moreover, since $V \in \mathcal{V}$, each $\{V_{\ell,j}\}_{j \in I_k}$ has cardinality at most $4^{-\ell+1} n$. Thus

$$\begin{aligned} & \sum_{k > 1} \sum_{j \in I_k} \sum_{u \in V_{\ell,i}} \sum_{v \in V_{\ell,j} \setminus \partial(V_{\ell,i}^c)} r_{u,v}^{-\alpha} \\ & \leq \sum_{k > 1} 8k (4^{-\ell+1} n)^2 ((k-1)(2^{-\ell} \sqrt{n}))^{-\alpha} \\ & = 128 (4^{-\ell} n)^{2-\alpha/2} \sum_{k > 1} k (k-1)^{-\alpha} \\ & = K'' (4^{-\ell} n)^{2-\alpha/2} \end{aligned} \quad (16)$$

for some $K'' > 0$, and where we have used that $\alpha > 2$. Substituting (15) and (16) into (14) yields

$$\begin{aligned} & \sum_{u \in V_{\ell,i}} \sum_{v \in V_{\ell,i}^c \setminus \partial(V_{\ell,i}^c)} r_{u,v}^{-\alpha} \\ & \leq 8K' \log^3(n) (4^{-\ell} n)^{2-\min\{3,\alpha\}/2} + K'' (4^{-\ell} n)^{2-\alpha/2} \end{aligned}$$

and, hence, by (12)

$$\begin{aligned} C(V_{\ell,i}, V_{\ell,i}^c \setminus \partial(V_{\ell,i}^c)) & \leq \tilde{K} \log^3(n) \left(8K' \log^3(n) (4^{-\ell} n)^{2-\min\{3,\alpha\}/2} \right. \\ & \quad \left. + K'' (4^{-\ell} n)^{2-\alpha/2} \right). \end{aligned} \quad (17)$$

Combining (9), (10), (11), and (17) shows that

$$\sum_{u \in V_{\ell,i}} \sum_{v \notin V_{\ell,i}} \lambda_{u,v}^{\text{UC}} \leq b(n) (4^{-\ell} n)^{2-\min\{3,\alpha\}/2}.$$

for every $\ell \in \{1, \dots, L(n)\}, i \in \{1, \dots, 4^\ell\}$, and under either fast or slow fading. \square

The following lemma bounds the maximal achievable sum rate across the boundary into the subsquares $V_{\ell,i}(n)$ for $\ell \in \{1, \dots, L(n)\}$, and $i \in \{1, \dots, 4^\ell\}$. Note that this lemma is only valid for $\alpha > 5$.

Lemma 8: Under either fast or slow fading, for any $\alpha > 5$, there exists $b(n) = O(\log^3(n))$ such that for all $V(n) \in \mathcal{V}(n)$, $\lambda^{\text{UC}} \in \Lambda^{\text{UC}}(n)$, $\ell \in \{1, \dots, L(n)\}$, and $i \in \{1, \dots, 4^\ell\}$, we have

$$\sum_{u \notin V_{\ell,i}(n)} \sum_{w \in V_{\ell,i}(n)} \lambda_{u,w}^{\text{UC}} \leq b(n) (4^{-\ell} n)^{1/2}.$$

Proof: By the cut-set bound [27, Theorem 14.10.1]

$$\sum_{u \notin V_{\ell,i}} \sum_{w \in V_{\ell,i}} \lambda_{u,w}^{\text{UC}} \leq C(V_{\ell,i}^c, V_{\ell,i}). \quad (18)$$

Denote by $\partial V_{\ell,i}$ the nodes in $V_{\ell,i}$ that are within distance one of the boundary between $A_{\ell,i}^c$ and $A_{\ell,i}$. Applying the generalized Hadamard inequality as in Lemma 7, we have under either fast or slow fading

$$\begin{aligned} C(V_{\ell,i}^c, V_{\ell,i}) & \leq C(V_{\ell,i}^c, \partial V_{\ell,i}) + C(V_{\ell,i}^c, V_{\ell,i} \setminus \partial V_{\ell,i}) \\ & \leq K \log^2(n) (4^{-\ell} n)^{1/2} + C(V_{\ell,i}^c, V_{\ell,i} \setminus \partial V_{\ell,i}) \end{aligned} \quad (19)$$

for some positive constant K .

For the second term in (19), we have by slightly adapting the upper bound from Theorem 2.1 in [3]

$$C(V_{\ell,i}^c, V_{\ell,i} \setminus \partial V_{\ell,i}) \leq \sum_{v \in V_{\ell,i}^c \setminus \partial V_{\ell,i}} \left(\sum_{u \in V_{\ell,i}^c} r_{u,v}^{-\alpha/2} \right)^2.$$

Now, consider $v \in V_{\ell,i} \setminus \partial V_{\ell,i}$ and let d_v be the distance of v from the closest node in $V_{\ell,i}^c$. Using $V \in \mathcal{V}$ and $\alpha > 5$

$$\sum_{u \in V_{\ell,i}^c} r_{u,v}^{-\alpha/2} \leq \tilde{K} \log(n) d_v^{2-\alpha/2}$$

for some positive constant \tilde{K} , and, hence

$$\begin{aligned} C(V_{\ell,i}^c, V_{\ell,i} \setminus \partial V_{\ell,i}) &\leq \sum_{v \in V_{\ell,i} \setminus \partial V_{\ell,i}} \tilde{K}^2 \log^2(n) d_v^{4-\alpha} \\ &\leq K' \log^3(n) (4^{-\ell} n)^{1/2} \end{aligned}$$

for some positive constant K' . Combined with (19) and (18), this proves Lemma 8. \square

C. Achievability Lemmas

In this section, we prove auxiliary achievability results. Recall that a permutation traffic is a traffic pattern in which each node is source and destination exactly once. Call the corresponding source–destination pairing $\Pi \subset V(n) \times V(n)$ a *permutation pairing*. The lemma below analyzes the performance achievable with either hierarchical relaying (for $\alpha \in (2, 3]$) or multihop communication (for $\alpha > 3$) applied simultaneously to transmit permutation traffic in several disjoint regions in the network. See Section V-C for a description of these communication schemes.

Lemma 9: Under fast fading, for any $\alpha > 2$, there exists $b(n) \geq n^{-o(1)}$ such that for all $V(n) \in \mathcal{V}(n)$, $\ell \in \{0, \dots, L(n)\}$, $i \in \{1, \dots, 4^\ell\}$, and permutation source–destination pairing Π_i on $V_{\ell,i}(n)$, there exists $\lambda_{u,w}^{\text{UC}} \in \Lambda^{\text{UC}}(n)$ such that

$$\min_{i \in \{1, \dots, 4^\ell\}} \min_{(u,w) \in \Pi_i} \lambda_{u,w}^{\text{UC}} \geq b(n) (4^{-\ell} n)^{1-\min\{3,\alpha\}/2}.$$

The same statement holds with probability $1 - o(1)$ as $n \rightarrow \infty$ in the slow fading case.

Consider the source–destination pairing $\Pi \triangleq \cup_i \Pi_i$ with $\{\Pi_i\}_i$ as in Lemma 9. This is a permutation pairing, since each Π_i is a permutation pairing on $V_{\ell,i}(n)$ and since the $\{V_{\ell,i}(n)\}_i$ are disjoint. Lemma 9 states that every source–destination pair in Π can communicate at a per-node rate of at least $n^{-o(1)} (4^{-\ell} n)^{1-\min\{3,\alpha\}/2}$. Note that, due to the locality of the traffic pattern, this can be considerably better than the $n^{1-\min\{3,\alpha\}/2-o(1)}$ per-node rate achieved by standard hierarchical relaying or multihop communication.

Proof: We shall use either hierarchical relaying (for $\alpha \in (2, 3]$) or multihop (for $\alpha > 3$) to communicate within each square $V_{\ell,i}$. We operate every fourth of the $V_{\ell,i}$ simultaneously, and show that the added interference due to this spatial re-use results only in a constant factor loss in rate.

