Hierarchical Cooperation in *Ad Hoc* Networks: Optimal Clustering and Achievable Throughput

Javad Ghaderi, *Student Member, IEEE*, Liang-Liang Xie, *Senior Member, IEEE*, and Xuemin (Sherman) Shen, *Fellow, IEEE*

***Abstract*—For a wireless network with n nodes distributed in an area A, and with n source–destination pairs communicating with each other at some common rate, the hierarchical cooperation scheme proposed in (Ozgur, Leveque, and Tse, 2007) is analyzed and optimized by choosing the number of hierarchical stages and the corresponding cluster sizes that maximize the total throughput. It turns out that increasing the number of stages does not necessarily improve the throughput, and the closed-form solutions for the optimization problem can be explicitly obtained. Based on the expression of the maximum achievable throughput, it is found that the hierarchical scheme achieves a scaling with the exponent depending on n. In addition, to apply the hierarchical cooperation scheme to random networks, a clustering algorithm is developed, which divides the whole network into quadrilateral clusters, each with exactly the number of nodes required.**

***Index Terms*—Ad hoc networks, hierarchical cooperation, optimal clustering, scaling laws, wireless networks.**

I. Introduction

WIRELESS networks formed by radio nodes are a subject of much topical interest, and they are found in various applications such as *ad hoc* networks, mesh networks, sensor networks, etc. For the optimal design and operation of such networks, it is of fundamental importance to determine the information-theoretic capacity of such networks, which, however, is a formidable task, since even for the simple three-node scenario [2], the exact capacity is still undetermined after several decades' effort.

Although the exact capacity is extremely difficult to determine, a lot of insightful upper and lower bounds on the capacity of large wireless networks have been obtained in recent years, e.g., [3]–[11]. The seminal work [3] initiated the study of scaling laws, and discovered a throughput scaling of $\Theta(\sqrt{n})$ under several communication models. Subsequently, a purely information-theoretic approach without any restrictions on the communication schemes was taken in [4], where a more fundamental connection between the total network transmit power and the transport capacity was discovered. As a consequence, when fixing the minimum separation distance and letting the number of nodes increase, the scaling law of $\Theta(\sqrt{n})$ was confirmed in the high signal attenuation regime. However, when the signal attenuation was low, higher scaling laws were shown to be possible for some special relay networks.

Therefore, an interesting question was raised as to what exactly the scaling laws are in the low signal attenuation regime. By incorporating long-range multiple-input multiple-output (MIMO) communications with local cooperations as proposed in [11], a recent work [1] developed a hierarchical architecture which was able to continually increase the scaling by adding more hierarchical stages. Specifically, for a network model where all the nodes are confined in a unit area but still with the far-field signal attenuation, the scaling with h hierarchical stages was claimed to be $\Theta(n^{\frac{h}{h+1}})$. Thus, by letting $h \to \infty$, any scaling of $\Theta(n^{1-\epsilon})$ is achievable, where $\epsilon > 0$ can be arbitrarily small. However, there is a fundamentally important issue that needed to be addressed, i.e., the pre-constant of the scaling. The pre-constants of the scalings for different h are different, and they are not even lower bounded from zero.

In this paper, we will show that the complete expression for the scaling with h hierarchical stages should be $c(h)n^{\frac{h}{h+1}}$, where, the pre-constant $c(h)$ not only depends on h, but also tends to zero as h goes to infinity. Since the preconstant affects the scaling behavior, we will present what can be achieved with the hierarchical scheme by providing an explicit expression of the pre-constant. It will become obvious that adding more stages does not necessarily increase the achievable rate for any fixed n. Actually, for each n, the optimal number of stages to choose is $\sqrt{\log_\beta(n/2)}$, where β is a constant to be defined later, and the corresponding maximum achievable throughput is

$$\frac{\beta R}{\sqrt{\log_\beta(n/2)}}(n/2)^{1-\frac{2}{\sqrt{\log_\beta(n/2)}}} \tag{1}$$

where R is another constant. Therefore, as shown in (1), the hierarchical scheme actually achieves a scaling with the exponent depending on n. Although the exponent converges to 1 as n increases, the convergence is not fast enough, and as a consequence, we will show that the average rate per source–destination pair tends to zero.

Generally, a network with area A is distinguished into two categories based on whether $A^{\alpha/2} \leq n$, where $\alpha \geq 2$ is the power path loss exponent. In the case where $A^{\alpha/2} \leq n$, (1) is

Manuscript received February 03, 2008; revised January 04, 2009. Current version published July 15, 2009. This work was partially supported by the NSERC of Canada. The material in this paper was presented in part at the IEEE International Conference on Communications, Beijing, China, May 2008.

J. Ghaderi was with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada. He is now with the Coordinated Science Laboratory, University of Illinois, Urbana, 61801 USA.

L.-L. Xie and X. Shen are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada.

Communicated by R. A. Berry, Associate Editor for Communication Networks.

Color versions of Figs. 3, 5, and 8 in this paper is available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2009.2023681

achievable. In the other case where $A^{\alpha/2} > n$, (1) has to be multiplied by $n/A^{\alpha/2}$ in order to meet the power constraint. It is worth pointing out that the results in this paper such as (1) apply to finite n. When trying to draw conclusions on scaling laws by taking $n \to \infty$, however, it should be noted that the results for the first case cannot remain valid if $\alpha > 2$, since the far-field model would fail to apply after some point.

For clarity, we will first present the results for regular networks. Then the extension to random networks is trivial after we introduce a clustering algorithm that divides the whole network into quadrilateral clusters, each with exactly the number of nodes required for carrying out the hierarchical cooperation scheme. This clustering algorithm is another contribution of this paper.

The remainder of the paper is organized as follows. In Section II, the wireless network model is described. Section III is devoted to the hierarchical cooperation scheme in regular networks, where we present the optimal throughput-delay results for the scheme with different stages. In Section IV, a clustering algorithm is developed to extend the results to general random networks. Some concluding remarks are presented in Section V.

II. Wireless Network Model

Consider the following standard additive white Gaussian noise channel model of wireless networks.

1) There are a set of n nodes $\mathcal{N}$ located on a plane.
2) Each node uses a common average power P to transmit.
3) At any time t, each node i transmits the signal $X_i(t) \in \mathbb{C}$, and receives the signal $Y_i(t) \in \mathbb{C}$. The received signal depends on the transmitted signals of all the other nodes as

$$Y_i(t) = \sum_{k \neq i} H_{ik}(t) X_k(t) + Z_i(t) \tag{2}$$

where $Z_i(t)$ is white circularly symmetric Gaussian noise of variance N, and the gain

$$H_{ik}(t) = \sqrt{G} r_{ik}^{-\alpha/2} \exp(j\theta_{ik}(t)) \tag{3}$$

where r_{ik} is the distance between nodes i and k, and $\theta_{ik}(t)$ is the is the random phase uniformly distributed in $[0, 2\pi]$. The phase varies with time according to a stationary ergodic random process $\{\theta_{ik}(t)\}$. Moreover, the random processes $\{\{\theta_{ij}(t)\}; i, j \in \mathcal{N}, i \neq j\}$ are independent and identically distributed (i.i.d.) across i and j. The parameter $\alpha \geq 2$ is the power path-loss exponent, and G is a constant depending on transmitter and receiver antenna gains and carrier wavelength λ_c (see [1]).

Note that the channel model (3) is based on *far-field* assumption [16]. Let r_f denote the far-field distance of a transmitter antenna. r_f is defined as

$$r_f = 2D^2/\lambda_c$$

where D is the largest physical linear dimension of the antenna and λ_c is the carrier wavelength. Moreover, r_f should satisfy

$$r_f \gg D \text{ and } r_f \gg \lambda_c.$$

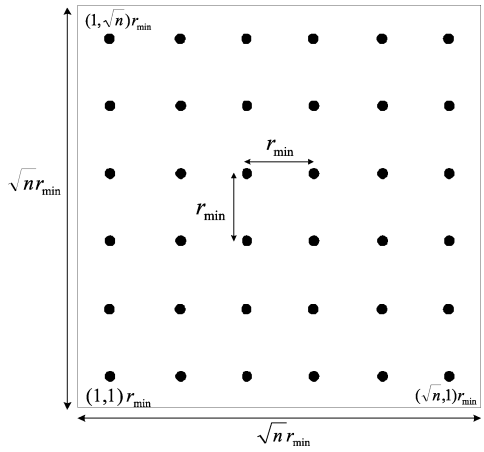

Fig. 1. A regular network with n nodes and a minimum distance $r_{\min}$.

