

# Cost of Local Cooperation in Hierarchical Virtual MIMO Transmission Schemes

Jinfeng Du  
 Nokia Bell Labs  
 Holmdel, 07733, NJ  
 Email: jinfeng.du@bell-labs.com

Muriel Médard  
 MIT  
 Cambridge, 02139, MA  
 Email: medard@mit.edu

Shlomo Shamai (Shitz)  
 Technion  
 Technion City, Haifa 32000, Israel  
 Email: sshlomo@ee.technion.ac.il

**Abstract**—Hierarchical cooperation schemes in wireless networks rely on local cooperation among neighboring nodes to create virtual multiple-input multiple-output (MIMO) connections between clusters of nodes. It was shown that, by applying the virtual MIMO technique recursively in a hierarchical manner, the sum rate of all source-destination pairs can scale linearly with the number of nodes in the network. In this paper we focus on the impact of local cooperation and establish new capacity scaling bounds for the virtual MIMO transmission taking into account the constraints of local communication both at the transmitters and the receivers. We show that the cost of local communication, which is inevitable to establish the virtual MIMO transmission, grows exponentially with the number of layers in the cooperation hierarchy and plays a vital role in determining the overall performance of the hierarchical virtual MIMO cooperation.

**Index Terms**—Scalability, distributed cooperation, wireless networks, virtual MIMO

## I. INTRODUCTION

Given a large wireless network consisting of  $N$  nodes randomly deployed within an area  $A$ , each of the nodes wants to transmit to a random destination node within the network at some equal rate. For *dense networks* where the area  $A$  is fixed but the density of nodes scales up, Gupta and Kumar [1] showed that, with multi-hop transmission, the maximum achievable sum rate  $C_N$  scales at most as  $O(\sqrt{N})$  as dense networks are essentially interference limited. For *extended networks* where the transmission distance scales up with the number of nodes but the density of nodes  $N/A_N$  is fixed, Xie and Kumar [2] provided an upper bound on the capacity scaling and showed that the multiple-hop transmission strategy is essentially optimal if the path loss attenuation factor is high. A hierarchical cooperation scheme proposed by Özgür et al. [3] employs local communication among neighboring nodes to create virtual multiple-input multiple-output (MIMO) connections between source-destination pairs, and such virtual MIMO technique is then applied recursively to solve the local communication problems by formulating new virtual MIMO connections at a smaller scale. It was shown in [3] that, with perfect channel state information at the receivers, the maximal sum rate  $C_N$  can scale as  $O(N^{1-\epsilon})$  for any  $\epsilon > 0$  in dense

This material is based upon work supported by the Air Force Office of Scientific Research (AFOSR) under award No. FA9550-13-1-0023, and by the MIT Wireless Center. The work of S. Shamai has been supported by the European Union's Horizon 2020 Research And Innovation Programme, grant agreement no. 694630.

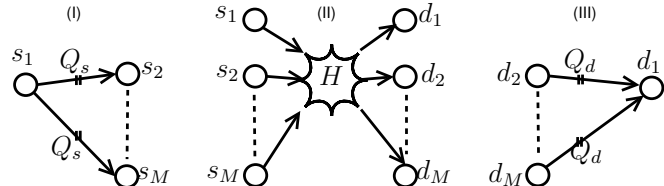


Fig. 1. Virtual MIMO model: local communication among transmitters (I) and among receivers (III) to facilitate the virtual MIMO transmission (II).

networks. This hierarchical cooperation scheme was later refined in [4], [5] to maximize the achievable rate. However, the cost of local communications, which grows exponentially with the number of layers in the cooperation hierarchy, was not fully explored in previous analysis [3]–[5]. For dense networks, the number of nodes  $N$  cannot be too large to ensure independent channels [6]–[8]. For extended networks, the received signal to noise ratio (SNR) becomes inevitably low and the channel knowledge can no longer be assumed available for free, which hinders the feasibility of quantization based joint detection in the distributed receivers [9]. Therefore the number of nodes  $N$  cannot be too large either.

In this paper we first quantify the cost of local communication for the hierarchical virtual MIMO transmission without imposing any limit on the total number of nodes  $N$  in the network, and then discuss the consequence of imposing such a constraint. We generalize the virtual MIMO model [3] to accommodate flexible local collaboration among neighboring nodes. As shown in Fig. 1, for each source-destination pair (take  $s_1$  and  $d_1$  for example), the transmission is split into three stages. In Stage I the source node  $s_1$  distributes its message to all the neighboring nodes ( $s_2, \dots, s_M$ ), each of  $Q_s$  bits, via local communication, to form a virtual multiple-antenna transmitter. In Stage II all the  $M$  nodes ( $s_1, \dots, s_M$ ) transmit to all the  $M$  nodes in the destination cluster ( $d_1, \dots, d_M$ ) using the same channel (i.e., in the same time and frequency resource block), and the transmission lasts a duration of one time slot (i.e., one channel use)<sup>1</sup>. In Stage III all nodes in the destination cluster first independently quantize their observation into  $Q_d$  bits and then forward them to the destination node via local communication. The destination node  $d_1$ , after