Consider first $\alpha \in (2, 3]$ and fast fading. The squares $A_{\ell,i}$ at level ℓ have an area of

$$n_\ell \triangleq 4^{-\ell} n.$$

In order to be able to use hierarchical relaying within each of the $\{A_{\ell,i}\}_i$, it is sufficient to show that we can partition each $A_{\ell,i}$ into

$$\frac{1}{1+\log^{-1/3}(n_\ell)}$$

squarelets, each of which contains a number of nodes proportional to the area (see Section V-C). In other words, we partition A into squarelets of size

$$\begin{aligned} n_\ell \frac{1 - \frac{1}{1+\log^{-1/3}(n_\ell)}}{1+\log^{-1/3}(n_\ell)} &\geq n_{\frac{L(n)}{1+\log^{-1/3}(n)}} \\ &\geq n_{\frac{L(n)}{2}}^{\frac{1}{2} \log^{-1/3}(n)} \\ &= 2^{\frac{1}{2} \log^{1/6}(n)} \\ &= n 4^{-\frac{1}{2} \log(n)} (1 - \frac{1}{2} \log^{-5/6}(n)) \end{aligned}$$

where we have assumed, without loss of generality, that $n \geq 2$. Since $V \in \mathcal{V}$, all these squarelets contain a number of nodes proportional to their area, and, hence, this shows that all

$$\{A_{i,\ell}\}_{\ell \in \{0, \dots, L(n)\}, i \in \{1, \dots, 4^\ell\}}$$

are simultaneously regular enough for hierarchical relaying to be successful under fast fading. This achieves a per-node rate of

$$\lambda_{u,w}^{\text{UC}} \geq n^{-o(1)} (4^{-\ell} n)^{1-\alpha/2} \quad (20)$$

for any $(u, w) \in \Pi_i$ (see Section V-C, or [13, Theorem 1]).

We now show that (20) holds with high probability also under slow fading. For $V \in \mathcal{V}$ hierarchical relaying is successful under slow fading for all permutation traffic on V with probability at least

$$1 - \exp\left(-2^K \log^{2/3}(n)\right)$$

for some positive constant K (see again Section V-C). Hence, hierarchical relaying is successful for all permutation traffic on $V_{\ell,i}$ with probability at least

$$\begin{aligned} 1 - \exp\left(-2^K \log^{2/3}(4^{-\ell} n)\right) \\ \geq 1 - \exp\left(-2^K \log^{2/3}(4^{-L(n)} n)\right) \\ = 1 - \exp\left(-2^K \log^{1/3}(n)\right) \end{aligned}$$

and, hence, hierarchical relaying is successful under slow fading for all $\ell \in \{1, \dots, L(n)\}$ and all permutation traffic on every $\{V_{\ell,i}\}_{i=1}^{4^\ell}$ with probability at least

$$\begin{aligned} 1 - L(n) 4^{L(n)} \exp\left(-2^K \log^{1/3}(n)\right) \\ \geq 1 - n^2 \exp\left(-2^K \log^{1/3}(n)\right) \\ \geq 1 - o(1) \end{aligned}$$

as $n \rightarrow \infty$.

We now argue that the additional interference from spatial re-use results only in a constant loss in rate. This follows from the same arguments as in the proof of [13, Theorem 1] (with the appropriate modifications for slow fading as described there). Intuitively, this is the case since the interference from a square at distance r is attenuated by a factor $r^{-\alpha}$, which, since $\alpha > 2$, is summable. Hence, the combined interference has power on the order of the receiver noise, resulting in only a constant factor loss in rate.

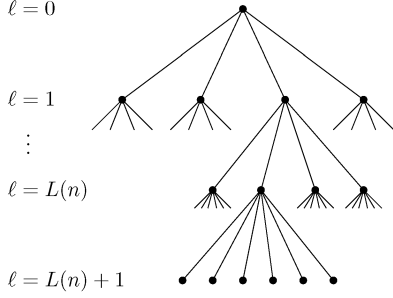


Fig. 7. Communication graph G constructed in the proof of Theorem 1. Nodes on levels $\ell \in \{0, \dots, L(n) - 1\}$ have four children each, nodes on level $\ell = L(n)$ have $\Theta(n^{\log^{-1/2}(n)})$ children each. The total number of leaf nodes is n , one representing each node in the wireless network $V(n)$. An internal node in G at level $\ell \in \{0, \dots, L(n)\}$ represents the collection of nodes in $V_{\ell,i}(n)$ for some i .

For $\alpha > 3$, the argument is similar—instead of hierarchical relaying we now use multihop communication. For $V \in \mathcal{V}$ and under either fast or slow fading, this achieves a per-node rate of

$$\lambda_{u,w}^{\text{UC}} \geq n^{-o(1)}(4^{-\ell}n)^{-1/2} \quad (21)$$

for any $(u, w) \in \Pi_i$. Combining (20) and (21) yields the lemma. \square

VII. PROOF OF THEOREM 1

The proof of Theorem 1 relies on the construction of a capacitated (noiseless, wireline) graph G and linking its performance under routing to the performance of the wireless network. This graph $G = (V_G, E_G)$ is constructed as follows. G is a full tree (i.e., all its leaf nodes are on the same level). G has n leaves, each of them representing an element of $V(n)$. To simplify notation, we assume that $V(n) \subset V_G$, so that the leaves of G are exactly the elements of $V(n) \subset V_G$. Whenever the distinction is relevant, we use u, v for nodes in $V(n) \subset V_G$ and μ, ν for nodes in $V_G \setminus V(n)$ in the following. The internal nodes of G correspond to $V_{\ell,i}(n)$ for all $\ell \in \{0, \dots, L(n)\}$, $i \in \{1, \dots, 4^\ell\}$, with hierarchy induced by the one on $A(n)$. In particular, let μ and ν be internal nodes in V_G and let $V_{\ell,i}(n)$ and $V_{\ell+1,j}(n)$ be the corresponding subsets of $V(n)$. Then ν is a child node of μ if $V_{\ell+1,j}(n) \subset V_{\ell,i}(n)$.

In the following, we will assume $V(n) \in \mathcal{V}(n)$, which holds with probability $1 - o(1)$ as $n \rightarrow \infty$ by Lemma 5. With this assumption, nodes in V_G at level $\ell < L(n)$ have 4 children each, nodes in V_G at level $\ell = L(n)$ have between $4^{-L(n)-1}n$ and $4^{-L(n)+1}n$ children, and nodes in V_G at level $\ell = L(n) + 1$ are the leaves of the tree (see Fig. 7 above and Fig. 3 in Section V-A).

For $\mu \in V_G$, denote by $\mathcal{L}(\mu)$ the leaf nodes of the subtree of G rooted at μ . Note that, by construction of the graph G , $\mathcal{L}(\mu) = V_{\ell,i}(n)$ for some ℓ and i . To understand the relation between V_G and $V(n)$, we define the *representative* $\mathcal{R} : V_G \rightarrow 2^{V(n)}$ of μ as follows. For a leaf node $u \in V(n) \subset V_G$ of G , let

$$\mathcal{R}(u) \triangleq \{u\}.$$

For $\mu \in V_G$ at level $L(n)$, choose $\mathcal{R}(\mu) \subset \mathcal{L}(\mu) \subset V(n)$ such that

$$|\mathcal{R}(\mu)| = 4^{-L(n)-1}n.$$

This is possible since $V(n) \in \mathcal{V}(n)$ by assumption. Finally, for $\mu \in V_G$ at level $\ell < L(n)$, and with children $\{\nu_j\}_{j=1}^4$, let

$$\mathcal{R}(\mu) \triangleq \bigcup_{j=1}^4 \mathcal{R}(\nu_j).$$

We now define an edge capacity $c_{\mu,\nu}$ for each edge $(\mu, \nu) \in E_G$. If μ is a leaf of G and ν its parent, set

$$c_{\mu,\nu} = c_{\nu,\mu} \triangleq 1. \quad (22)$$

If μ is an internal node at level ℓ in G and ν its parent, then set

$$c_{\mu,\nu} = c_{\nu,\mu} \triangleq (4^{-\ell}n)^{2-\min\{3,\alpha\}/2}. \quad (23)$$

Having chosen edge capacities on G , we can now define the set $\Lambda_G^{\text{UC}}(n) \subset \mathbb{R}_+^{n \times n}$ of feasible unicast traffic matrices between leaf nodes of G . In other words, $\lambda^{\text{UC}} \in \Lambda_G^{\text{UC}}(n)$ if messages at the leaf nodes of G can be routed to their destinations (which are also leaf nodes) over G at rates λ^{UC} while respecting the capacity constraints on the edges of G . Define

$$\Lambda_G^{\text{BUC}}(n) \triangleq \Lambda_G^{\text{UC}}(n) \cap \mathcal{B}^{\text{UC}}(n).$$

We first prove the achievability part of Theorem 1. The next lemma shows that if traffic can be routed over the tree G then approximately the same traffic can be transmitted reliably over the wireless network.