This imposes the following constraint on the minimum separation distance $r_{\min}$ between nodes

$$r_{\min} \geq r_f. \tag{4}$$

We have basically used the channel model of [1] and refer the reader to [1] for more discussions in Section VI there.

Consider the problem of n source–destination pairs in the network, where each node is a source, with its destination node arbitrarily chosen from the other nodes. For simplicity, assume that each node chooses a different node as its destination, although this requirement can be relaxed to some extend as we can see from the coding strategy described later. Therefore, each node is a source and also a destination for another source. We only consider the case where all pairs communicate at the same rate.

For the simplicity of presentation, and in order to expose the key features of the coding strategy, we will first consider a regular network as depicted in Fig. 1, where nodes are located at the grid points $(xr_{\min}, yr_{\min})$ for $1 \leq x, y \leq \sqrt{n}$ in an area $A = nr_{\min}^2$. Then the results can be easily extended to general random networks with high probability, where n nodes are randomly and uniformly distributed inside a square of area A.

III. Hierarchical Cooperation in Regular Networks

A. Double-Stage Cooperation Scheme

As a prelude, consider only two stages for the scheme and assume $A = 1$ unit. We will follow [1], but show what is really achievable by presenting a more transparent description. Divide the regular network into clusters of size M nodes (See Fig. 2). The *double stage scheme* is based on local transmit and receive cooperation in clusters and MIMO transmissions between clusters. Consider one source node s and its destination node d. The goal of s is to send M subblocks of length L bits (in overal, ML bits) to d.

Let these bits be arranged in a *data matrix* $B \in \{0,1\}^{M \times L}$

$$B = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1L} \\ b_{21} & b_{22} & \dots & b_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ b_{M1} & b_{M2} & \dots & b_{ML} \end{pmatrix}$$

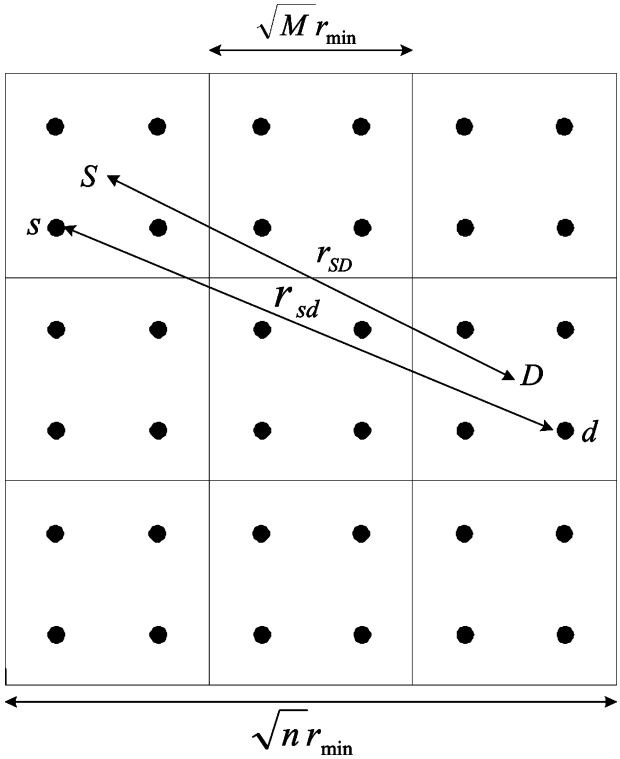

Fig. 2. Dividing the network of n nodes into clusters of size M nodes: a source node s and its cluster S and the destination d within cluster D have been depicted.

which corresponds to choose one message W from 2^{LM} possible messages $\{1, \ldots, 2^{LM}\}$. Denote the ith row by B^i (ith *data subblock*) and the mth column by B_m (mth *data vector*). The node s sends its data matrix to the node d in three steps:

1) s distributes its subblocks among the M nodes in its cluster by using TDMA. For this purpose, for each node k in the source cluster, s encodes the data subblock B^k to a codeword of length C_0 chosen from a randomly generated Gaussian codebook $\mathcal{C}_0 \overset{\text{i.i.d.}}{\sim} \mathcal{N}_{\mathbb{C}}(0, \sigma_0^2)$ where $\sigma_0^2 = Pr_{sk}^{\alpha}$. Sending one subblock requires C_0 time slots and distributing all subblocks needs MC_0 time slots. At the end, each node in the cluster obtains one data subblock of s.
2) The nodes of the source cluster form a distributed array antenna and send the LM bits of information to the destination cluster by MIMO transmissions. To accomplish this step, each node i encodes its subblock B^i to a codeword $X_i^{C_1} = (X_{i1}X_{i2}\cdots X_{iC_1})$ of C_1 symbols by using a randomly generated Gaussian codebook $\mathcal{C}_1 \overset{\text{i.i.d.}}{\sim} \mathcal{N}_{\mathbb{C}}(0, \sigma_1^2)$ where $\sigma_1^2 = P\frac{r_{SD}^{\alpha}}{M}$ and r_{SD} is the distance between the centers of two clusters. Then nodes of the source cluster send their codewords simultaneously to the destination cluster. Therefore this step needs C_1 time slots to complete. Each node k in destination cluster receives an observation Y_{kt} from the MIMO transmission at time t for $1 \le t \le C_1$ according to (2) or the following vector form:

$$Y^M(t) = H(t)X^M(t) + Z(t) \tag{5}$$

where $X^M(t) = (X_{1t}X_{2t}\cdots X_{Mt})^T$ and $Y^M(t) = (Y_{1t}Y_{2t}\cdots Y_{Mt})^T$ is the *observation vector* at time t. $Z(t) = (Z_{1t}Z_{2t}\cdots Z_{Mt})^T$ is uncorrelated noise at the receiver nodes, and $H_{ik}(t)$ are given by (3). The nodes simply store their observations. At the end of this step, each node k in destination cluster has accumulated an *observation subblock* $Y_k^{C_1} = (Y_{k1}Y_{k2}\cdots Y_{kC_1})$ of C_1 observations.
3) Each node k in the destination cluster quantizes its observations with Q bits per observation to obtain a *quantized observation subblock* $V_k^{C_1Q} = (V_{k1}V_{k2}\cdots V_{k(C_1Q)})$ of length C_1Q bits. From now on, the step is similar to step 1 but in reverse order. The cluster nodes send their quantized observation subblocks V_k to d by using the codewords of length C_0C_1Q/L chosen from a randomly generated Gaussian codebook with power σ_0^2 where $\sigma_0^2 = Pr_{kd}^{\alpha}$. The destination d can decode the quantized observations and estimate the observation subblocks and consequently, the observation vector $Y^M(t)$ by an *estimated observation vector* $\widehat{Y}^M(t)$. Then d can decode the transmitted data vectors $\widehat{B}_t$. The required number of slots for this step is MC_0C_1Q/L.

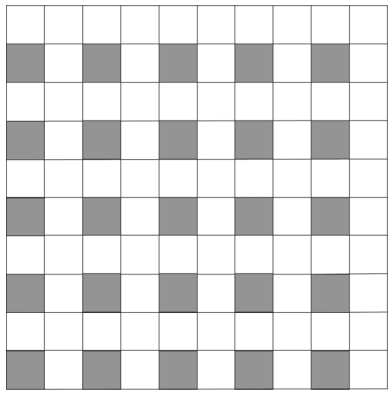

Fig. 3. Parallel operating clusters according to 4-TDMA: Gray clusters are active.