<sup>1</sup>Any predefined time slot works here. For example, choosing one second as the time unit will facilitate direct comparison of virtual MIMO transmission rate [bits per second] and local cooperation message size  $Q_s$  and  $Q_d$  [bits].

receiving all the quantized bits from its neighboring nodes ( $d_2, \dots, d_M$ ) in the cluster, will perform joint detection to retrieve the  $MQ_s$  bits sent by the source node  $s_1$ . To focus on the impact of local communication, we assume that the virtual MIMO transmission in Stage II is capacity achieving in the sense that we can recover the transmitted message successfully if decoding is done based on the original observations (without quantization) of all the receiving nodes in the cluster. The virtual MIMO model in Fig. 1 nicely captures the essence of multiple-cell cooperation [10], [11] in cellular networks, where base stations are connected through finite-rate backhaul and the user terminals are capable of communicating directly with other users in its close vicinity.

## II. SYSTEM MODEL

We inherit the grid topology from [3] such that there are  $N$  nodes evenly distributed in a square and every node wants to transmit the same amount data to a destination node randomly chosen from the rest of the  $N-1$  nodes. All the transmissions are carried out in the same frequency band and the aggregate throughput of all the  $N$  source-destination pairs,  $C_N$ , is referred as the sum rate. For a pair of transmit and receive clusters at distance  $d_{st}$  meters apart, the SNR for the transmission between node  $i$  in the transmit cluster and node  $j$  in the receive cluster is written as

$$\text{SNR}_{i,j} = P_i(d_{st})d_{ij}^{-\alpha}\gamma_{i,j}E[|h_{i,j}|^2] \approx \gamma(d_{st}), \quad \forall i, j, \quad (1)$$

where  $P_i(d_{st})$  is the transmit power of node  $i$ ,  $d_{ij}$  is the actual distance between node  $i$  in the transmit cluster and node  $j$  in the receive cluster, and  $\alpha \geq 2$  is the attenuation factor. We assume that all channels  $h_{i,j}$  are i.i.d. with  $E[|h_{i,j}|^2]=1$ . All other factors such as the noise power, antenna gains, and other loss are modeled through  $\gamma_{i,j}$ , which has marginal variation. The approximation in (1) is taken to ignore the minor variation of SNR across different transmit-receive pairs as we are aiming at the capacity scaling behavior with very large number of nodes rather than the exact capacity.

Given two clusters each of  $M$  nodes at distance of  $d_{st}$  meters apart, the capacity of a  $K_1 \times K_2$  MIMO channel between  $K_1 \in [1:M]$  nodes in the transmit cluster and  $K_2 \in [1:M]$  nodes in the receive cluster, assuming full cooperation among transmit/receive nodes, can be denoted as  $\mathcal{C}(K_1, K_2, \gamma)$  [bits per channel use]. Without specifying the expression of the channel capacity  $\mathcal{C}(K_1, K_2, \gamma)$ , which is channel assumption dependent, the following fact will be used in our analysis: given any fixed  $K_1$  (resp.  $K_2$ )  $\in [1:M]$ , the capacity  $\mathcal{C}(K_1, K_2, \gamma)$  is monotonically increasing and concave w.r.t.  $K_2$  (resp.  $K_1$ ) if we ignore the continuity constraint.

The hierarchical cooperation protocol proposed in [3] rests on a layered structure, where the local cooperation tasks of Stage I and Stage III of an upper layer in the hierarchy are treated as the communication problem to be solved by the lower layer, which itself also consists of three stages partitioned in the same fashion. Such recursion is applied

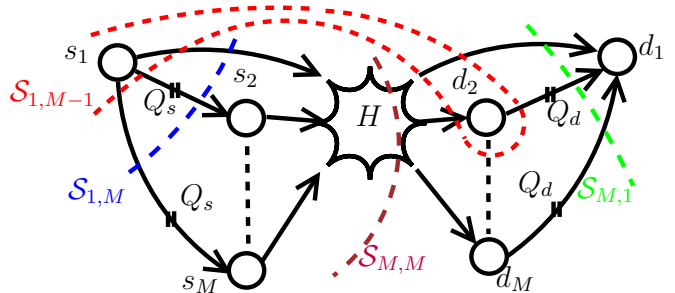


Fig. 2. A cut  $\mathcal{S}_{K_1, K_2}$  applied to the virtual MIMO setup puts  $K_1$  transmit nodes (including the source node  $s_1$ ) on the left side of the cut and  $K_2$  receive nodes (including the destination node  $d_1$ ) on the right side.

repeatedly until the communication task for the lower layer becomes trivial or can be solved by local transmission. Our analysis is bottom-up, from the lowest layer (only incurs local transmission) to the highest layer (all nodes in the network are included). Note that at each layer, the local communication in Stage I (and Stage III) are carried out concurrently in all clusters. In this paper we artificially ignore the inter-cluster interference despite the fact that we do not resort to spatial reuse. Such “ignorance” of interference rests on two key observations: the power of inter-cluster interference can be limited to a low level via power control as stated in [3]; the penalty of spatial reuse can be greatly reduced by intelligent scheduling mechanism [4], [5]. Therefore we do not expect our results will change drastically should the residual interference (after power control) be accounted into analysis.