Lemma 10: Under fast fading, for any $\alpha > 2$, there exists $b(n) \geq n^{-o(1)}$ such that for any $V(n) \in \mathcal{V}(n)$

$$b(n)\Lambda_G^{\text{UC}}(n) \subset \Lambda^{\text{UC}}(n).$$

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

Proof: Assume $\lambda^{\text{UC}} \in \Lambda_G^{\text{UC}}$, i.e., traffic can be routed between the leaf nodes of G at a rate λ^{UC} , we need to show that $n^{-o(1)}\lambda^{\text{UC}} \in \Lambda^{\text{UC}}$ (i.e., almost the same flow can be reliably transmitted over the wireless network). We use the three-layer communication architecture introduced in Section V-A to establish this result.

Recall the three layers of this architecture: the routing, cooperation, and physical layers. The layers of this communication scheme operate as follows. In the routing layer, we treat the wireless network as the graph G and route the messages between nodes over the edges of G . The cooperation layer provides this tree abstraction to the routing layer by distributing and concentrating messages over subsets of the wireless networks. The physical layer implements this distribution and concentration of messages by dealing with interference and noise.

Consider first the routing layer, and assume that the tree abstraction G can be implemented in the wireless network with only a $n^{-o(1)}$ factor loss. Since $\lambda^{\text{UC}} \in \Lambda_G^{\text{UC}}$ by assumption, we then know that the routing layer will be able to reliably transmit messages at rates $n^{-o(1)}\lambda^{\text{UC}}$ over the wireless network. We now

show that the tree abstraction can indeed be implemented with a factor $n^{-o(1)}$ loss in the wireless network.

This tree abstraction is provided to the routing layer by the cooperation layer. We will show that the operation of the cooperation layer satisfies the following invariance property: If a message is located at a node $\mu \in G$ in the routing layer, then the same message is evenly distributed over all nodes in $\mathcal{R}(\mu)$ in the wireless network. In other words, all nodes $u \in \mathcal{R}(\mu) \subset V$ contain a distinct part of length $1/|\mathcal{R}(\mu)|$ of the message.

Consider first a leaf node $u \in V \subset V_G$ in G , and assume the routing layer calls upon the cooperation layer to send a message to its parent $\nu \in V_G$ in G . Note first that u is also an element of V , and it has access to the entire message to be sent over G . Since for leaf nodes $\mathcal{R}(u) = \{u\}$, this shows that the invariance property is satisfied at u . The message is split at u into $|\mathcal{R}(\nu)|$ parts of equal length, and one part is sent to each node in $\mathcal{R}(\nu)$ over the wireless network. In other words, we distribute the message over the wireless network by a factor of $|\mathcal{R}(\nu)|$. Hence, the invariance property is also satisfied at ν .

Consider now an internal node $\mu \in V_G$, and assume the routing layer calls upon the cooperation layer to send a message to its parent node $\nu \in V_G$. Note that since all traffic in G originates at the leaf nodes of G (which are the actual nodes in the wireless network), a message at μ had to traverse all levels below μ in the tree G . We assume that the invariance property holds up to the level of μ in the tree, and show that it is then also satisfied at the level of ν . By the induction hypothesis, each node $u \in \mathcal{R}(\mu)$ has access to a distinct part of length $1/|\mathcal{R}(\mu)|$. Each such node u splits its message part into four distinct parts of equal length. Node u keeps one part for itself, and sends the other three parts to nodes in $\mathcal{R}(\nu)$. Since $|\mathcal{R}(\nu)| = 4|\mathcal{R}(\mu)|$, this can be performed such that each node in $\mathcal{R}(\nu)$ obtains exactly one message part. In other words, we distribute the message by a factor four over the wireless network, and the invariance property is satisfied at $\nu \in V_G$.

Operation along edges down the tree (i.e., towards the leaf nodes) is similar, but instead of distributing messages, we now concentrate them over the wireless network. To route a message from a node $\mu \in V_G$ with internal children $\{\nu_j\}_{j=1}^4$ to one of them (say ν_1) in the routing layer, the cooperation layer sends the message parts from each $\{\mathcal{R}(\nu_j)\}_{j=2}^4$ to a corresponding node in $\mathcal{R}(\nu_1)$ and combines them there. In other words, we concentrate the message by a factor four over the wireless network.

To route a message to a leaf node $u \in V \subset V_G$ from its parent ν in G in the routing layer, the cooperation layer sends the corresponding message parts at each node $\mathcal{R}(\nu)$ to u over the wireless network. Thus, again we concentrate the message over the network, but this time by a factor of $|\mathcal{R}(\nu)|$. Both these operations along edges down the tree preserve the invariance property. This shows that the invariance property is preserved by all operations induced by the routing layer in the cooperation layer.

Finally, to actually implement this distribution and concentration of messages, the cooperation layer calls upon the physical layer. Note that at the routing layer, all edges of the tree can be routed over simultaneously. Therefore, the cooperation layer can potentially call the physical layer to perform distribution and

concentration of messages over all sets $\{\mathcal{R}(\mu)\}_{\mu \in V_G}$ simultaneously. The function of the physical layer is to schedule all these operations and to deal with the resulting interference as well as with channel noise.

This scheduling is done as follows. First, the physical layer time shares between communication up the tree and communication down the tree (i.e., between distribution and concentration of messages). This results in a loss of a factor $1/2$ in rate. The physical layer further time shares between all the $L(n) + 1$ internal levels of the tree, resulting in a further $\frac{1}{L(n)+1}$ factor loss in rate. Hence, the total rate loss by this time sharing is

$$\frac{1}{2(L(n) + 1)}. \tag{24}$$

Consider now the operations within some level $\ell \in 1, \dots, L(n)$ in the tree (i.e., for edge (μ, ν) on this level, neither μ nor ν is a leaf node). We show that the rate at which the physical layer implements the edge (μ, ν) is equal to $n^{-o(1)}c_{\mu, \nu}$, i.e., only a small factor less than the capacity of the edge (μ, ν) in the tree G . Note first that the distribution or concentration of traffic induced by the cooperation layer to implement one edge e at level ℓ (i.e., between node levels ℓ and $\ell - 1$) is restricted to $V_{\ell-1, i}$ for some $i = i(e)$. We can thus partition the edges at level ℓ into $\{E_G^j\}_{j=1}^4$ such that for each j

$$V_{\ell-1, i(e)}_{e \in E_G^j}$$

partitions $V(n)$. Time sharing between the four values of j yields an additional loss of a factor $1/4$ in rate. Fix one such value of j , and consider the operations induced by the cooperation layer in the set corresponding to j . We consider communication up the tree (i.e., distribution of messages), the analysis for communication down the tree is similar. For a particular edge $(\mu, \nu) \in E_G^j$ with ν the parent of μ , each node $u \in \mathcal{R}(\mu)$ has split its message part into four parts, three of which need to be sent to the nodes in $\mathcal{R}(\nu) \setminus \mathcal{R}(\mu)$. Moreover, this assignment of destination nodes in $\mathcal{R}(\nu) \setminus \mathcal{R}(\mu)$ to u is performed such that no node in $\mathcal{R}(\nu) \setminus \mathcal{R}(\mu)$ is destination more than once. In other word, each node in $\mathcal{R}(\mu)$ is source exactly three times and each node in $\mathcal{R}(\nu) \setminus \mathcal{R}(\mu)$ is destination exactly once. This can be written as three source–destination pairings $\{\prod_{i(\mu, \nu)}^k\}_{k=1}^3$, on $V_{\ell-1, i(\mu, \nu)}$. Moreover, each such $\prod_{i(\mu, \nu)}^k$ can be understood as a subset of a permutation source–destination pairing. We time share between the three values of k (yielding an additional loss of a factor $1/3$ in rate). Now, for each value of k , Lemma 9 shows that by using either hierarchical relaying (for $\alpha \in (2, 3]$) or multihop communication (for $\alpha > 3$), we can communicate according to $\{\prod_{i(e)}^k\}_{e \in E_G^j}$ at a per-node rate of

$$n^{-o(1)}(4^{-\ell-1}n)^{1-\min\{3, \alpha\}/2}$$

under fast fading, and with probability⁴ $1 - o(1)$ also under slow fading. Since $\mathcal{R}(\mu)$ contains $4^{-\ell-1}n$ nodes, and accounting for the loss (24) for time sharing between the levels in G and the