In the double stage cooperation strategy, the power of each observation must be upper bounded independent of cluster size, which leads to quantization with a fixed number of bits for an average distortion Δ^2. When two clusters are neighbor, using the power assignment of $\sigma^2 = Pr_{SD}^{\alpha}/M$ yields an unbounded received power when the cluster size increases. A simple solution is to divide these clusters into two equal halves, each with $M/2$ nodes. The source node s distributes its subblocks among $M/2$ nodes of the half located farther to the border. Then these $M/2$ nodes form a distributed antenna and perform MIMO between the halves located farther away. Now, the required time for the step 2) is twice the time needed for disjoint clusters, i.e., the required time is $2C_1$ slots. In step 3), $M/2$ nodes take part in delivering the observations to the destination. For source and destination nodes located in the same cluster, we can simply ignore the second step. According to Lemma 4.5 of [1], the power received by each node in destination cluster P^{rec} in the step 2 is lower and upper bounded independent of cluster size such that

$$GPa^2 + N \le P^{\text{rec}} \le GPb^2 + N \tag{6}$$

where

$$a = (2 - \sqrt{2})^{\alpha/2} \tag{7}$$

$$b = (2 + \sqrt{2})^{\alpha/2}. \tag{8}$$

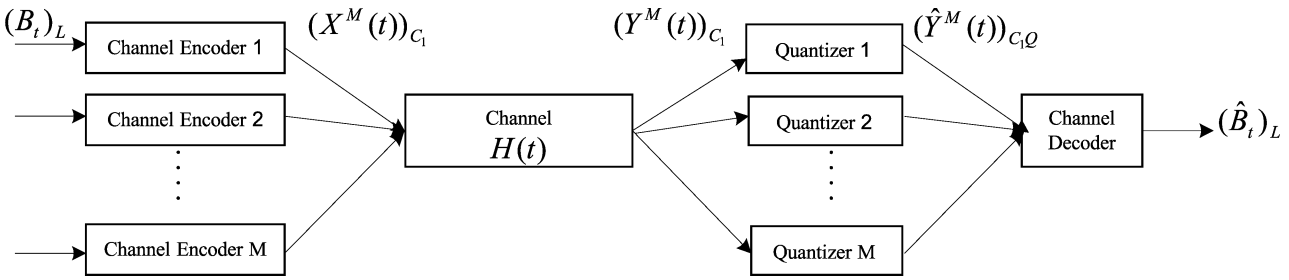

Fig. 4. A model for the quantized MIMO channel.

Each source–destination pair must accomplish the three steps. Clustering also enables spatial reuse in the sense that clusters can work in parallel for local cooperations (step 1 and step 3) provided they locate far enough from each other. This leads to three phases in the operation of the network:

Phase 1: Setting Up Transmit Cooperation: Clusters work in parallel according to the 4-TDMA scheme in Fig. 3 (as opposed to 9-TDMA scheme in [1][1]) where each cluster is active a fraction $1/4$ of the total time of this phase.

When a cluster becomes active, its source nodes must perform the first step, i.e., distributing their subblocks to the other nodes of the cluster by a simple time-division multiple-access (TDMA). Each source node needs MC_0 slots, hence the required time for source nodes of one cluster to exchange their bits is at most M^2C_0 slots and due to 4-TDMA, the whole phase needs $4M^2C_0$ slots to complete. Each node transmits with power σ_0^2 in at most fraction $\frac{1}{4M}$ of the total time of the phase. It can be shown that this power assignment satisfies an overall average power consumption less than P/n. Using the 4-TDMA ensures us that the interference power P_I each node received from all simultaneously transmitting nodes is bounded according to the following Lemma.[2]

Lemma 3.1: The interference signals received by different nodes, due to parallel operating clusters using 4-TDMA, are independent and for $\alpha > 2$ the interference power P_I that each node is received is given by

$$P_I \leq GP\left(8 + \frac{2}{\alpha - 2} + \frac{2}{\alpha - 1}\right).$$

Phase 2: MIMO Transmissions: We perform successive MIMO transmissions according to the step 2, one MIMO for each source–destination pair from source cluster to destination cluster in one time slot, hence we need at most $2nC_1$ slots. Each node encodes the subblocks by using a Gaussian code of power σ_1^2 as defined earlier. Since at most M MIMO transmissions are originated from each cluster, each node is active at most a fraction M/n of the total time of this phase and remains silent during the rest of the phase which yields an average power consumption less than P/n.

Phase 3: Cooperate to Decode: After the first two phases, each source–destination pair has completed the steps 1 and 2. Each cluster should accomplish the step 3 by conveying the quantized observations to the corresponding destination nodes located in the cluster. This phase is identical to the first phase, except that each node has C_1Q bits to transmit to each node in the same cluster instead of L bits. Therefore, this phase needs $4M^2C_0C_1Q/L$ slots to complete.

In summary, the required time D_2 for the double-stage scheme is

$$\begin{aligned} D_2 &= D(\text{phase } 1) + D(\text{phase } 2) + D(\text{phase } 3) \\ &= 4M^2C_0 + 2nC_1 + 4M^2C_0C_1Q/L \\ &= 4M^2C_0(1 + C_1Q/L) + 2nC_1. \end{aligned}$$

Assume the channel gains are known at all nodes. All communication links in the first phase can operate at any rate less than the following:

$$R_0 \leq \log\left(1 + \frac{GP}{P_I + N}\right). \tag{9}$$

Communications in the second phase are performed over the quantized MIMO channel of Fig. 4 where the notation $(\cdots)_L$ is used for an i.i.d. sequence of L random variables. The following lemma asserts that a spatial multiplexing gain of M is achievable for this channel.

Lemma 3.2: Define the average probability of error for the quantized MIMO channel by

$$P_e^L = \frac{1}{2^{LM}} \sum_{k=1}^{2^{LM}} \mathbb{P}(\widehat{W} \neq k | W = k)$$

then there exists a strategy to quantize the observations with Q bits per observation and a codebook $\mathcal{C}_1$ satisfying power constraint $\sigma_1^2 = Pr_{SD}^{\alpha}/M$ to encode the data subblocks such that arbitrary low P_e^L is feasible. Moreover, the minimum quantization rate Q and the maximum achievable rate R_1 of the codebook satisfy

$$Q > \log\left(1 + \frac{GPb^2 + N}{\Delta^2}\right)$$

and

$$\begin{aligned} \log\left(1 + \tau\frac{GP}{N + \Delta^2}\right)\frac{(a^2 - \tau)^2}{2b^4} \\ \leq R_1 < \log\left(1 + b^2\frac{GP}{N + \Delta^2}\right) \end{aligned} \tag{10}$$

for any $0 \leq \tau \leq a^2$.

[1]4-TDMA actually saves time compared to 9-TDMA. However, the scaling won't be changed.

[2]The analysis for $\alpha = 2$ is basically similar, but the results are different, e.g., the interference is bounded by a multiple of $\log n$ (see [1]). We ignore this singular case, and, for simplicity, assume $\alpha \in (2, \infty)$.

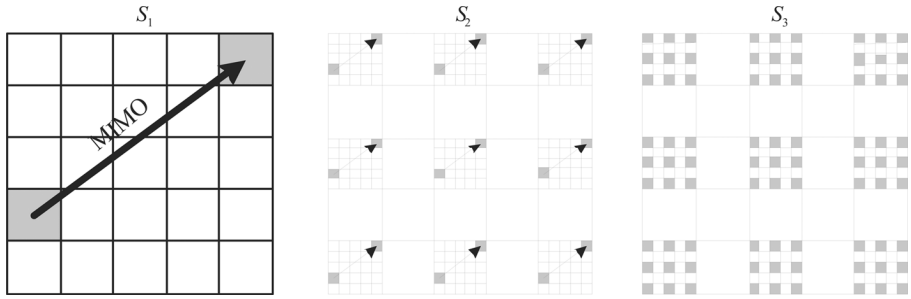

Fig. 5. The three stages of the triple-stage cooperation scheme.

Recall that a and b are two constants given respectively by (7) and (8), and Δ^2 is the average distortion. Note that there is a trade-off in choosing Δ^2. Although Δ^2 needs to be reduced for maximizing R_1, a small Δ^2 leads to a high quantization rate Q, and vice versa. Also note that in (10), it is reasonable to choose τ to maximize the lower bound. In the sequel, we consider fixed values for Δ^2 and τ to get nonzero and fixed rates R_1 and Q. In addition, for simplicity, all nodes use the same rate for their codewords $\mathcal{C}_0$ and $\mathcal{C}_1$, i.e., $\frac{L}{C_0} = \frac{L}{C_1} = R$, where

$$R \leq \min\{\text{LHS of (10), RHS of (9)}\}.$$

Hence, the required time D_2 can be written as

$$D_2(n, M, L) = 4M^2 \frac{L}{R}\left(1 + \frac{Q}{R}\right) + 2n\frac{L}{R}.$$

We call this quantity delay because each destination can decode its intended bits only after receiving all the corresponding observations, i.e., after the step 3. At the end of this time, each node has delivered ML bits to its destination which yields a total throughput of

$$T_2 = \frac{nML}{D_2}$$

which is maximized by choosing $M = \sqrt{\frac{R}{2(R+Q)}}\sqrt{n}$

$$T_2^{\text{opt}}(n) = \frac{R}{4\sqrt{2(1+Q/R)}} n^{1/2} \tag{11}$$

and the corresponding delay is

$$D_2^{\text{opt}}(n, L) = 4\frac{L}{R}n. \tag{12}$$

Obviously, by repeating n times, the double stage scheme can also be used for the problem where each node needs to send different information to all the other nodes. The achievable rate is as the following.