## III. RATE UPPER BOUNDS OF A VIRTUAL MIMO SESSION

For each source-destination pair, say  $s_1 \rightarrow d_1$ , the data transmission takes three steps:  $s_1$  first transmits  $Q_s$  bits each to its  $M-1$  neighbors via orthogonal channels;  $M \times M$  virtual MIMO transmission is performed for a duration of one time slot (i.e., one channel use); each node in the receiving cluster forward  $Q_d$  bits to the destination node  $d_1$  via orthogonal channels. Therefore, the amount of data that transmits from  $s_1$  to  $d_1$  can be upper bounded by the cut-set bounds illustrated in Fig. 2. A cut for the virtual MIMO setup may put  $K_1$  transmit nodes (including  $s_1$ ) on the left side of the cut and  $K_2$  receive nodes (including  $d_1$ ) on the right side. Since we do not distinguish a node from its neighbors, it is the number of nodes rather than their specific identity that matters. All cuts that have  $K_1$  transmit nodes on its left side and  $K_2$  receive nodes on its right side will be denoted as cut  $\mathcal{S}_{K_1, K_2}$  as they will incur the same upper bound. For  $K_1, K_2 = 1, \dots, M$ , the cut  $\mathcal{S}_{K_1, K_2}$  incurs a constraint from the data transmission via the  $K_1 \times K_2$  MIMO channel,  $(M-K_1)$  constraints of  $Q_s$  bits each from the orthogonal channels originated from  $s_1$ , and  $(M-K_2)$  constraint of  $Q_d$  bits each from the orthogonal channels ended at  $d_1$ . The corresponding upper bound on the transmitted bits  $D$  from  $s_1$  to  $d_1$  can be written as

$$D \leq 1 * \mathcal{C}(K_1, K_2, \gamma) + (M-K_1)Q_s + (M-K_2)Q_d \text{ [bits]}.$$

Since  $\mathcal{C}(K_1, K_2, \gamma)$  is monotonically increasing and concave, the minimum value of the above combined function is

<sup>2</sup>The SNR  $\gamma$  depends on  $(M, d_{st})$ , which is not highlighted here.

obtained at its boundaries, which leads to the following cut-set bound (after applying all the cuts)

$$D = \min\{\mathcal{C}(M, M, \gamma), \mathcal{C}(1, 1, \gamma) + (M-1)Q_s + (M-1)Q_d, \mathcal{C}(1, M, \gamma) + (M-1)Q_s, \mathcal{C}(M, 1, \gamma) + (M-1)Q_d\}, \quad (2)$$

Furthermore, with per-antenna power constraint (in contrast to the sum power constraint), we have

$$\begin{aligned} \mathcal{C}(M, 1, \gamma) - \mathcal{C}(1, 1, \gamma) &\leq \mathcal{C}(M, M, \gamma) - \mathcal{C}(1, M, \gamma), \\ \mathcal{C}(1, M, \gamma) - \mathcal{C}(1, 1, \gamma) &\leq \mathcal{C}(M, M, \gamma) - \mathcal{C}(M, 1, \gamma), \end{aligned} \quad (3)$$

where the inequality is due to the fact that, when channel knowledge is known at the receiver, larger capacity gain can be obtained by increasing the degree-of-freedom than by increasing the diversity order. To maximize the cut-set bound in (2), we choose  $Q_s$  and  $Q_d$  such that

$$\begin{aligned} Q_s &\geq \frac{\mathcal{C}(M, M, \gamma) - \mathcal{C}(1, M, \gamma)}{M-1}, \\ Q_d &\geq \frac{\mathcal{C}(M, M, \gamma) - \mathcal{C}(M, 1, \gamma)}{M-1}, \end{aligned} \quad (4)$$

which, together with the condition (3), implies that

$$\mathcal{C}(1, 1, \gamma) + (M-1)Q_s + (M-1)Q_d \geq \mathcal{C}(M, M, \gamma).$$