⁴Note that Lemma 9 actually shows that all permutation traffic for every value of ℓ can be transmitted with high probability under slow fading. In other words, with high probability all levels of G can be implemented successfully under slow fading.

additional loss of factors $1/4$ and $1/3$ for time sharing between j and k , the physical layer implements an edge capacity for e at level ℓ of

$$\begin{aligned} & \frac{1}{2(L(n)+1)} \cdot \frac{1}{4} \cdot \frac{1}{3} \cdot 4^{-\ell-1} n \\ & \cdot n^{-o(1)} (4^{-\ell+1} n)^{1-\min\{3,\alpha\}/2} \\ & = n^{-o(1)} (4^{-\ell} n)^{2-\min\{3,\alpha\}/2} \\ & = n^{-o(1)} c_e. \end{aligned}$$

Consider now the operations within level $\ell = L(n) + 1$ in the tree (i.e., for edge (u, ν) on this level, u is a leaf node). We show that the rate at which the physical layer implements the edge (u, ν) is equal to $n^{-o(1)} c_{u,\nu}$. We again consider only communication up the tree (i.e., distribution of messages in the cooperation layer), communication down the tree is performed in a similar manner. The traffic induced by the cooperation layer at level $L(n) + 1$ is within the sets $V_{L(n),i}$ for $i = \{1, \dots, 4^{L(n)}\}$. Consider now communication within one $V_{L(n),i}$, and assume without loss of generality that in the routing layer every node $u \in V_{L(n),i}$ needs to send traffic along the edge (u, ν) . In the physical layer, we need to distribute a $1/|\mathcal{R}(\nu)|$ fraction of this traffic from each node $u \in V_{L(n),i}$ to every node in $\mathcal{R}(\nu) \subset V_{L(n),i}$. This can be expressed as $|V_{L(n),i}|$ source–destination pairings, and we time share between them. Accounting for the fact that only $1/|\mathcal{R}(\nu)|$ of traffic needs to be sent according to each pairing and since $V \in \mathcal{V}$, this results in a time sharing loss of at most a factor

$$\frac{|\mathcal{R}(\nu)|}{|V_{L(n),i}|} \leq \frac{1}{16}.$$

Now, using Lemma 9, all these source–destination pairings in all subsquares $\{V_{L(n),i}\}$ can be implemented simultaneously at a per node rate of

$$\begin{aligned} n^{-o(1)} \left(4^{-L(n)} n\right)^{1-\min\{3,\alpha\}/2} & \geq n^{-o(1)} \left(n^{\log^{-1/2}(n)}\right)^{-1/2} \\ & \geq n^{-o(1)}. \end{aligned}$$

Accounting for the loss (24) for time sharing between the levels in G , the additional factor $1/16$ loss for time sharing within each $V_{L(n),i}$, the physical layer implements an edge capacity for e at level $\ell = L(n) + 1$ of

$$\frac{1}{2(L(n)+1)} \cdot \frac{1}{16} \cdot n^{-o(1)} = n^{-o(1)} = n^{-o(1)} c_e$$

under either fast or slow fading.

Together, this shows that the physical and cooperation layers provide the tree abstraction G to the routing layer with edge capacities of only a factor $n^{-o(1)}$ loss. Hence, if messages can be routed at rates λ^{UC} between the leaf nodes of G , then messages can be reliably transmitted over the wireless network at rates $n^{-o(1)} \lambda^{\text{UC}}$. Hence

$$\lambda^{\text{UC}} \in \Lambda_G^{\text{UC}} \Rightarrow n^{-o(1)} \lambda^{\text{UC}} \in \Lambda^{\text{UC}}.$$

Noting that the $n^{-o(1)}$ factor is uniform in λ^{UC} , this shows that

$$n^{-o(1)} \Lambda_G^{\text{UC}} \subset \Lambda^{\text{UC}}.$$

□

We have seen that the unicast capacity region $\Lambda_G^{\text{UC}}(n)$ of the graph G under routing is (appropriately scaled) an inner bound to the unicast capacity region $\Lambda^{\text{UC}}(n)$ of the wireless network. Taking the intersection with the set of balanced traffic matrices $\mathcal{B}^{\text{UC}}(n)$ yields that the same holds for $\Lambda_G^{\text{BUC}}(n)$ and $\Lambda^{\text{BUC}}(n)$. The next lemma shows that $(\gamma(n) + 1) \Lambda_G^{\text{BUC}}(n)$ (with $\gamma(n) = n^{o(1)}$ as in the definition of $\mathcal{B}^{\text{UC}}(n)$ in (1)) is an outer bound to the approximate unicast capacity region $\hat{\Lambda}^{\text{BUC}}(n)$ of the wireless network as defined in (3). Combining Lemmas 5, 10, and 11 below, yields that with high probability

$$n^{-o(1)} \hat{\Lambda}^{\text{BUC}}(n) \subset n^{-o(1)} \Lambda_G^{\text{BUC}}(n) \subset \Lambda_G^{\text{BUC}}(n)$$

proving the achievability part of Theorem 1.

Lemma 11: For any $\alpha > 2$ and any $V(n) \in \mathcal{V}(n)$

$$\hat{\Lambda}^{\text{BUC}}(n) \subset (\gamma(n) + 1) \Lambda_G^{\text{BUC}}(n)$$

where $\gamma(n) = n^{o(1)}$ is the factor in the definition of $\mathcal{B}^{\text{UC}}(n)$ in (1).

Proof: We first relate the total traffic across an edge e in the graph G to the total traffic across a cut $V_{\ell,i}$ for some ℓ and i .

Consider an edge $e = (\mu, \nu) \in E_G$, and assume first that e connects nodes at level ℓ and $\ell - 1$ in the tree with $\ell \geq L(n)$. We slight abuse of notation, set

$$c_e \triangleq c_{\mu,\nu}.$$

Note first that by (23) we have

$$c_e = (4^{-\ell} n)^{2-\min\{3,\alpha\}/2}. \quad (25)$$

Moreover, since G is a tree, removing the edge e from E_G separates the tree into two connected components, say $S_1, S_2 \subset V_G$. Consider now the leaf nodes in S_1 . By the construction of the tree structure of G , these leaf nodes are either equal to $V_{\ell,i}$ or $V_{\ell,i}^c$ for some $i \in \{1, \dots, 4^\ell\}$. Assume without loss of generality that they are equal to $V_{\ell,i}$. Then $V_{\ell,i}^c$ are the leaf nodes in S_2 . Now since traffic is only assumed to be between leaf nodes of G , the total traffic demand between S_1 and S_2 is equal to

$$\sum_{u \in V_{\ell,i}} \sum_{w \notin V_{\ell,i}^c} (\lambda_{u,w}^{\text{UC}} + \lambda_{w,u}^{\text{UC}}). \quad (26)$$

By the tree structure of G , all this traffic has to be routed over edge e .

Consider now an edge e connecting a node at level $L(n) + 1$ and $L(n)$, i.e., a leaf node u to its parent ν . Then, by (22)

$$c_e = 1. \quad (27)$$

The total traffic crossing the edge e is equal to

$$\sum_{w \neq u} (\lambda_{u,w}^{\text{UC}} + \lambda_{w,u}^{\text{UC}}). \quad (28)$$

We now show that

$$\hat{\Lambda}^{\text{BUC}} \subset (\gamma(n) + 1) \Lambda_G^{\text{BUC}}. \quad (29)$$

Assume $\lambda^{\text{UC}} \in \hat{\Lambda}^{\text{BUC}}$, then

$$\sum_{u \in V_{\ell,i}} \sum_{w \notin V_{\ell,i}} \lambda_{u,w}^{\text{UC}} \leq (4^{-\ell} n)^{2-\min\{3,\alpha\}}$$

for all $\ell \in \{1, \dots, L(n)\}$, $i \in \{1, \dots, 4^\ell\}$, and

$$\sum_{w \neq u} (\lambda_{u,w}^{\text{UC}} + \lambda_{w,u}^{\text{UC}}) \leq 1$$

for all $u \in V$. Since λ^{UC} is balanced, this implies that

$$\frac{1}{\gamma(n)+1} \sum_{u \in V_{\ell,i}} \sum_{w \notin V_{\ell,i}} (\lambda_{u,w}^{\text{UC}} + \lambda_{w,u}^{\text{UC}}) \leq (4^{-\ell} n)^{2-\min\{3,\alpha\}}$$

for $\ell \leq L(n)$. By (25), (26), (27), and (28), we obtain that the traffic demand across each edge e of the graph G is less than $\gamma(n)+1$ times its capacity c_e . Therefore, using that G is a tree, $\frac{1}{\gamma(n)+1} \lambda^{\text{UC}}$ can be routed over G , i.e., $\lambda^{\text{UC}} \in (\gamma(n)+1) \Lambda_G^{\text{BUC}}$. This proves (29). \square

We now turn to the converse part of Theorem 1. The next lemma shows that $\hat{\Lambda}^{\text{UC}}(n)$ (appropriately scaled) is an outer bound to the unicast capacity region $\Lambda^{\text{UC}}(n)$ of the wireless network. Taking the intersection with the collection of balanced traffic matrices $\mathcal{B}^{\text{UC}}(n)$ and combining with Lemma 5, this shows that with high probability

$$\Lambda^{\text{BUC}}(n) \subset O(\log^6(n)) \hat{\Lambda}^{\text{BUC}}(n)$$

proving the converse part of Theorem 1.