Lemma 3.3: For a regular network of size n, by the double-stage cooperation scheme with clusters of size M, each node can deliver ML different bits to each of the other nodes in a time block of

$$nD_2(n, M, L) = 4nM^2 \frac{L}{R}\left(1 + \frac{Q}{R}\right) + 2n^2 \frac{L}{R}.$$

Remark 3.1: Note that L denotes the number of bits to be transmitted in a basic time block, and is proportional to the block length for any fixed communication rate. Although for the interest of delay, it is better to choose smaller L as shown in Lemma 3.3, shorter block length leads to higher decoding errors. Hence, there is always a minimum L required to ensure enough reliability.

B. Triple-Stage Cooperation Scheme

Is it possible to achieve a better throughput by local cooperation and MIMO transmissions? Recall that in Phase 1 and Phase 3 of the double stage scheme, TDMA was used in each cluster to deliver the bits. Since each cluster itself is a network similar to the original network only with a smaller number of nodes, this implies that one can use the double stage scheme in each cluster to exchange the bits as well. Next, we analyze the throughput and delay of this new *triple stage scheme* when the double stage scheme is used in Phase 1 and Phase 3.

First, divide the whole network into clusters of size M_1, and then divide each cluster of size M_1 into sub-clusters of size M_2. Apply the double stage scheme to each cluster of size M_1. To avoid the interference from neighboring clusters, use 4-TDMA as before. Hence, according to Lemma 3.3, it takes $M_1 D_2(M_1, M_2, L)$ time slots for each node to deliver $M_2 L$ bits to each node in the same cluster and this phase needs $4M_1 D_2(M_1, M_2, L)$ time slots to complete.

In Phase 2, as before, it takes $2n\frac{M_2 L}{R}$ time slots to complete.

In Phase 3, same as phase 1 except that there are $\frac{Q}{R}$ times as many bits to transmit, it takes $4M_1 D_2(M_1, M_2, L)\frac{Q}{R}$ time slots to complete.

Totally, with the triple stage scheme, it takes

$$D_3(n, M_1, M_2, L) = 4M_1 D_2(M_1, M_2, L)\left(1 + \frac{Q}{R}\right) + 2n\frac{M_2 L}{R}$$

time slots to communicate $M_1 M_2 L$ bits for each source–destination pair. This yields a throughput of

$$T = \frac{nM_1 M_2 L}{D_3(n, M_1, M_2, L)}. \tag{13}$$

The three stages of the scheme, namely S_1, S_2 and S_3 have been depicted in Fig. 5. In stage S_1, global MIMO transmissions are performed between clusters of size M_1. In stage S_2, clusters M_1 work in parallel and local MIMO transmissions are performed between subclusters of size M_2. S_3 is the bottom stage where point-to-point communications take place between nodes of subclusters.

Using the partial derivatives with respect to M_1 and M_2 to maximize the throughput in (13), the optimal cluster sizes are given by

$$M_2 = \frac{1}{2(1+Q/R)}\left(\frac{n}{2}\right)^{1/3}$$
$$M_1 = \frac{1}{2(1+Q/R)}\left(\frac{n}{2}\right)^{2/3}$$

and consequently the optimal throughput and the delay of the triple-stage scheme are given by

$$T_3^{\text{opt}}(n) = \frac{2^{1/3}}{24(1+Q/R)}Rn^{2/3}$$
$$D_3^{\text{opt}}(n,L) = \frac{3}{2^{1/3}(1+Q/R)}\frac{L}{R}n^{4/3}.$$

Remark 3.2: It is easy to prove as an extension of Lemma 3.1 that for the triple stage cooperation scheme, the received interference signals by different nodes of the network are uncorrelated in all the stages. Moreover the stage S_3 has the largest interference power which can be bounded by P_I [3]. Hence, the following coding rate R and quantization rate Q can be used in all the stages

$$R \le \max_{0\le\tau\le a^2} \log\left(1+\tau\frac{GP}{N+P_I+\Delta^2}\right)\frac{(a^2-\tau)^2}{2b^4} \tag{14}$$

$$Q > \log\left(1+\frac{GPb^2+N+P_I}{\Delta^2}\right) \tag{15}$$

Compared to the double stage scheme, the triple-stage scheme can achieve a higher order of n for throughput (an order of $n^{2/3}$ for the triple stage scheme in contrast with an order of $n^{1/2}$ for the double-stage scheme), but the preconstant of throughput decreases by increasing the number of stages. The desirable and adverse effects of increasing the number of stages can be explained as follows.

- Increasing the number of stages results in a better use of the degrees of freedom as the network transports more portion of the traffic by MIMO transmissions and less by TDMA. This in turn leads to an increase in order of n in the throughput.
- For a higher stage scheme, one should be able to bound the interference power due to parallel operating clusters which invokes running 4-TDMA in the network and at the same time inside the clusters. This yields an increase in the delay and consequently a reduction in the throughput. Another overhead arises from quantizing and reencoding the observations at different stages which further increases the delay and reduces the throughput.

C. H-Stage Hierarchical Cooperation Scheme

Generally, suppose that with the $(h-1)$-stage hierarchical cooperation scheme with cluster sizes $M_1, M_2, \ldots, M_{h-2}$, it takes $D_{h-1}(n, M_1, M_2, \ldots, M_{h-2}, L)$ time slots to communicate $M_1M_2\cdots M_{h-2}L$ bits for each source–destination pair.

[3]Since the scheme runs 4-TDMA in both network and clusters, the exact interference power is less than P_I, nevertheless this bound is sufficient to verify that a universal coding rate R is feasible.

Replacing phase 1 and phase 3 of the double stage scheme with the $(h-1)$-stage scheme, we have the h-stage scheme. Obviously, for the h-stage scheme with cluster sizes $M_1, M_2, \ldots, M_{h-1}$, it takes

$$\begin{aligned} &D_h(n, M_1, M_2, \ldots, M_{h-1}, L) \\ &\quad = 4M_1 D_{h-1}(M_1, M_2, \ldots, M_{h-1}, L)\left(1+\frac{Q}{R}\right) \\ &\quad + 2n\frac{M_2\cdots M_{h-1}L}{R} \end{aligned}$$

time slots to communicate $M_1M_2\cdots M_{h-1}L$ bits for each source–destination pair.

It can be verified that the general formula is

$$D_h(n, M_1, M_2, \ldots, M_{h-1}, L) = M_1M_2\cdots M_{h-1}\frac{L}{R}$$
$$\times\left\{[4(1+Q/R)]^{h-1}M_{h-1} + 2\sum_{i=0}^{h-2}[4(1+Q/R)]^i\frac{M_i}{M_{i+1}}\right\}.$$

Consequently, the throughput is given by

$$T_h(n, M_1, M_2, \ldots, M_{h-1}) = \frac{nM_1M_2\cdots M_{h-1}L}{D_h(n, M_1, M_2, \ldots, M_{h-1})}$$

which in general is a function of all the cluster sizes.

We maximize the throughput by using the partial derivatives. Solving $\partial T_h/\partial M_i = 0$ for $1 \le i \le h-1$ yields

$$M_i^2 = \frac{M_{i-1}M_{i+1}}{4(1+Q/R)}$$

where let $M_0 = n$ and $M_h = 2$. Therefore, the optimal choices of the cluster sizes are

$$M_i = \frac{2(n/2)^{(h-i)/h}}{[4(1+Q/R)]^{i(h-i)/2}} \quad \text{for } 1 \le i \le h-1. \tag{16}$$

Next we present one of our main results.