Therefore, the cut-set bound (2) can now be written as

$$D \leq \mathcal{C}(M, M, \gamma). \quad (5)$$

If the channels are i.i.d. Rayleigh fading with  $\mathbf{H} = [h_{i,j}]$ ,  $h_{i,j} \sim \mathcal{CN}(0, 1)$ , and the receive nodes have perfect channel state information, the capacity  $\mathcal{C}(M, M, \gamma)$  can be written as

$$\begin{aligned} \mathcal{C}(M, M, \gamma) &= \mathbb{E}[\log \det(\mathbf{I}_M + \gamma \mathbf{H} \mathbf{H}^\dagger)] \\ &= \mathbb{E} \left[ \sum_{i=1}^M \log(1 + M\gamma \lambda_i^2) \right] \simeq M \int \log(1 + M\gamma x) f(x) dx, \end{aligned} \quad (6)$$

where the first equality is from the MIMO capacity with per-antenna power constraint, the second step is by rewriting the capacity with singular values of the random matrix  $\frac{1}{\sqrt{M}} \mathbf{H}$ , and the third step is due to that, for large  $M$ , the empirical distribution of  $\lambda$  converges to a limiting density function  $f(x)$  [12, (8.23)] as prescribed by the *Random Matrix Theory*. For both  $M\gamma \gg 1$  (e.g., dense networks) and  $M\gamma \ll 1$  (e.g., extended networks), we have the following approximation

$$\mathcal{C}(M, M, \gamma) \simeq M \log(1 + M\gamma).$$

When  $M$  is large, we can write  $\mathcal{C}(1, M, \gamma)$  and  $\mathcal{C}(M, 1, \gamma)$  as

$$\mathcal{C}(M, 1, \gamma) = \mathbb{E}[\log(1 + \gamma \sum_i |h_{i,j}|^2)] \simeq \log(1 + M\gamma), \quad (7)$$

$$\mathcal{C}(1, M, \gamma) = \mathbb{E}[\log(1 + \gamma \sum_j |h_{i,j}|^2)] \simeq \log(1 + M\gamma), \quad (8)$$

where the approximation is from the *Law of Large Numbers* for large  $M$ . Note that all the capacities  $\mathcal{C}(1, M, \gamma)$ ,  $\mathcal{C}(M, 1, \gamma)$ , and  $\mathcal{C}(M, M, \gamma)$  grow with  $M$  if the SNR  $\gamma$  does not change with  $M$ , then the requirement of local communication  $Q_s$  and  $Q_d$  as stated in (4) will not be satisfied as  $M$  scales up. If we instead allocate power such that  $M\gamma(d_{st}) = \gamma_0$  is a constant, the condition (4) always hold and we can rewrite the condition (for both  $\gamma_0 \gg 1$  and  $\gamma_0 \ll 1$ ) as

$$Q_s \gtrsim \log(1 + \gamma_0), \quad Q_d \gtrsim \log(1 + \gamma_0). \quad (9)$$

#### IV. AGGREGATE THROUGHPUT UNDER HIERARCHICAL COOPERATION PROTOCOL

By choosing power allocation such that  $M\gamma(d_{st}) = \gamma_0 > 0$  as discussed in Sec. III, the aggregate throughput  $\mathcal{C}(M, M, \gamma)$  in (6) can be written as  $M \int \log(1 + \gamma_0 x) f(x) dx = MR^*$ , where

$$R^* \triangleq \int \log(1 + \gamma_0 x) f(x) dx \leq \log(1 + \gamma_0).$$

In the high SNR regime we have  $R^* \simeq \log(1 + \gamma_0) - \log(e)$  and the requirement in (9) becomes

$$Q_s \gtrsim R^* + \log(e), \quad Q_d \gtrsim R^* + \log(e). \quad (10)$$

In the low SNR regime we have  $R^* \simeq \log(1 + \gamma_0)$  and the requirement in (9) becomes

$$Q_s \gtrsim R^*, \quad Q_d \gtrsim R^*. \quad (11)$$

##### A. How Much Time Is Required in the Hierarchical Protocol

Assume there exists a scheme that can support  $N_0$  source-destination pairs at an aggregate throughput  $C_0 = \frac{1}{S_0} N_0^{b_0}$  within a cluster of  $N_0$  nodes, where  $S_0 > 0$  and  $b_0 \geq 0$  are two parameters depending on the communication scheme.

1) *Layer  $\ell=1$* : Set the cluster size  $M_1 = N_0$ . Each source node within a cluster needs to transmit  $Q_s$  bits to the rest of  $(M_1 - 1)$  nodes within the cluster, resulting in  $M_1(M_1 - 1)$  transmission sessions each of  $Q_s$  bits. Using the transmission scheme from layer  $\ell=0$ , it takes

$$\frac{(M_1 - 1)M_1 Q_s}{C_0} = \frac{(M_1 - 1)M_1 Q_s}{N_0^{b_0}/S_0} \simeq S_0 Q_s M_1^{2-b_0}$$