Lemma 12: Under either fast or slow fading, for any $\alpha > 2$, there exists $b(n) = O(\log^6(n))$ such that for any $V(n) \in \mathcal{V}(n)$,

$$\Lambda^{\text{UC}}(n) \subset b(n) \hat{\Lambda}^{\text{UC}}(n).$$

Proof: Assume $\lambda^{\text{UC}} \in \Lambda^{\text{UC}}$. By Lemma 7, we have for any $\ell \in \{1, \dots, L(n)\}$ and $i \in \{1, \dots, 4^\ell\}$

$$\sum_{u \in V_{\ell,i}} \sum_{w \notin V_{\ell,i}} \lambda_{u,w}^{\text{UC}} \leq K \log^6(n) (4^{-\ell} n)^{2-\min\{3,\alpha\}/2} \quad (30)$$

for some constant K not depending on λ^{UC} .

Consider now $u \in V$. Lemma 6 shows that

$$\begin{aligned} \sum_{w \neq u} \lambda_{u,w}^{\text{UC}} &\leq \tilde{K} \log(n) \\ \sum_{w \neq u} \lambda_{w,u}^{\text{UC}} &\leq \tilde{K} \log(n) \end{aligned}$$

with constant \tilde{K} not depending on λ^{UC} and, therefore

$$\sum_{w \neq u} (\lambda_{u,w}^{\text{UC}} + \lambda_{w,u}^{\text{UC}}) \leq 2\tilde{K} \log(n). \quad (31)$$

Combining (30) and (31) proves that there exists $b(n) = O(\log^6(n))$ such that $\lambda^{\text{UC}} \in \Lambda^{\text{UC}}$ implies $\lambda^{\text{UC}} \in b(n) \hat{\Lambda}^{\text{UC}}$, proving the lemma. \square

VIII. PROOF OF THEOREM 2

Consider again the tree graph $G = (V_G, E_G)$ with leaf nodes $V(n) \subset V_G$ constructed in Section VII. As before, we consider traffic between leaf nodes of G . In particular, any multicast traffic matrix $\lambda^{\text{MC}} \in \mathbb{R}_+^{n \times 2^n}$ for the wireless network is also a multicast traffic matrix for the graph G . Denote by $\Lambda_G^{\text{MC}}(n) \subset \mathbb{R}_+^{n \times 2^n}$ the set of feasible (under routing) multicast traffic matrices between leaf nodes of G , and set

$$\Lambda_G^{\text{BMC}}(n) \triangleq \Lambda_G^{\text{MC}}(n) \cap \mathcal{B}^{\text{MC}}(n).$$

The next lemma shows that if multicast traffic can be routed over G then approximately the same multicast traffic can be transmitted reliably over the wireless network. Taking the intersection with $\mathcal{B}^{\text{MC}}(n)$ implies that the same result holds also for balanced traffic.

Lemma 13: Under fast fading, for any $\alpha > 2$, there exists $b(n) \geq n^{-o(1)}$ such that for all $V(n) \in \mathcal{V}(n)$

$$b(n) \Lambda_G^{\text{MC}}(n) \subset \Lambda^{\text{MC}}(n).$$

The same statement holds under slow fading with probability $1 - o(1)$ as $n \rightarrow \infty$.

Proof: The proof follows using the same construction as in Lemma 10. \square

We now show that, since G is a tree graph, $\hat{\Lambda}^{\text{BMC}}(n)$ is an inner bound (up to a factor $\gamma(n)+1$) to the the multicast capacity region $\Lambda_G^{\text{BMC}}(n)$. The fact that G is a tree is critical for this result to hold.

Lemma 14: For any $\alpha > 2$

$$\hat{\Lambda}^{\text{BMC}}(n) \subset (\gamma(n)+1) \Lambda_G^{\text{BMC}}(n).$$

where $\gamma(n) = n^{o(1)}$ is the factor in the definition of $\mathcal{B}^{\text{MC}}(n)$ in (2).

Proof: Assume $\lambda^{\text{MC}} \in \mathcal{B}^{\text{MC}} \setminus \Lambda_G^{\text{BMC}}$. Since G is a tree, there is only one way to route multicast traffic from u to W , namely along the subtree $G(\{u\} \cup W)$ induced by $\{u\} \cup W$ (i.e., the smallest subtree of G that covers $\{u\} \cup W$). Hence, for any edge $e \in E_G$, the traffic $d_{\lambda^{\text{MC}}}(e)$ that needs to be routed over e is equal to

$$d_{\lambda^{\text{MC}}}(e) = \sum_{\substack{u \in V, W \subset V: \\ e \in E_G(\{u\} \cup W)}} \lambda_{u,W}^{\text{MC}}.$$

Now, since $\lambda^{\text{MC}} \in \mathcal{B}^{\text{MC}} \setminus \Lambda_G^{\text{BMC}}$, there exists $e \in E_G$ such that

$$d_{\lambda^{\text{MC}}}(e) > c_e. \quad (32)$$

Let ℓ be the level of this edge e in G . We have

$$c_e = \begin{cases} (4^{-\ell} n)^{2-\min\{3,\alpha\}/2}, & \text{if } \ell \leq L(n) \\ 1, & \text{else.} \end{cases} \quad (33)$$

Assume first that $\ell \leq L(n)$ and let i be such that the removal of the edge e in G disconnects the leaf nodes in $V_{\ell,i}$ from the ones in $V_{\ell,i}^c$. Then we have

$$d_{\lambda^{\text{MC}}}(e) = \sum_{u \in V_{\ell,i}} \sum_{\substack{W \subset V: \\ W \setminus V_{\ell,i} \neq \emptyset}} \lambda_{u,W}^{\text{MC}} + \sum_{u \notin V_{\ell,i}(n)} \sum_{\substack{W \subset V: \\ W \cap V_{\ell,i} \neq \emptyset}} \lambda_{u,W}^{\text{MC}}. \quad (34)$$

Assume then that $\ell = L(n) + 1$, and assume e separates the leaf node u from $\{u\}^c$ in G . Then

$$d_{\lambda^{\text{MC}}}(e) = \sum_{\substack{W \subset V: \\ W \setminus \{u\} \neq \emptyset}} \lambda_{u,W}^{\text{MC}} + \sum_{\tilde{u} \neq u} \sum_{\substack{W \subset V: \\ u \in W}} \lambda_{\tilde{u},W}^{\text{MC}}. \quad (35)$$

If $\ell = L(n) + 1$, then (32), (33), and (35) imply that $\lambda^{\text{MC}} \notin \hat{\Lambda}^{\text{BMC}}$ and, therefore, $\lambda^{\text{MC}} \notin \frac{1}{\gamma(n)+1} \hat{\Lambda}^{\text{BMC}}$. If $\ell \leq L(n)$ then, since λ^{MC} is $\gamma(n)$ -balanced, we have

$$\begin{aligned} \sum_{u \in V_{\ell,i}} \sum_{\substack{W \subset V: \\ W \setminus V_{\ell,i} \neq \emptyset}} \lambda_{u,W}^{\text{MC}} + 1 \sum_{u \notin V_{\ell,i}(n)} \sum_{\substack{W \subset V: \\ W \cap V_{\ell,i} \neq \emptyset}} \lambda_{u,W}^{\text{MC}} \\ \leq (\gamma(n) + 1) \sum_{u \in V_{\ell,i}} \sum_{\substack{W \subset V: \\ W \setminus V_{\ell,i} \neq \emptyset}} \lambda_{u,W}^{\text{MC}}. \end{aligned} \quad (36)$$

Combining (32), (33), (34), and (36) shows that $\lambda^{\text{MC}} \notin \frac{1}{\gamma(n)+1} \hat{\Lambda}^{\text{BMC}}$ for $\ell \leq L(n)$, as well.