Theorem 3.1: For a regular network of n nodes in a unit area, by the h-stage hierarchical cooperation scheme with the optimal cluster sizes (16), the throughput is given by

$$T_h^{\text{opt}}(n) = \frac{R}{h(2\sqrt{1+Q/R})^{h-1}}(n/2)^{1-\frac{1}{h}} \tag{17}$$

and the corresponding delay is

$$D_h^{\text{opt}}(n,L) = \frac{h2^{(h+2)(h-1)/(2h)}}{(2\sqrt{1+Q/R})^{(h+3)(h-2)(h-1)/6}}\frac{L}{R}n^{\frac{h-1}{2}+\frac{1}{h}}.$$

For any fixed n, we can find the optimal h to maximize $T_h^{\text{opt}}(n)$. Let

$$\frac{dT_h^{\text{opt}}(n)}{dh} = 0$$

which leads to

$$h^2\ln(2\sqrt{1+Q/R}) + h - \ln(n/2) = 0.$$

Hence, the optimal number of stages to choose is

$$h^* = \frac{\sqrt{1+4\ln(2\sqrt{1+Q/R})\ln(n/2)}-1}{2\ln(2\sqrt{1+Q/R})}. \tag{18}$$

In order to obtain a simple formula, let

$$h^* = \sqrt{\frac{\ln(n/2)}{\ln(2\sqrt{1+Q/R})}} = \sqrt{\log_\beta(n/2)} \tag{19}$$

where $\beta := 2\sqrt{1+Q/R}$. Note that

$$\beta^h = \beta^{\log_\beta(n/2)\frac{h}{\log_\beta(n/2)}} = (n/2)^{\frac{h}{\log_\beta(n/2)}}.$$

Therefore,

$$\begin{aligned} T_h^{\text{opt}}(n) &= \frac{R}{h(2\sqrt{1+Q/R})^{h-1}}(n/2)^{1-\frac{1}{h}} \\ &= \frac{\beta R}{h\beta^h}(n/2)^{1-\frac{1}{h}} \\ &= \frac{\beta R}{h}(n/2)^{1-\frac{1}{h}-\frac{h}{\log_\beta(n/2)}} \end{aligned} \tag{20}$$

where choosing h as in (19), we have

$$T^{\text{opt}}(n) = \frac{\beta R}{\sqrt{\log_\beta(n/2)}}(n/2)^{1-\frac{2}{\sqrt{\log_\beta(n/2)}}}. \tag{21}$$

Obviously (21) is a very accurate estimation, although we made some approximation in (19) and h^* should always be an integer.

Theorem 3.2: For a regular network of n nodes in the unit area, by the hierarchical cooperation scheme with the optimal number of stages (18) and the optimal cluster sizes (16), the maximum throughput is approximately given by (21).

Actually, we can provide an exact upper bound of $T^{\text{opt}}(n)$. It follows from (20) that

$$\begin{aligned} T_h^{\text{opt}}(n) &\le \beta R(n/2)^{1-\frac{1}{h}-\frac{h}{\log_\beta(n/2)}} \\ &\le \beta R(n/2)^{1-\frac{2}{\sqrt{\log_\beta(n/2)}}} \end{aligned} \tag{22}$$

where, in the last inequality, " $=$ " holds if $h = \sqrt{\log_\beta(n/2)}$.

To check how much different (22) is from the linear scaling law $\Theta(n)$, we take the ratio

$$\begin{aligned} \frac{n/2}{(n/2)^{1-\frac{2}{\sqrt{\log_\beta(n/2)}}}} &= (n/2)^{\frac{2}{\sqrt{\log_\beta(n/2)}}} \\ &= \left(\beta^{\log_\beta(n/2)}\right)^{\frac{2}{\sqrt{\log_\beta(n/2)}}} \\ &= \beta^{2\sqrt{\log_\beta(n/2)}} \\ &\to \infty. \end{aligned}$$

Hence, the hierarchical cooperation scheme cannot achieve arbitrarily close to linear scaling. Instead, the difference grows to infinity as n increases. Consequently, the average rate per source–destination pair tends to zero.

D. Hierarchical Cooperation for Networks With Area A

Generally, consider a regular network with area A. Note that distance affects the power loss. We can scale down the general regular network with area A to a regular network with unit area, but with the power constraint $\frac{P}{(\sqrt{A})^\alpha}$ at each node, since the distance between nodes is reduced by a factor of $\sqrt{A}$. Recall that when $A = 1$ unit, running the hierarchy does not need the whole power budget P and the average power consumption is less than P/n per node. Thus, a general network can be dichotomized based on the relation between its area and the number of nodes into two cases.

- Dense network: The network is called dense when $A^{\alpha/2} \le n$. Then the nodes have enough power to run the hierarchical scheme and get the throughput-delay results as discussed above.
- Sparse network: The network is called sparse when $A^{\alpha/2} > n$. Then the nodes do not have sufficient power to run the hierarchical scheme all the time. Instead, they run the scheme in a fraction $n/A^{\alpha/2}$ of the time with power $PA^{\alpha/2}/n$ and remain silent during the rest of the time. Obviously this bursty modification satisfies the original average power constraint P, and correspondingly, the achieved throughput is modified by a factor of $n/A^{\alpha/2}$, e.g., in (17) and (21).

IV. Extension to Random Networks

In this section, we extend the results of regular networks to random networks. We first review the extension method of [1]: Consider a random network of unit area with n nodes. Since the average number of nodes in a cluster of area $A_c = \frac{M}{n}$ is M, the hierarchical scheme was applied to this random network by dividing the network into the clusters of area $\frac{M_1}{n}$ and proceeding to clusters of area $\frac{M_{h-1}}{n}$, for the h-stage scheme, and get the throughput-delay of the regular network but with a failure probability. Failure arises from the deviation of number of nodes in each cluster from its average. By a simple Chernoff bound argument, the probability of having large deviations from the average can be bounded (see [1, Lemma 4.1]). As $n \to \infty$, this probability goes to zero.

The above *clustering* method is not sufficient for the following reasons.

1) The clusters of area $A_c = \frac{M}{n}$ are required to contain *exactly* M nodes to perform the hierarchical scheme. A deviation from the average number of nodes M, even very small, results in failure of the scheme. However, [1] only bounded the probability of large deviation.
2) The probability of having exactly M nodes in a cluster of area $A_c = \frac{M}{n}$ is given by the binomial distribution $\mathfrak{p}(M; n, M/n) = \binom{n}{M}(\frac{M}{n})^M (1-\frac{M}{n})^{n-M}$. Using the Stirling's formula to approximate the factorial terms, as $n \to \infty$, yields

$$\mathfrak{p}(M; n, M/n) \approx \frac{M^M}{e^M M!}.$$

Recall that for the optimal operation of the scheme, the cluster sizes M are chosen proportional to n^γ where $0 < \gamma < 1$. Hence, the probability of having M nodes is proportional to $\frac{1}{\sqrt{2\pi M}}$ which, in fact, goes to zero.

To resolve the issue of making clusters of exactly M nodes, we will develop a clustering algorithm in this paper. To achieve high probability, we need to consider simultaneously the probabilities of events of the entire class of clusters, which invokes

a sort of uniform convergence (in probability) of law of large numbers over the entire class. To resolve this, we will resort to the *Vapnik–Chervonekis* theorem.

A. Choosing an Appropriate Cluster Shape

We use the Vapnik–Chervonekis theorem [13], [14] to find the appropriate cluster shape. Let $\mathcal{F}$ be a set of subsets and A a finite set of points. First, we recall some definitions as follows.

Definition 1: $\mathrm{Proj}_{\mathcal{F}}(A)$ is the projection of $\mathcal{F}$ on A which is defined as $\{F \cap A \ : \ F \in \mathcal{F}\}$.

Definition 2: A is *shattered* by $\mathcal{F}$ if $\mathrm{Proj}_{\mathcal{F}}(A) = 2^A$, i.e., if the projection of $\mathcal{F}$ on A includes all possible subsets of A.

Definition 3: The *VC-dimension of* $\mathcal{F}$, denoted by VC-d($\mathcal{F}$) is the cardinality of the largest set A that $\mathcal{F}$ shatters. It may be infinite.

The Vapnik–Chervonekis Theorem: If $\mathcal{F}$ is a set of finite VC-dimension and $\{X_i\}$ is a sequence of n i.i.d. random variables with common probability distribution P, then for every $\epsilon, \ \delta > 0$

$$\mathrm{Prob}\left\{\sup_{F\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n} I(X_i \in F) - P(F)\right| \le \epsilon\right\} > 1-\delta \tag{23}$$

whenever

$$n > \max\left\{\frac{8\,\mathrm{VC-d}(\mathcal{F})}{\epsilon}\log\frac{16e}{\epsilon}, \frac{4}{\epsilon}\log\frac{2}{\delta}\right\}. \tag{24}$$

An application of this theorem has been already presented in [3] for the set of disks on the plane. In this section, we consider a more general case; we apply the Vapnik-Chervonekis theorem to the set of all the clusters that partition the given random network with n nodes in the unit area. Note that a finite VC-dimension, for the set of clusters $\mathcal{F}$, is a sufficient condition for the uniform convergence in the weak law of large numbers. Assume that this condition is satisfied and the set of clusters has a finite VC-dimension (We will later derive a sufficient condition for the cluster shapes to make the VC-dimension finite). Denote the area of each cluster $c \in \mathcal{F}$ by A_c and its number of nodes with N_c, then we have the following lemma.