time slots (channel use) to carry out the message exchange within each cluster. In Stage II there are  $N_1$  sessions, each corresponding to a  $M_1 \times M_1$  virtual MIMO transmission that lasts for one time slot. Therefore Stage II takes  $N_1$  time slots. In Stage III, a destination node receives  $Q_d$  bits from each of its  $M_1 - 1$  neighbors, and on average every node inside the cluster will be a destination node, which results in  $M_1(M_1 - 1)$  sessions each of  $Q_d$  bits. Using the transmission scheme from layer  $\ell=0$  it takes  $(M_1 - 1)M_1 Q_d / C_0 \simeq S_0 Q_d M_1^{2-b_0}$  time slots. The aggregate throughput at layer  $\ell=1$  is therefore

$$C_1 \simeq \frac{N_1 M_1 R^*}{N_1 + S_0(Q_s + Q_d) M_1^{2-b_0}} = \frac{R^* N_1^{\frac{1}{2-b_0}}}{1 + S_0(Q_s + Q_d)} = \frac{N_1^{b_1}}{S_1}, \quad (12)$$

where the second step is obtained by choosing  $N_1 = M_1^{2-b_0}$  and the last step is obtained by setting

$$b_1 = \frac{1}{2-b_0}, \quad S_1 = S_0 \left( \frac{Q_s + Q_d}{R^*} \right) + \frac{1}{R^*}. \quad (13)$$

2) *Layer  $\ell=2$* : Set  $M_2 = N_1$ . In Stage I each source node transmits  $Q_s$  bits to all the  $M_2 - 1$  neighboring nodes and every destination node receives  $Q_d$  bits from each of its  $M_2 - 1$  neighbors. The transmission scheme developed at layer  $\ell=1$ , whose rate is  $C_1$  as in (12), is used to accomplish the transmission tasks of Stage I & III. Hence Stage I takes  $M_2(M_2 - 1)Q_s / C_1 \simeq S_1 Q_s M_2^{2-b_1}$  time slots and Stage III takes about  $S_1 Q_d M_2^{2-b_1}$  slots, whereas Stage II takes  $N_2$  time

slots, one for each source node. The aggregate throughput at layer  $\ell=2$  is

$$C_2 \simeq \frac{N_2 M_2 R^*}{N_2 + S_1(Q_s + Q_d) M_2^{2-b_1}} = \frac{R^* N_2^{\frac{1}{2-b_1}}}{1 + S_1(Q_s + Q_d)} = \frac{N_2^{b_2}}{S_2},$$

where the second step is by choosing  $N_2 = M_2^{2-b_1}$ , and

$$b_2 = \frac{1}{2-b_1}, \quad S_2 = S_1 \left( \frac{Q_s + Q_d}{R^*} \right) + \frac{1}{R^*}. \quad (14)$$

3) *Layer  $\ell > 2$* : Assume at layer  $(\ell-1)$  we have

$$N_{\ell-1} = M_{\ell-1}^{2-b_{\ell-2}}, \quad C_{\ell-1} = \frac{1}{S_{\ell-1}} N_{\ell-1}^{b_{\ell-1}}, \quad (15)$$

$$b_{\ell-1} = \frac{1}{2-b_{\ell-2}}, \quad S_{\ell-1} = S_{\ell-2} \left( \frac{Q_s + Q_d}{R^*} \right) + \frac{1}{R^*}. \quad (16)$$

Setting  $M_\ell = N_{\ell-1}$ , applying the layer  $(\ell-1)$  scheme in the Stage I and Stage III of layer  $\ell$ , and resorting to the  $M_\ell \times M_\ell$  virtual MIMO transmission in Stage II, we have

$$C_\ell \simeq \frac{N_\ell M_\ell R^*}{N_\ell + S_{\ell-1}(Q_s + Q_d) M_\ell^{2-b_{\ell-1}}} = \frac{R^* N_\ell^{\frac{1}{2-b_{\ell-1}}}}{1 + S_{\ell-1}(Q_s + Q_d)} = \frac{N_\ell^{b_\ell}}{S_\ell},$$

where the second step is by choosing  $N_\ell = M_\ell^{2-b_{\ell-1}}$  and

$$b_\ell = \frac{1}{2-b_{\ell-1}}, \quad S_\ell = S_{\ell-1} \left( \frac{Q_s + Q_d}{R^*} \right) + \frac{1}{R^*}. \quad (17)$$

### B. Aggregate Throughput (Sum Capacity)

Given  $b_0 \in [0, 1)$  and  $S_0 > 0$ ,  $b_\ell$  and  $S_\ell$  can be determined explicitly for all  $\ell = 1, 2, 3, \dots$ , as follows

$$b_\ell = \frac{\ell - (\ell-1)b_0}{(\ell+1) - \ell b_0}, \quad (18)$$

$$S_\ell = \left( \frac{Q_s + Q_d}{R^*} \right)^\ell \left( S_0 + \frac{1}{Q_s + Q_d - R^*} \right) - \frac{1}{Q_s + Q_d - R^*}, \quad (19)$$

where  $b_\ell$  is determined by observation and induction, whereas  $S_\ell$  is determined by the fact that

$$S_\ell + \frac{1}{Q_s + Q_d - R^*} = \left( \frac{Q_s + Q_d}{R^*} \right) (S_{\ell-1} + \frac{1}{Q_s + Q_d - R^*}).$$