Hence, we have shown that $\lambda^{\text{MC}} \in \mathcal{B}^{\text{MC}} \setminus \Lambda_G^{\text{BMC}}$ implies $\lambda^{\text{MC}} \notin \frac{1}{\gamma(n)+1} \hat{\Lambda}^{\text{BMC}}$, proving the lemma. \square

Combining Lemmas 13, and 14, and 5 shows that, with probability $1 - o(1)$ as $n \rightarrow \infty$

$$n^{-o(1)} \hat{\Lambda}^{\text{BMC}}(n) \subset n^{-o(1)} \Lambda_G^{\text{BMC}}(n) \subset \Lambda^{\text{BMC}}(n)$$

proving the inner bound in Theorem 2.

We now turn to the proof of the outer bound to $\Lambda^{\text{MC}}(n)$. The next lemma combined with Lemma 5, and taking the intersection with $\mathcal{B}^{\text{MC}}(n)$, proves the outer bound in Theorem 2.

Lemma 15: Under fast fading, for any $\alpha > 2$, there exists $b(n) = O(\log^6(n))$ such that for all $V(n) \in \mathcal{V}(n)$

$$\Lambda^{\text{MC}}(n) \subset b(n) \hat{\Lambda}^{\text{MC}}(n).$$

The same statement holds under slow fading with probability $1 - o(1)$ as $n \rightarrow \infty$.

Proof: We say that a unicast traffic matrix λ^{UC} is *compatible* with a multicast traffic matrix λ^{MC} if there exists a mapping $f : V(n) \times 2^{V(n)} \rightarrow V(n)$ such that $f(u, W) \in W \cup \{u\}$, for all (u, W) , and

$$\lambda_{u,w}^{\text{UC}} = \sum_{\substack{W \subset V(n): \\ f(u,W)=w}} \lambda_{u,W}^{\text{MC}}$$

for all (u, w) . In words, λ^{MC} is compatible with λ^{UC} if we can create the unicast traffic matrix λ^{UC} from λ^{MC} by simply dis-

carding the traffic for the pair (u, W) at all the nodes $W \setminus \{f(u, W)\}$.

Note that if $\lambda^{\text{MC}} \in \Lambda^{\text{MC}}$ and if λ^{UC} is compatible with λ^{MC} then $\lambda^{\text{UC}} \in \Lambda^{\text{UC}}$. Indeed, we can reliably transmit at rate λ^{UC} by using the communication scheme for λ^{MC} and discarding all the unwanted messages delivered by this scheme. Now consider a cut $V_{\ell,i}$ with $\ell \leq L(n)$ in the wireless network, and choose a mapping $f : V(n) \times 2^{V(n)} \rightarrow V(n)$ such that

$$\sum_{u \in V_{\ell,i}} \sum_{w \notin V_{\ell,i}} \lambda_{u,w}^{\text{UC}} = \sum_{u \in V_{\ell,i}} \sum_{\substack{W \subset V: \\ W \setminus V_{\ell,i} \neq \emptyset}} \lambda_{u,W}^{\text{MC}}.$$

Since $\lambda^{\text{UC}} \in \Lambda^{\text{UC}}$, we can apply Lemma 7 to obtain

$$\begin{aligned} \sum_{u \in V_{\ell,i}} \sum_{\substack{W \subset V: \\ W \setminus V_{\ell,i} \neq \emptyset}} \lambda_{u,W}^{\text{MC}} \\ = \sum_{u \in V_{\ell,i}} \sum_{w \notin V_{\ell,i}} \lambda_{u,w}^{\text{UC}} \leq b(n) (4^{-\ell} n)^{2 - \min\{3, \alpha\}/2} \end{aligned}$$

with $b(n) = O(\log^6(n))$. Repeating the same argument for cuts of the form $\{u\}$ and $\{u\}^c$ and using Lemma 6, shows that $\lambda^{\text{MC}} \in b(n) \hat{\Lambda}^{\text{MC}}$. Noting that the $b(n)$ term is uniform in λ^{MC} yields that

$$\Lambda^{\text{MC}} \subset b(n) \hat{\Lambda}^{\text{MC}}$$

concluding the proof of the lemma. \square

IX. DISCUSSION

We discuss several aspects and extensions of the three-layer architecture introduced in Section V-C and used in the achievability parts of Theorems 2 and 1. In Section IX-A, we comment on the various tree structures used in the three-layer architecture. In Section IX-B we show that for certain values of α the bounds in the theorems can be significantly sharpened. In Section IX-C, we discuss bounds for traffic that is not balanced. In Section IX-D, we show that for large values of path-loss exponent ($\alpha > 5$) these bounds are tight. Hence, in the large path-loss regime the requirement of balanced traffic is not necessary, and we obtain a scaling characterization of the entire unicast and multicast capacity regions. In Section IX-E, we point out how the results discussed so far can be used to obtain the scaling of the unicast and multicast capacity regions of dense networks (where n nodes are randomly placed on a square of unit area).

A. Tree Structures

There are two distinct tree structures that are used in the construction of the three-layer communication scheme proposed in this paper—one explicit and one implicit. These two tree structures appear in different layers of the communication scheme and serve different purposes.

The first (explicit) tree structure is given by the tree G utilized in the routing layer and implemented in the cooperation layer. The main purpose of this tree structure is to perform localized load balancing. In fact, the distribution and concentration of traffic is used to avoid unnecessary bottlenecks. Note that the tree G is used by the scheme for any value of α .

The second (implicit) tree structure occurs in the physical layer. This tree structure appears only for $\alpha \in (2, 3]$. In this regime, the physical layer uses the hierarchical relaying scheme. It is the hierarchical structure of this scheme that can equivalently be understood as a tree. The purpose of this second tree structure is to enable distributed multiple-antenna communication, i.e., to perform cooperative communication.

B. Second-Order Asymptotics

The scaling results in Theorems 1 and 2 are up to a factor $n^{\pm o(1)}$ and, hence, preserve information at scale n^β for constant β (see also the discussion in Section III-C). Here, we examine in more detail the behavior of this $n^{\pm o(1)}$ factor and show that in certain situations it can be significantly sharpened.

Note first that the outer bounds in Theorems 1 and 2 hold up to a factor $O(\log^6(n))$, i.e., poly-logarithmic in n . However, the inner bounds hold only up to the aforementioned $n^{-o(1)}$ factor. A closer look at the proofs of the two theorems reveals that the precise inner bound is of order

$$\gamma^{-1}(n)n^{-O(\log^{-1/3}(n))}$$

where $\gamma(n)$ is the factor in the definition of $\mathcal{B}^{\text{UC}}(n)$ and $\mathcal{B}^{\text{MC}}(n)$ [see (1) and (2)]. With a more careful analysis (see [13] for the details), this can be sharpened to essentially

$$\gamma^{-1}(n)n^{-O(\log^{-1/2}(n))}.$$

The exponent $\log^{-1/2}(n)$ in the inner bound has two causes. The first is the use of hierarchical relaying (for $\alpha \in (2, 3]$). The second is the operation of the physical layer at level $L(n) + 1$ of the tree (i.e., to implement communication between the leaf nodes of G and their parents). Indeed at that level, we are operating on a square of area

$$4^{-L(n)}n = n^{\log^{-1/2}(n)}$$

and the loss is essentially inversely proportional to that area. Now, the reason why $L(n)$ can not be chosen to be larger (to make this loss smaller), is because hierarchical relaying requires a certain amount of regularity in the node placement, which can only be guaranteed for large enough areas.