Lemma 4.1: For every cluster $c \in \mathcal{F}$ that contains exactly M nodes

$$\frac{M-\xi\log n}{n} < A_c < \frac{M+\xi\log n}{n} \tag{25}$$

with probability larger than $1-\frac{\xi\log n}{n}$ where $\xi = \max\{8\,\mathrm{VC-d}(\mathcal{F}), 16e\}$.

Proof: Let $\mathcal{F}$ denote the class of clusters with finite VC-dimension VC-d($\mathcal{F}$). To satisfy (24), let $\epsilon = \delta = \frac{\xi\log n}{n}$. Then the Vapnik–Chervonekis theorem states that

$$\mathrm{Prob}\left\{\sup_{c\in\mathcal{F}}\left(\left|\frac{N_c}{n} - A_c\right| \le \frac{\xi\log n}{n}\right)\right\} > 1-\frac{\xi\log n}{n}. \tag{26}$$

Therefore, if a cluster c contains exactly M nodes, i.e., $N_c = M$, then its area must satisfy (24) with high probability. □

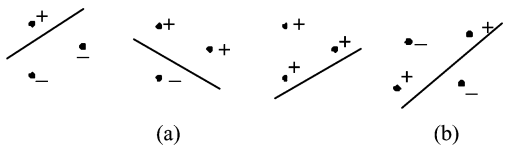

Fig. 6. VC-dimension for the set of half-spaces is 3. (a) A set of 3 points is shattered and (b) no set of 4 points can be shattered.

Note that if a cluster has an area less than $\frac{M-\xi\log n}{n}$, then with high probability it contains less than M nodes. Similarly, if its area is greater than $\frac{M+\xi\log n}{n}$, with high probability, it contains more than M nodes. Next, we need to choose a right shape for clusters to make the VC-dimension finite. We will make use of the following lemma, due to [15], in finding the appropriate shape. We have presented the sketch of the proof in the Appendix for completeness.

Lemma 4.2: Let $\mathcal{F}$ be a set of subsets with VC dimension d. Consider another set $\mathcal{F}_{\cap r}$ which consists of r-wise intersections of subsets in $\mathcal{F}$. The VC-dimension of the new set is at most $2rd\log(3r)$.

Corollary 4.1: The VC-dimension of the set of convex r-laterals is finite and upper bounded by $6r\log(3r)$ where r is the number of sides.

Proof: Consider a line in the plane. It divides the plane into two half-spaces. Choose one of the half-spaces as subset. Define $\mathcal{F}'$ as the set of all half-spaces produced by considering different lines in the plane. It is easy to prove that VC-d($\mathcal{F}'$) = 3 since a set of 3 nodes that are not collinear can be shattered (see Fig. 6(a)) but it is impossible to find a set of 4 nodes that are shattered by $\mathcal{F}'$ (see Fig. 6(b)). The labels "+" and "−" in Fig. 6 have been used to specify different subsets of points. The key observation is that any convex r-lateral is an intersection of r half-spaces. In the light of this observation and by using Lemma 4.2, it is concluded that the VC-dimension of the set of convex r-laterals is at most $6r\log(3r)$. □

We will use a set of quadrilaterals as the clusters. Since the VC-dimension is at most $24\log 12$, we can apply Lemma 4.2 with $\xi = 800$ to these clusters. Next, we develop an algorithm to make clusters of exactly M nodes.

B. Clustering Algorithm

Divide the network into squares of area $\frac{M}{n}$, and start from the square located on the top left corner. Depending on how many nodes are within this square, three situations may arise:

1) if the number of nodes in the square is exactly M, ignore this square and go to the next one;
2) if the number of nodes in the square is less than M, make a quadrilateral cluster by expanding the square: Move the top right vertex of the square to the right such that the created quadrilateral cluster contains exactly M nodes;
3) if the number of nodes in the square is more than M, make a quadrilateral by shrinking the square: Move the top right vertex of the square to the left such that the resultant quadrilateral cluster contains exactly M nodes.

After making the first cluster, go to the second cluster on the right side and make it a quadrilateral with exactly M nodes

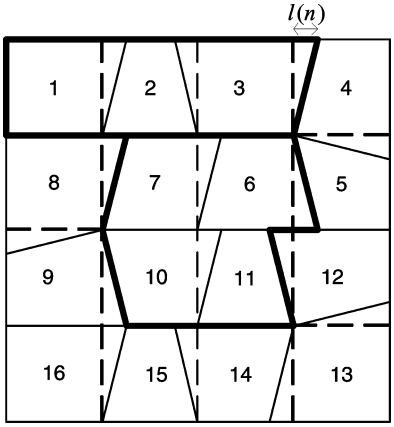

Fig. 7. Clustering of a random network with exactly M nodes in each quadrilateral cluster.

by expanding or shrinking as discussed above. Repeat the procedure for all the squares in the first row. For the top right square, use its bottom right vertex to do expanding/shrinking. For the second row, starting from the right square, move to the left side, and make the quadrilateral clusters of M nodes by expanding-shrinking. Perform the same procedure for all the rows, and we will have a set of quadrilateral clusters; each one contains exactly M nodes. One instance of such a clustering algorithm has been depicted in Fig. 7. Note that according to Lemma 4.1, the amount of expanding/shrinking in the areas of the squares is less than $\frac{\xi \log n}{n}$ with high probability.

C. Network Operation

The operation of random networks is similar to the operation of the regular networks. The centers of the quadrilateral clusters are defined as the centers of the original squares. Note that the new quadrilateral cluster will include the center of its original square with high probability. To observe this property of our clustering algorithm, consider the combination of the clusters 1, 2, and 3 in Fig. 7. This combination gives a larger quadrilateral cluster with $N_c = 3M$, hence according to (26) the deviation of the area of this cluster from its average $(3M/n)$ must be less than $\frac{\xi \log n}{n}$ and consequently $l(n) \leq \frac{2\xi \log n}{\sqrt{nM}}$. Therefore $l(n)$ is much smaller than the square side $\sqrt{\frac{M}{n}}$ (recall that $M = n^\gamma$ for $0 < \gamma < 1$) and the quadrilaterals are concentrated on the squares. In other words, each quadrilateral corresponds only to one square, and *vice versa*. Hence, the hierarchical scheme can be applied to the random networks by using the corresponding quadrilateral of each square instead of original square cluster. By making clusters of M_{h-1} nodes for the bottom stage of the hierarchy using the clustering algorithm, these clusters can be combined to make larger clusters of M_{h-2} nodes for the upper stage. Following the same procedure, make clusters of exactly M_1 nodes for the top stage. It is worth noting that for combined clusters, for example, combination of clusters 6, 7, 10, and 11 in Fig. 7, we can define the same deviation factor $l(n)$ as defined for the clusters of the bottom stage. As the result, the received power of each MIMO transmission will be lower-bounded and upper-bounded by (6). The only difference is that the coefficients a and b in (7)–(8) should be replaced by

$$a' = \left(1 + \frac{\sqrt{2}}{2}\frac{\sqrt{\frac{M}{n}} + l(n)}{\sqrt{\frac{M}{n}} - l(n)}\right)^{-\alpha/2}$$

$$b' = \left(1 - \frac{\sqrt{2}}{2}\frac{\sqrt{\frac{M}{n}} + l(n)}{\sqrt{\frac{M}{n}} - l(n)}\right)^{-\alpha/2}$$

or equivalently

$$a' = \left(1 + \frac{\sqrt{2}}{2}\frac{M + 2\xi \log n}{M - 2\xi \log n}\right)^{-\alpha/2}$$

$$b' = \left(1 - \frac{\sqrt{2}}{2}\frac{M + 2\xi \log n}{M - 2\xi \log n}\right)^{-\alpha/2}.$$

M can be chosen as the optimal cluster size for different stages, i.e., M_i for $1 \leq i \leq h-1$. But it holds for any i that

$$\frac{M_i + 2\xi \log n}{M_i - 2\xi \log n} \leq \frac{M_{h-1} + 2\xi \log n}{M_{h-1} - 2\xi \log n}.$$

Moreover, since $M_{h-1} \propto n^{1/h}$, the right-hand side (RHS) of this inequality is a decreasing function of n for large values of n and approaches to one. In fact, for any given $\eta > 1$, there exists a n_0 such that for all $n > n_0$, the RHS is less than η. Hence, for all $n > n_0$ the coefficients can be chosen as

$$a' = \left(1 + \frac{\sqrt{2}}{2}\eta\right)^{-\alpha/2}$$

$$b' = \left(1 - \frac{\sqrt{2}}{2}\eta\right)^{-\alpha/2}. \tag{27}$$

Consequently, the required quantization rate Q and the channel coding rate R can be defined based on the above coefficients. Obviously as $n \to \infty, a' \to a$ and $b' \to b$ and we can use the same quantization and coding rates as the rates already used for regular networks.