Therefore we have

$$N_\ell = M_\ell^{1/b_\ell} = N_{\ell-1}^{1/b_\ell} = N_0^{1/(\prod_{i=1}^\ell b_i)} = N_0^{(\ell+1) - \ell b_0}, \quad (20)$$

$$M_\ell = N_{\ell-1} = N_0^{\ell - (\ell-1)b_0}, \quad (21)$$

and the aggregate throughput (i.e., sum rate)  $C_\ell$  is given by

$$C_\ell = \frac{N_\ell^{b_\ell}}{S_\ell} = \frac{N_0^{\ell - (\ell-1)b_0}}{\left( \frac{Q_s + Q_d}{R^*} \right)^\ell \left( S_0 + \frac{1}{Q_s + Q_d - R^*} \right) - \frac{1}{Q_s + Q_d - R^*}}, \quad (22)$$

where the last step is obtained by applying (18), (19) and (20).

## V. CAPACITY SCALING UNDER VARIOUS CONSTRAINTS

### A. When the Total Number of Nodes $N$ is Unbounded

When the total number of nodes  $N$  is unbounded, we can choose  $N_0$  and  $\ell$  independently to scale up  $N$ .

1) *If we fix  $N_0$  and let  $\ell \rightarrow \infty$* : We can rewrite (22) as

$$C_\ell \simeq \frac{N_0^{b_0 + \ell(1-b_0) - \ell \log_{N_0} \left( \frac{Q_s + Q_d}{R^*} \right)}}{S_0 + \frac{1}{Q_s + Q_d - R^*}} = \frac{C_0 N_\ell^{\beta_\ell}}{1 + \frac{1}{S_0(Q_s + Q_d - R^*)}}, \quad (23)$$

where  $C_0 = N_0^{b_0}/S_0$  is the rate at which local message change can be performed at layer  $\ell=0$  as defined at the beginning of Sec. IV-A. The scaling factor  $\beta_\ell$  in (23) is given by

$$\beta_\ell = 1 - \frac{\log_{N_0} \left( \frac{Q_s + Q_d}{R^*} \right)}{1 - b_0} - \frac{1 - \frac{1}{(1-b_0)} \log_{N_0} \left( \frac{Q_s + Q_d}{R^*} \right)}{\ell(1-b_0) + 1}. \quad (24)$$

As  $\ell \rightarrow \infty$ , the scaling factor  $\beta_\ell$  increases monotonically to

$$\beta_\infty \triangleq 1 - \frac{\log_{N_0} \left( \frac{Q_s + Q_d}{R^*} \right)}{1 - b_0} = 1 - \log_{N_0^{1-b_0}} \left( \frac{Q_s + Q_d}{R^*} \right). \quad (25)$$

Since we have  $Q_s \geq R^*$  and  $Q_d \geq R^*$  as stated in (10) and (11), the penalty term on the scaling factor  $\beta_\infty$  is not negligible for  $b_0 \in [0, 1)$  with a fixed  $N_0$ .

2) *If we fix  $\ell$  and let  $N_0$  grows*: From [1] we know that  $C_0 \leq O(\sqrt{N_0})$  if the optimal multi-hop strategy is used for message exchange in the initial cluster. We can therefore rewrite (22) (setting  $b_0 = 1/2$  and  $S_0 = 1$ ) as

$$C_\ell \simeq \frac{N_0^{\ell - (\ell-1)/2}}{\left( \frac{Q_s + Q_d}{R^*} \right)^\ell \left( 1 + \frac{1}{Q_s + Q_d - R^*} \right)} = \frac{N_\ell^{\frac{\ell/2 + 1/2}{\ell/2 + 1}}}{\left( \frac{Q_s + Q_d}{R^*} \right)^\ell \left( 1 + \frac{1}{Q_s + Q_d - R^*} \right)} \simeq \left( \frac{Q_s + Q_d}{R^*} \right)^{-\ell} N_\ell^{1 - \frac{1}{\ell+2}}. \quad (26)$$

Here the cost of local communication is not significant as long as  $\ell$  is not too large.