This suggests that for the $\alpha > 3$ regime, where multihop communication is used at the physical layer instead of hierarchical relaying, we might be able to significantly improve the inner bound. To this end, we have to choose more levels in the tree G , such that at the last level before the tree nodes, we are operating on a square that has an area of order $\log(n)$. Changing the three-layer architecture in this manner, and choosing $\gamma(n)$ appropriately, for $\alpha > 3$ the inner bound can be improved to $\Omega(\log^{-2}(n))$ in n . Combined with the poly-logarithmic outer bound, this yields a $O(\log^8(n))$ approximation of the balanced unicast and multicast capacity regions for $\alpha > 3$.

C. Nonbalanced Traffic

Theorems 1 and 2 describe the scaling of the balanced unicast and multicast capacity regions $\Lambda^{\text{BUC}}(n)$ and $\Lambda^{\text{BMC}}(n)$, respectively. As we have argued, the balanced unicast region $\Lambda^{\text{BUC}}(n)$

coincides with the unicast capacity region $\Lambda^{\text{UC}}(n)$ along at least $n^2 - n$ out of n^2 total dimensions, and the balanced multicast region $\Lambda^{\text{BMC}}(n)$ coincides with the multicast capacity region $\Lambda^{\text{MC}}(n)$ along at least $n2^n - n$ out of $n2^n$ total dimensions. However, the proofs of these results provide also bounds for traffic that is not balanced, i.e., for the remaining $2n$ dimensions.

Define the following two regions:

$$\hat{\Lambda}_1^{\text{UC}}(n) \triangleq \left\{ \lambda^{\text{UC}} \in \mathbb{R}_+^{n \times n} : \sum_{u \in V_{\ell,i}(n)} \sum_{w \notin V_{\ell,i}(n)} (\lambda_{u,w}^{\text{UC}} + \lambda_{w,u}^{\text{UC}}) \leq (4^{-\ell}n)^{2-\min\{3,\alpha\}/2} \forall \ell \in \{1, \dots, L(n)\}, i \in \{1, \dots, 4^\ell\} \right. \\ \left. \sum_{w \neq u} (\lambda_{u,w}^{\text{UC}} + \lambda_{w,u}^{\text{UC}}) \leq 1 \forall u \in V(n) \right\}$$

and

$$\hat{\Lambda}_1^{\text{MC}}(n) \triangleq \left\{ \lambda^{\text{MC}} \in \mathbb{R}_+^{n \times 2^n} : \sum_{u \in V_{\ell,i}(n)} \sum_{\substack{W \subset V(n): \\ W \setminus V_{\ell,i}(n) \neq \emptyset}} \lambda_{u,W}^{\text{MC}} + \sum_{u \notin V_{\ell,i}(n)} \sum_{\substack{W \subset V(n): \\ W \cap V_{\ell,i}(n) \neq \emptyset}} \lambda_{u,W}^{\text{MC}} \leq (4^{-\ell}n)^{2-\min\{3,\alpha\}/2} \forall \ell \in \{1, \dots, L(n)\}, i \in \{1, \dots, 4^\ell\} \right. \\ \left. \sum_{\substack{W \subset V(n): \\ W \setminus \{u\} \neq \emptyset}} \lambda_{u,W}^{\text{MC}} + \sum_{\substack{\tilde{u} \neq u \\ W \subset V(n): \\ u \in W}} \lambda_{\tilde{u},W}^{\text{MC}} \leq 1 \forall u \in V(n) \right\}.$$

$\hat{\Lambda}_1^{\text{UC}}(n)$ and $\hat{\Lambda}_1^{\text{MC}}(n)$ differ in that for $\ell \in \{1, \dots, L(n)\}$, $\hat{\Lambda}_1^{\text{UC}}(n)$ only bounds traffic flow out of $V_{\ell,i}(n)$, whereas $\hat{\Lambda}_1^{\text{MC}}(n)$ bounds traffic in both directions across $V_{\ell,i}(n)$ (and similar for $\hat{\Lambda}_1^{\text{MC}}(n)$ and $\hat{\Lambda}_1^{\text{UC}}(n)$).

The analysis in Sections VII and VIII shows that

$$n^{-o(1)}\hat{\Lambda}_1^{\text{UC}}(n) \subset \Lambda^{\text{UC}}(n) \subset O(\log^6(n))\hat{\Lambda}_1^{\text{UC}}(n) \\ n^{-o(1)}\hat{\Lambda}_1^{\text{MC}}(n) \subset \Lambda^{\text{MC}}(n) \subset O(\log^6(n))\hat{\Lambda}_1^{\text{MC}}(n)$$

with probability $1 - o(1)$ as $n \rightarrow \infty$. In other words, we obtain an inner and an outer bound on the capacity regions $\Lambda^{\text{UC}}(n)$ and $\Lambda^{\text{MC}}(n)$. These bounds coincide in the scaling sense for balanced traffic, for which we recover Theorems 1 and 2.

D. Large Path-Loss Exponent Regime

The discussion in Section IX-C reveals that in order to obtain scaling information for traffic that is not balanced, a stronger

version of the converse results in Lemma 7 is needed. In particular, Lemma 7 bounds the sum-rate

$$\sum_{u \in V_{\ell,i}(n)} \sum_{w \notin V_{\ell,i}(n)} \lambda_{u,w}^{\text{UC}}$$

for $\lambda^{\text{UC}} \in \Lambda^{\text{UC}}(n)$. The required stronger version of the lemma would also need to bound sum rates in the other direction, i.e.,

$$\sum_{u \notin V_{\ell,i}(n)} \sum_{w \in V_{\ell,i}(n)} \lambda_{u,w}^{\text{UC}}.$$

For large path-loss exponents $\alpha > 5$, such a stronger version of Lemma 7 holds (see Lemma 8). With this, we obtain that for $\alpha > 5$

$$\begin{aligned} n^{-o(1)} \hat{\Lambda}_1^{\text{UC}}(n) &\subset \Lambda^{\text{UC}}(n) \subset O(\log^6(n)) \hat{\Lambda}_1^{\text{UC}}(n) \\ n^{-o(1)} \hat{\Lambda}_1^{\text{MC}}(n) &\subset \Lambda^{\text{MC}}(n) \subset O(\log^6(n)) \hat{\Lambda}_1^{\text{MC}}(n) \end{aligned}$$

with probability $1 - o(1)$ as $n \rightarrow \infty$. In other words, in the high path-loss exponent regime $\alpha > 5$, $\hat{\Lambda}_1^{\text{UC}}(n)$ and $\hat{\Lambda}_1^{\text{MC}}(n)$ characterize the scaling of the entire unicast and multicast capacity regions, respectively.

E. Dense Networks

So far, we have only discussed *extended* networks, i.e., n nodes are located on a square of area n . We now briefly sketch how these results can be recast for *dense* networks, in which n nodes are located on a square of unit area.

Note first that by rescaling power by a factor $n^{-\tilde{\alpha}/2}$, a dense network with any path-loss exponent α can essentially be transformed into an equivalent extended network with path-loss exponent $\tilde{\alpha}$. In particular, any scheme for extended networks with path-loss exponent $\tilde{\alpha}$ yields a scheme with same performance for dense networks with any path-loss exponent α (see also [12, Section V-A]). To optimize the resulting scheme for the dense network, we start with the scheme for extended networks corresponding to $\tilde{\alpha}$ close to 2. Hence an inner bound for the unicast and multicast capacity regions for dense networks with path-loss exponent α can be obtained from the ones for extended networks by taking a limit as $\tilde{\alpha} \rightarrow 2$. Moreover, an application of Lemma 6 yields a matching (in the scaling sense) outer bound.