V. Conclusion

In this paper, the exact achievable throughput of the hierarchical scheme with any number of stages is derived. The optimal cluster sizes for all the stages are found to maximize the total throughput. We also find the optimal number of stages to choose for any network of size n. We observe that linear scaling is not achievable via the hierarchical scheme. As one increases the number of stages of the hierarchy to achieve a scaling closer to the linear one, the overhead due to using 4-TDMA scheme for parallel operating clusters and quantizing and re-encoding the observations at different stages, reduces the performance significantly. It also leads to an exponential growth for the delay. Finally, it is worth pointing out that the results presented here provide solid conclusions to networks with finite sizes, not only limiting behaviors.

APPENDIX A

Proof of Lemma 3.1: The proof follows in parallel with Lemma 4.2 of [1] for 9-TDMA. Consider Fig. 8 for the regular network in the unit area. The interference signal received by each node v is given by

$$I_v = \sum_{u \in \mathcal{U}_v} H_{vu} X_u$$

where H_{vu} is given by (3) and X_u is the signal transmitted by an active node located in a simultaneously operating cluster u with power σ_0^2. $\mathcal{U}_v$ is the set of clusters operating simultaneously with node v which can be grouped such that a group $\mathcal{U}_v(k)$ contains $8k$ clusters which are separated from v by a distance larger than $(2k-1)r$ where $r = \sqrt{\frac{M}{n}}$. The number of such groups can be easily bounded by $k \leq 1/4\sqrt{f}$ where $f = n/M$ is the number of clusters

$$P_I < \sum_{k=1}^{1/4\sqrt{f}} \sum_{U \in \mathcal{U}_v(k)} \frac{G\sigma_0^2}{((2k-1)r)^\alpha}$$

where we used the assumption that channel gains are independent. Substituting the value of σ_0^2 yields

$$P_I < 8GP \sum_{k=1}^{1/4\sqrt{f}} \frac{k}{(2k-1)^\alpha}.$$

When $\alpha > 2$, the above summation can be bounded by

$$\begin{aligned}
\sum_{k=1}^{\infty} \frac{k}{(2k-1)^\alpha} &= \frac{1}{2^\alpha} \sum_{k=1}^{\infty} \frac{1}{(k-1/2)^{\alpha-1}} \\
&\quad + \frac{1}{2^{\alpha+1}} \sum_{k=1}^{\infty} \frac{1}{(k-1/2)^\alpha} \\
&\leq \frac{1}{2^\alpha} \left(\frac{1}{(1/2)^{\alpha-1}} + \int_{1/2}^{\infty} \frac{dx}{x^{\alpha-1}} \right) \\
&\quad + \frac{1}{2^{\alpha+1}} \left(\frac{1}{(1/2)^\alpha} + \int_{1/2}^{\infty} \frac{dx}{x^\alpha} \right) \\
&= 1 + \frac{1}{4} \left(\frac{1}{\alpha-2} + \frac{1}{\alpha-1} \right)
\end{aligned}$$

which concludes the proof. □

APPENDIX B

Proof of Lemma 3.2: Consider Fig. 4. In the simple strategy of [1], each node simply quantizes the observations with rate Q bits per observation. Let the observations be encoded independently with a distortion constraint Δ^2. Since each observation is $\mathcal{N}_{\mathbb{C}}(0, P^{\text{rec}})$, Q must satisfy

$$Q > \log \left(1 + \frac{GPb^2 + N}{\Delta^2} \right). \tag{28}$$

Now consider the quantized MIMO channel which can be written as

$$\hat{Y}^M = HX^M + Z + D \tag{29}$$

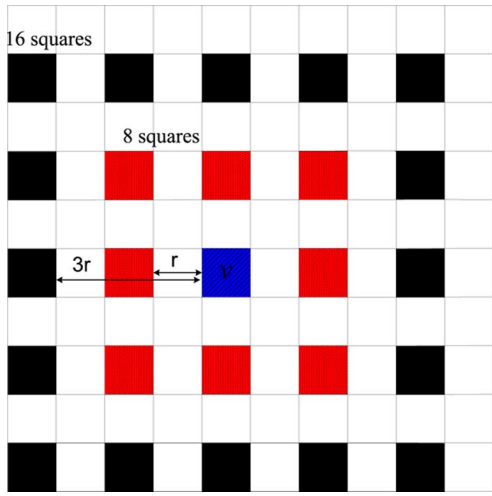

Fig. 8. Grouping of interfering clusters in 4-TDMA.

and $D \sim \mathcal{N}_{\mathbb{C}}(0, \Delta^2 I)$. The mutual information of this channel with CSI at receiver is given by $\mathcal{I}(X^M; Y^M, \mathbf{H})$ which can be written as $\log\det(I + \frac{1}{N'} H Q_X H^*)$ where $N' = N + \Delta^2$ (Noise and distortion are assumed to be uncorrelated). When H varies in a stationary ergodic manner, in general Q_X is chosen to maximize the expectation. Recall that in our model, H varies according to a stationary ergodic process, and elements of H are independent with mean zero, and different variances, such that the distributions of real and imaginary parts of the elements of H are symmetric around the origin. In this case, this is a well known result that the optimal Q_X must be diagonal. In other words, independent signaling can achieve the capacity. Now, consider the strategy of [1] when the elements of transmitted vector X^M are i.i.d. $\sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2)$, i.e., nodes use the same power $\sigma^2 = P \frac{r_{SD}^\alpha}{M}$. In this case, the mutual information is given by

$$\mathcal{I}(X^M; Y^M, \mathbf{H}) = \mathbb{E}\left[\log\det\left(I + \frac{\sigma^2}{N'} HH^* \right) \right].$$

Define ρ_{ik} as $(\frac{r_{SD}}{r_{ik}})^{\alpha/2}$, then the above mutual information can be written as

$$\mathbb{E}\left[\log\det\left(I + \frac{\text{SNR}}{M} FF^* \right) \right]$$

where $\text{SNR} = \frac{GP}{N'}$ and $F_{ik} = \rho_{ik} \exp(j\theta_{ik})$. Let $\lambda_1, \lambda_2, \cdots \lambda_M$ be the M eigenvalues of $\frac{1}{M} FF^*$, then the mutual information is given by

$$\begin{aligned}
\mathcal{I}(X^M; Y^M, \mathbf{H}) &= \sum_{k=1}^{M} \mathbb{E}\left[\log\left(1 + \text{SNR}\lambda_k \right) \right] \\
&\leq M\mathbb{E}\left[\log\left(1 + \frac{\text{SNR}}{M} \sum_{k=1}^{M} \lambda_k \right) \right] \\
&\leq M \log\left(1 + \frac{\text{SNR}}{M} \mathbb{E}\left[\sum_{k=1}^{M} \lambda_k \right] \right) \\
&= M \log\left(1 + \frac{\text{SNR}}{M^2} \mathbb{E}\left[\text{Tr}\left(FF^* \right) \right] \right)
\end{aligned}$$

where

$$\mathbb{E}\left[\mathrm{Tr}\ (FF^*)\right] = \sum_{i,k=1}^{M} \mathbb{E}[|F_{ik}|^2] = \sum_{i,k=1}^{M} \rho_{ik}^2 \leq M^2 b^2.$$

Here, we have used the fact that $a \leq \rho_{ik} \leq b$, and a and b are given by (8). Therefore the mutual information can be upper-bounded by

$$\mathcal{I}(X^M; Y^M, \mathbf{H}) \leq M \log(1 + \mathrm{SNR} b^2).$$

This upper bound, along with the lower bound of Lemma 4.3 of [1], yields the given bounds on the maximum achievable rate $R_1 = L/C_1$. The destination can decode the message with an average error probability arbitrary close to zero if $2^{LM+\epsilon} = 2^{C_1 \mathcal{I}(X^M;Y^M,\mathbf{H})}$ for any $\epsilon > 0$ as $C_1 \to \infty$. □

APPENDIX C

Proof of Lemma 4.2: The proof is based on the following Lemma.