3) *If both  $N_0$  and  $\ell$  grow*: Since  $\beta = \ln(C)/\ln(N)$ , where

$$N = N_0^{\ell(1-b_0)+1}, \quad C \simeq \left( \frac{Q_s + Q_d}{R^*} \right)^{-\ell} N_0^{\ell(1-b_0)+b_0},$$

grow with  $N_0$  and  $\ell$ , we take the partial derivatives of  $\beta$  w.r.t.  $N_0$  and  $\ell$ , relaxing the integer constraints, and set them to zero to determine the relationship between  $N_0$  and  $\ell$  as

$$N_0 = \left( \frac{Q_s + Q_d}{R^*} \right)^{[\ell(1-b_0)+1]/(1-b_0)^2}. \quad (27)$$

The maximized  $\beta$ , possible only if (27) is met, is

$$\beta = \frac{[\ell(1-b_0) + b_0]^2 + b_0(1-b_0)}{[\ell(1-b_0) + 1]^2}, \quad (28)$$

and the required total number of nodes  $N$  to sustain (28) is

$$N = \left( \frac{Q_s + Q_d}{R^*} \right)^{[\ell(1-b_0)+1]^2/(1-b_0)^2}. \quad (29)$$

### B. When $N$ is Large but Fixed

Substituting (20) into (22), we have

$$C \simeq \left( \frac{Q_s + Q_d}{R^*} \right)^{-\ell} N^{\frac{\ell - (\ell-1)b_0}{(\ell+1) - \ell b_0}} = \left( \frac{Q_s + Q_d}{R^*} \right)^{-\ell} N^{1 + \frac{b_0 - 1}{\ell(1-b_0) + 1}}, \quad (30)$$

which prescribes the optimal  $\ell$  (to maximize  $C$ ) as

$$\ell^* = \sqrt{\frac{\ln(N)}{\ln\left(\frac{Q_s + Q_d}{R^*}\right)}} - \frac{1}{1 - b_0}, \quad (31)$$

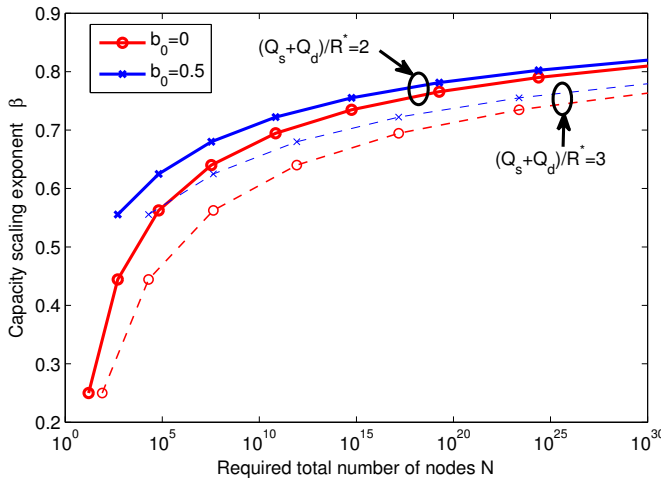


Fig. 3. The maximized capacity scaling exponent  $\beta$  of (28) as a function of the required total number of nodes  $N$  specified in (29).

and the corresponding optimal initial cluster size  $N_0^*$  as

$$\ln(N_0^*) = \frac{\ln(N)}{\ell^*(1-b_0) + 1} = \frac{\sqrt{\ln(N) \ln\left(\frac{Q_s+Q_d}{R^*}\right)}}{1-b_0}. \quad (32)$$

The optimal “scaling” exponent can now be derived as

$$\beta^* = \frac{\ln(C)}{\ln(N)} = 1 - 2\sqrt{\frac{\ln\left(\frac{Q_s+Q_d}{R^*}\right)}{\ln(N)}} + \frac{\ln\left(\frac{Q_s+Q_d}{R^*}\right)}{(1-b_0)\ln(N)}. \quad (33)$$

## VI. NUMERICAL ILLUSTRATION

To have a quantitative sense of how local communication cost may degrade the overall performance, in Fig. 3 we plot the maximized capacity scaling exponent  $\beta$  stated in (28) as a function of the required total number of nodes  $N$  specified in (29). Since the local communication cost must satisfy the constraints  $Q_s \geq R^*$  and  $Q_d \geq R^*$  to not hurt the virtual MIMO capacity, we plot for two cases with  $\frac{Q_s+Q_d}{R^*} = 2$  and  $\frac{Q_s+Q_d}{R^*} = 3$ . The former represents an idealized power control mechanism such that those constraints are always satisfied with equality, and the latter puts up some margin to tolerate rate mismatch. For the initial stage,  $b_0 = 0$  corresponds to the situation with TDMA type of message exchanging methods and  $b_0 = 0.5$  corresponds to the case where ideal multi-hop routing is utilized. From Fig. 3 we can see that targeting at a capacity scaling exponent  $\beta > \frac{3}{4}$  would require the total number of nodes  $N > 10^{14}$  even for the case with ideal local communication cost  $\frac{Q_s+Q_d}{R^*} = 2$ .