The resulting approximate balanced capacity regions $\hat{\Lambda}^{\text{BUC}}(n)$ and $\hat{\Lambda}^{\text{BMC}}(n)$ have particularly simple shapes in this limit. In fact, the only constraints in (3) and (4) that can be tight are at level $\ell = \log(n)$. Moreover, as in Section IX-D, it can be shown that the restriction of balanced traffic is not necessary for dense networks. This results in the following approximate capacity regions for dense networks:

$$\hat{\Lambda}^{\text{UC}}(n) \triangleq \left\{ \lambda^{\text{UC}} \in \mathbb{R}_+^{n \times n} : \sum_{w \neq u} (\lambda_{u,w}^{\text{UC}} + \lambda_{w,u}^{\text{UC}}) \leq 1, \forall u \in V(n) \right\}$$

for unicast, and

$$\hat{\Lambda}^{\text{MC}}(n) \triangleq \left\{ \lambda^{\text{MC}} \in \mathbb{R}_+^{n \times 2^n} : \sum_{\substack{W \subset V(n): \\ W \setminus \{u\} \neq \emptyset}} \lambda_{u,W}^{\text{MC}} + \sum_{\tilde{u} \neq u} \sum_{\substack{W \subset V(n): \\ u \in W}} \lambda_{\tilde{u},W}^{\text{MC}} \leq 1 \right. \\ \left. \forall u \in V(n) \right\}$$

for multicast. We obtain that for dense networks, for any $\alpha > 2$

$$\begin{aligned} n^{-o(1)} \hat{\Lambda}^{\text{UC}}(n) &\subset \Lambda^{\text{UC}}(n) \subset O(\log^6(n)) \hat{\Lambda}^{\text{UC}}(n) \\ n^{-o(1)} \hat{\Lambda}^{\text{MC}}(n) &\subset \Lambda^{\text{MC}}(n) \subset O(\log^6(n)) \hat{\Lambda}^{\text{MC}}(n) \end{aligned}$$

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

X. CONCLUSION

In this paper, we have obtained an explicit information-theoretic characterization of the scaling of the n^2 -dimensional balanced unicast and $n2^n$ -dimensional balanced multicast capacity regions of a wireless network with n randomly placed nodes and assuming a Gaussian fading channel model. These regions span at least $n^2 - n$ and $n2^n - n$ dimensions of $\mathbb{R}_+^{n \times n}$ and $\mathbb{R}_+^{n \times 2^n}$, respectively, and, hence, determine the scaling of the unicast capacity region along at least $n^2 - n$ out of n^2 dimensions and the scaling of the multicast capacity region along at least $n2^n - n$ out of $n2^n$ dimensions. The characterization is in terms of $2n$ weighted cuts, which are based on the geometry of the locations of the source nodes and their destination nodes and on the traffic demands between them, and thus can be readily evaluated.

This characterization is obtained by establishing that the unicast and multicast capacity regions of a capacitated (wireline, noiseless) tree graph under routing have essentially the same scaling as that of the original network. The leaf nodes of this tree graph correspond to the nodes in the wireless network, and internal nodes of the tree graph correspond to hierarchically growing sets of nodes.

This equivalence suggests a three-layer communication architecture for achieving the entire balanced unicast and multicast capacity regions (in the scaling sense). The top or routing layer establishes paths from each of the source nodes to its destination (for unicast) or set of destinations (for multicast) over the tree graph. The middle or cooperation layer provides this tree abstraction to the routing layer by distributing the traffic among the corresponding set of nodes as a message travels up the tree graph, and by concentrating the traffic on to the corresponding set of nodes as a message travels down the tree. The bottom or physical layer implements this distribution and concentration of traffic over the wireless network. This implementation depends on the path-loss exponent: For low path loss, $\alpha \in (2, 3]$, hierarchical relaying is used, while for high path loss ($\alpha > 3$), multihop communication is used.

This scheme also establishes that a separation based approach, in which the routing layer works essentially independently of the physical layer, can achieve nearly the entire balanced unicast and multicast capacity regions in the scaling sense. Thus, for balanced traffic, such techniques as network coding can provide at most a small increase in the scaling.

ACKNOWLEDGMENT

The authors would like to thank D. Tse, G. Wornell, and L. Zheng for helpful discussions, and the anonymous reviewers for their help in improving the presentation of this paper.

REFERENCES

- [1] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.
- [2] L.-L. Xie and P. R. Kumar, "A network information theory for wireless communication: Scaling laws and optimal operation," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 748–767, May 2004.
- [3] A. Jovičić, P. Viswanath, and S. R. Kulkarni, "Upper bounds to transport capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2555–2565, Nov. 2004.
- [4] F. Xue, L.-L. Xie, and P. R. Kumar, "The transport capacity of wireless networks over fading channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 834–847, Mar. 2005.
- [5] O. Lévêque and I. E. Telatar, "Information-theoretic upper bounds on the capacity of large extended ad hoc wireless networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 858–865, Mar. 2005.
- [6] L.-L. Xie and P. R. Kumar, "On the path-loss attenuation regime for positive cost and linear scaling of transport capacity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2313–2328, Jun. 2006.
- [7] M. Franceschetti, O. Dousse, D. N. C. Tse, and P. Thiran, "Closing the gap in the capacity of wireless networks via percolation theory," *IEEE Trans. Inf. Theory*, vol. 53, no. 3, pp. 1009–1018, Mar. 2007.
- [8] P. Gupta and P. R. Kumar, "Towards an information theory of large networks: An achievable rate region," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1877–1894, Aug. 2003.
- [9] L.-L. Xie and P. R. Kumar, "An achievable rate for the multiple-level relay channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1348–1358, Apr. 2005.
- [10] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.
- [11] S. Aeron and V. Saligrama, "Wireless ad hoc networks: Strategies and scaling laws for the fixed SNR regime," *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 2044–2059, Jun. 2007.
- [12] A. Özgür, O. Lévêque, and D. N. C. Tse, "Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3549–3572, Oct. 2007.
- [13] U. Niesen, P. Gupta, and D. Shah, "On capacity scaling in arbitrary wireless networks," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 3959–3982, Sep. 2009.
- [14] R. Madan, D. Shah, and O. Lévêque, "Product multicommodity flow in wireless networks," *IEEE Trans. Inf. Theory*, vol. 54, no. 4, pp. 1460–1476, Apr. 2008.
- [15] F. T. Leighton and S. Rao, "Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms," *J. ACM*, vol. 46, no. 6, pp. 787–832, Nov. 1999.
- [16] N. Linial, E. London, and Y. Rabinovich, "The geometry of graphs and some of its algorithmic applications," *Combinatorica*, vol. 15, no. 2, pp. 215–245, Jun. 1995.
- [17] S. Subramanian, S. Shakkottai, and P. Gupta, "On optimal geographic routing in wireless networks with holes and non-uniform traffic," in *Proc. IEEE Infocom*, May 2007, pp. 1328–1336.
- [18] S. Subramanian, S. Shakkottai, and P. Gupta, "Optimal geographic routing for wireless networks with near-arbitrary holes and traffic," in *Proc. IEEE Infocom*, Apr. 2008, pp. 1019–1027.
- [19] S. H. A. Ahmad, A. Jovičić, and P. Viswanath, "On outer bounds to the capacity region of wireless networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2770–2776, Jun. 2006.
- [20] X.-Y. Li, S.-J. Tang, and O. Frieder, "Multicast capacity for large scale wireless ad hoc networks," in *Proc. ACM MobiCom*, Sep. 2007, pp. 263–277.
- [21] S. Shakkottai, X. Liu, and R. Srikant, "The multicast capacity of large multihop wireless networks," in *Proc. ACM MobiHoc*, Sep. 2007, pp. 247–255.
- [22] A. Keshavarz-Haddad, V. Ribeiro, and R. Riedi, "Broadcast capacity in multihop wireless networks," in *Proc. ACM MobiCom*, Sep. 2006, pp. 239–250.
- [23] P. Chaporkar and S. Sarkar, "Wireless multicast: Theory and approaches," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1954–1972, Jun. 2005.
- [24] P. Jacquet and G. Rodolakis, "Multicast scaling properties in massively sense ad hoc networks," in *Proc. Int. Conf. Parallel and Distributed Systems*, Jul. 2005, vol. 2, pp. 93–99.
- [25] M. Franceschetti, M. D. Migliore, and P. Minero, "The capacity of wireless networks: Information-theoretic and physical limits," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3413–3424, Aug. 2009.
- [26] M. Franceschetti, M. D. Migliore, and P. Minero, "The degrees of freedom of wireless networks: Information theoretic and physical limits," in *Proc. Allerton Conf. Communication, Control, and Computing*, Sep. 2008, pp. 1578–1581.
- [27] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

Urs Niesen received the M.S. degree from the School of Computer and Communication Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland, in 2005, and the Ph.D. degree from the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, in 2009.

He is currently a member of technical staff at Alcatel-Lucent, Bell Laboratories, Murray Hill, NJ. His research interests are in the areas of communication and information theory.

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

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

Devavrat Shah received the B.Tech. from the Indian Institute of Technology, Bombay, India, in 1999, and the Ph.D. degree from Stanford University, Stanford, CA, in 2004.

He is currently a Jamieson Career Development Associate Professor with the Department of Electrical Engineering and Computer Science at the Massachusetts Institute of Technology, Cambridge. His research interests include statistical inference and network algorithms.

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