Lemma C.1: Consider a set A of cardinality m and define $\Phi_d(m) = \sum_{i=0}^{d} \binom{m}{d}$ for $d \geq 0$. If the VC dimension of $\mathcal{F}$ is d, then

i) $|\mathrm{Proj}_A(\mathcal{F})| \leq \Phi_d(m)$;
ii) $\Phi_d(m) \leq 2(m^d/d!) \leq (em/d)^d$ for all $m \geq d \geq 1$.

Proof of Part (i): We show that for any $\mathcal{H} \subset 2^A$ that has VC dimension d, $|\mathcal{H}| \leq \Phi_d(m)$. Letting $\mathcal{H} = \mathrm{Proj}_A(\mathcal{F})$, we get the result. The proof of the latter is based on induction. Consider any point $x \in A$. Define the following sets:

$$\mathcal{H} - x = \{H - \{x\} : H \in \mathcal{H}\}$$
$$\mathcal{H}^{(x)} = \{H \in \mathcal{H} : x \notin H,\ H \cup \{x\} \in \mathcal{H}\}$$

Note that $\mathcal{H} - x$ and $\mathcal{H}^{(x)}$ are families of subsets of A and that $|\mathcal{H}| = |\mathcal{H} - x| + |\mathcal{H}^{(x)}|$. Obviously $\mathcal{H} - x$ has VC dimension at most d and therefore $|\mathcal{H} - x| \leq \Phi_d(m-1)$. If we prove $\mathcal{H}^{(x)}$ has VC-dimension less than $d-1$, then the lemma follows since $\Phi_d(m-1) + \Phi_{d-1}(m-1) = \Phi_d(m)$. The VC-dimension of $\mathcal{H}^{(x)}$ is at most $d-1$ since if its VC-dimension is d, there exists a set $B \subset A - \{x\}$ such that it is shattered by $\mathcal{H}^{(x)}$ and $|B| = d$. But in this case, $B \cup \{x\} \subset A$ can be shattered by $\mathcal{H}$; it means that VC-dimension of $\mathcal{H}$ is $d+1$ which is impossible.

Proof of Part (ii): The second inequality of part (ii) is based on Stirling's approximation for $d!$ and the proof of the first inequality is by induction on d and m.

To prove Lemma 4.2, suppose A has m elements. Note that, according to the above lemma, $|\mathrm{Proj}_A(\mathcal{F})| \leq \Phi_d(m)$. Every set in $\mathrm{Proj}_A(\mathcal{F}_{\cap r})$ is of the form $\bigcap_{i=1}^{r} A_i$ where $A_i \in \mathrm{Proj}_A(\mathcal{F})$. This shows that $|\mathrm{Proj}_A(\mathcal{F}_{\cap r})| \leq \binom{|\mathrm{Proj}_A(\mathcal{F})|}{r}$. If $|\mathrm{Proj}_A(\mathcal{F}_{\cap r})| < 2^m$, A can not be shattered by $\mathcal{F}_{\cap r}$. Therefore, by part (ii) of the lemma, it suffices to choose m such that

$$\binom{\Phi_d(m)}{r} \leq \left(\frac{em}{d}\right)^{dr} < 2^m$$

which is satisfied when $m = 2dr\log(3r)$. This concludes the proof. □

REFERENCES

[1] A. Ozgur, O. Leveque, and D. Tse, "Hierarchical cooperation achieves optimal capacity scaling in Ad Hoc networks," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3549–3572, Oct. 2007.

[2] E. C. van der Meulen, "Three-terminal communication channels," *Adv. Appl. Prob.*, vol. 3, pp. 120–154, 1971.

[3] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, pp. 388–404, Mar. 2000.

[4] 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, Feb. 2004.

[5] A. Jovicic, P. Viswanath, and S. R. Kulkarni, "Upper bounds to transport capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 50, pp. 2555–2565, Nov. 2004.

[6] O. Leveque and 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.

[7] 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.

[8] 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, Joint Special Issue of IEEE Trans. on Information Theory and IEEE/ACM Trans. on Networking on Networking and Information Theory, no. 6, pp. 2313–2328, Jun. 2006.

[9] S. Ahmad, A. Jovicic, and P. Viswanath, "Outer bounds to the capacity region of wireless networks," *IEEE Trans. Inf. Theory*, vol. 52, pp. 2770–2776, Jun. 2006.

[10] M. Franceschetti, O. Dousse, D. Tse, and P. Thiran, "Closing the gap in the capacity of wireless networks via percolation theory," *IEEE Trans. Inf. Theory*, vol. 53, no. 3, pp. 1009–1018, Mar. 2007.

[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] J. Ghaderi, L.-L. Xie, and X. Shen, "Throughput optimization for hierarchical cooperation in Ad Hoc networks," in *Proc. 2008 IEEE Int. Conf. Commun.*, Beijing, China, May 19–23, 2008.

[13] V. N. Vapnik and A. Chervonenkis, "On the uniform convergence of relative frequencies of events to their probabilities," *Theory Prob. Applicat.*, vol. 16, no. 2, pp. 264280–, 1971.

[14] V. N. Vapnik, *Estimation of Dependences Based on Empirical Data*. New York: Springer-Verlag, 1982.

[15] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth, "Learnability and the Vapnik-Chervonenkis dimension," *J. ACM*, vol. 36, no. 4, pp. 929–965, 1989.

[16] C. A. Balanis, *Antenna Theory: Analysis and Design*. New York: Wiley, 2005.

Javad Ghaderi (S'09) received the B.Sc. degree from the University of Tehran, Iran, in 2006 and the M.Sc. degree from the University of Waterloo, Canada, in 2008, both in electrical engineering. He is currently working towards the Ph.D. degree at the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign.

His research interests include wireless networks, network control and optimization, and network information theory.

Liang-Liang Xie (M'03–SM'09) received the B.Sc. degree in mathematics from Shandong University, Jinan, China, and the Ph.D. degree in control theory from the Chinese Academy of Sciences, Beijing, China, in 1995 and 1999, respectively.

He held postdoctoral positions at the Automatic Control Group, Linköping University, Sweden during 1999–2000 and the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, during 2000–2002. He was a faculty member of the Institute of Systems Science, Chinese Academy of Sciences before joining the Department of Electrical and Computer Engineering, University of Waterloo, Canada in 2005, where he is currently an Associate Professor. His research interests include network information theory, wireless networks, adaptive control and system identification.

Xuemin (Sherman) Shen (M'97–SM'02–F'09) received the B.Sc. degree from Dalian Maritime University, China, in 1982 and the M.Sc. and Ph.D. degrees from Rutgers University, NJ, in 1987 and 1990, respectively, all in electrical engineering.

He is a Professor and University Research Chair, Department of Electrical and Computer Engineering, University of Waterloo, Canada. Dr. Shen's research focuses on mobility and resource management in interconnected wireless/wired networks, UWB wireless communications networks, wireless network security, wireless body area networks and vehicular *ad hoc* and sensor networks. He is a coauthor of three books, and has published more than 400 papers and book chapters in wireless communications and networks, control and filtering.

Dr. Shen served as the Tutorial Chair for IEEE ICC'08, the Technical Program Committee Chair for IEEE Globecom'07, the General Co-Chair for Chinacom'07 and QShine'06, the Founding Chair for IEEE Communications Society Technical Committee on P2P Communications and Networking. He also serves as a Founding Area Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS; Editor-in-Chief for *Peer-to-Peer Networking and Application*; Associate Editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY; *KICS/IEEE Journal of Communications and Networks*, *Computer Networks*; *ACM/Wireless Networks*; and *Wireless Communications and Mobile Computing* (New York: Wiley), etc. He has also served as Guest Editor for IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, IEEE WIRELESS COMMUNICATIONS, IEEE COMMUNICATIONS MAGAZINE, and *ACM Mobile Networks and Applications*, etc. He received the Excellent Graduate Supervision Award in 2006, and the Outstanding Performance Award in 2004 and 2008 from the University of Waterloo, the Premier's Research Excellence Award (PREA) in 2003 from the Province of Ontario, Canada, and the Distinguished Performance Award in 2002 and 2007 from the Faculty of Engineering, University of Waterloo. He is a registered Professional Engineer of Ontario, Canada, and a Distinguished Lecturer of IEEE Communications Society.