In Fig. 4 we plot optimal capacity scaling exponent  $\beta^*$  specified in (33) as a function of the total number of nodes  $N$ , where we maximize the aggregate throughput  $C_N$  for each fixed  $N$  and ignore the integer constraint on  $\ell$  and  $N_0$ . From Fig. 4 we can see that better message exchanging methods at the initial stage (i.e., larger  $b_0$ ) helps to improve the sum rate  $C_N$  and hence also the scaling exponent  $\beta^*$  (blue vs. red), and more than “necessary” local communication cost  $Q_s + Q_d$  severely degrades the overall performance. Since the idealized case  $\frac{Q_s+Q_d}{R^*} = 2$  may require sophisticated (if not impossible) power control at every node on each stage, appropriate tradeoff

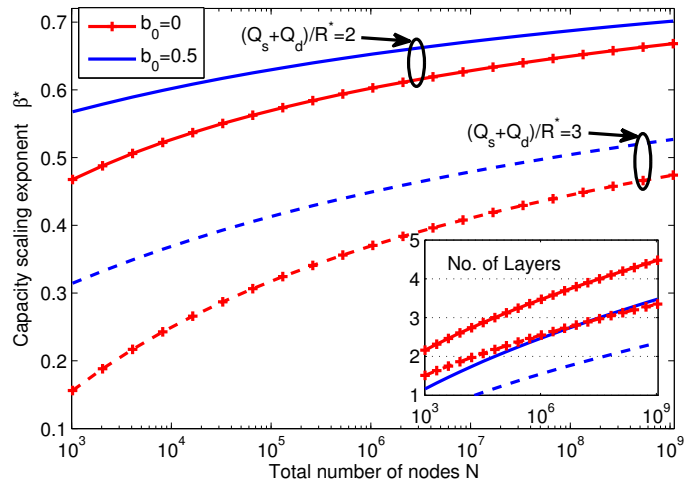


Fig. 4. The optimal capacity scaling exponent  $\beta^*$  of (33) with the aggregate throughput  $C_N$  maximized at each fixed  $N$ , ignoring the integer constraint on  $\ell$  and  $N_0$ . More than necessary local communication cost  $Q_s + Q_d > 2R^*$  severely degrades the sum rate  $C_N$  hence also the scaling exponent  $\beta^*$ . Note that the number of hierarchical layers  $\ell^*$  of (31), at which the sum rate is maximized, is less than 5 even for networks consisting of  $N = 10^9$  nodes.

between the local communication margin and power control complexity is to be investigated. Note that for the range of network size we have investigated,  $10^3 \leq N \leq 10^9$  as in Fig. 4, the number of hierarchical layers  $\ell^*$  at which the sum rate is maximized is very small (less than 5 even for  $N = 10^9$ ).

## REFERENCES

- [1] P. Gupta and P. R. Kumar, “The capacity of wireless networks,” *IEEE Transactions on Information Theory*, vol. 46, pp. 388–404, Mar. 2000.
- [2] 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 Transactions on Information Theory*, vol. 52, pp. 2313–2328, Jun. 2006.
- [3] A. Özgür, O. Lévêque, and D. N. C. Tse, “Hierarchical cooperation achieves optimal capacity scaling in Ad Hoc networks,” *IEEE Transactions on Information Theory*, vol. 53, pp. 3549–3572, Oct. 2007.
- [4] J. Ghaderi, L.-L. Xie, and X. Shen, “Hierarchical cooperation in ad hoc networks: optimal clustering and achievable throughput,” *IEEE Transactions on Information Theory*, vol. 55, pp. 3425–3436, Aug. 2009.
- [5] S.-N. Hong and G. Caire, “On the performance of optimized dense device-to-device wireless networks,” *arXiv:1402.1815v2*, Jun. 2015.
- [6] M. Franceschetti, M. D. Migliore, and P. Minero, “The capacity of wireless networks: Information-theoretic and physical limits,” *IEEE Transactions on Information Theory*, vol. 55, pp. 3413–3424, Aug. 2009.
- [7] S.-H. Lee and S.-Y. Chung, “Capacity scaling of wireless ad hoc networks: Shannon meets Maxwell,” *IEEE Transactions on Information Theory*, vol. 58, pp. 1702–1715, Mar. 2012.
- [8] A. Özgür, O. Lévêque, and D. N. C. Tse, “Spatial degrees of freedom of large distributed MIMO systems and wireless ad hoc networks,” *IEEE Journal on Selected Areas in Communications*, vol. 31, pp. 202–214, Feb. 2013.
- [9] N. Fawaz and M. Médard, “A Converse for the Wideband Relay Channel with Physically Degraded Broadcast,” in *Proceedings IEEE Information Theory Workshop (ITW)*, Oct. 2011.
- [10] M. K. Karakayali, G. J. Foschini, and R. A. Valenzuela, “Network coordination for spectrally efficient communications in cellular systems,” *IEEE Wireless Communications*, vol. 13, pp. 56–61, Aug. 2006.
- [11] D. Gesbert, S. Hanly, H. Huang, S. Shamai, O. Simeone, and W. Yu, “Multi-cell MIMO cooperative networks: A new look at interference,” *IEEE Journal on Selected Areas in Communications (JSAC)*, vol. 28, pp. 1380–1408, Dec. 2010.
- [12] D. N. C. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, 2005.