A Scaling Law Approach to Wireless Relay Networks

Ayfer Özgür
Stanford University
Stanford, California 94305
aozgur@stanford.edu

***Abstract*—There has been significant recent progress in understanding approximately optimal relaying strategies for wireless networks, such as quantize-map-and-forward, noisy network coding, amplify-and-forward, etc. While one can construct specific instances of networks, with specific topology and channel configurations, where each one of these strategies can provide significant gains over simple strategies such as routing and direct transmission, it is not clear how much gain these more sophisticated strategies can provide in generic setups. In this paper, we follow a scaling law approach and assume that nodes are randomly distributed over the network area and the channels between them are governed by a path-loss model. This approach has been used to demonstrate the benefits of sophisticated cooperation in networks with multiple unicast flows both in the high and the low SNR regimes. However, for a single unicast flow, we show that more sophisticated relaying can not provide significant gain over simple multi-hop or direct transmission both in the high and the low-SNR regimes, if the nodes are uniformly distributed over the network area. More sophisticated relaying is needed in networks where nodes are clustered in the low to moderate SNR regimes. We propose a cluster decode-and-forward strategy that is scaling optimal.**

I. Introduction

Motivated by the relaying opportunities provided by the massive proliferation of wireless devices, capacity study of Gaussian relay networks has received significant attention over the last decade. The work of Avestimehr, Diggavi and Tse [1] has shown that a compress-and-forward type of strategy at the relays is universally good for dealing with the broadcast and superposition of wireless signals and can achieve close to the capacity of relay networks in the high capacity regime, across different channel configurations and topologies. Amplify-and-forward type of strategies have been investigated for the low-SNR regime in [12]. Can such more sophisticated relaying strategies provide significant rate gains over the traditional routing (milti-hop) approach in wireless adhoc networks? If so, what are the operating regimes and scenarios where sophisticated relaying is most useful? These are important questions concerning future communication architectures for such networks.

It is easy to construct specific instances of networks where each one of these strategies can provide gains over routing. For example, when the channel gain parameter t is large in the diamond network in Figure 1-(b), a compress-and-forward type of strategy [1], [2], [3] that employs both the relays can approximately achieve twice the rate achievable by routing over a single relay. As $t \to \infty$, the rate difference between the two strategies increases unboundedly. However, the ratio of the rates achieved by the two strategies remains bounded by 2. Indeed, [4] shows that this bound holds even with n-relays and arbitrary channel configurations: in the high-capacity regime, the capacity of any n-relay diamond network is approximately bounded by twice the rate achieved by routing over the best relay. Therefore, for the diamond topology, a compress-and-forward type of strategy employing all the n relays can provide a rather modest multiplicative gain over routing. Moreover, for most configurations the gain is much smaller than 2. For example, routing over one of the relays approximately achieves capacity in Figure 1-(a) when t is large. Similarly, one can construct specific examples of networks where amplify-and-forward relaying can provide large gains over routing, but usually such large gains only pertain to very specific channel configurations [12].

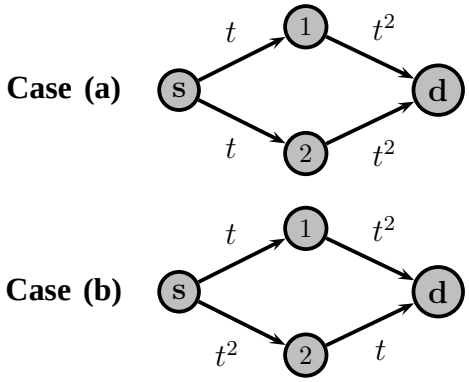

Fig. 1. Two instantiations of a diamond relay network.

Evaluating the benefits of different relaying strategies not only requires to understand the gains they provide in specific network instances, but also the likelihood of each instance. In this paper, we follow the scaling law approach of Gupta and Kumar [6]. This approach adopts a random model for the location of the nodes and assumes that channel gains are governed by a path loss and fading model which is a function of the node locations. The scaling law approach has been used in [7], [8], [9], [10] to show that when there are multiple source-destination pairs in the network (multiple unicast traffic), sophisticated cooperation techniques, such as hierarchical cooperation in [7], can provide a multiplicative gain as large as $\Theta(\sqrt{n})$ in the system capacity in various

operating regimes, where n is the number of nodes in the network. The gain follows from the ability of cooperation to harness interference between different source-destination pairs. Can sophisticated relaying strategies provide similar gains when there is only a single-source destination pair in the network?

This is the question we investigate in this paper. We show that the answer is no when the wireless nodes are uniformly distributed over the network area. Direct transmission from the source node to the destination and multi-hopping are sufficient to approximately achieve the capacity in both the high and the low SNR regimes. Sophisticated relaying is needed only when the nodes in the network are clustered together and the direct channel between the source and the destination is weak, i.e. the SNR is not too large. We propose a cluster decode-and-forward strategy to achieve the optimal scaling of the capacity in this case. This strategy uses multi-hop communication inside the clusters to facilitate MIMO communication across clusters. Information is then routed from one cluster to the next via successive MIMO transmissions, by being decoded and re-encoded at each intermediate cluster. The strategy relies on spatial reuse to divide the end-to-end communication problem to successive steps. Decoding at each step prevents noise accumulation and makes the strategy good also in the low-SNR regime as opposed to compress-and-forward type of strategies [1], [2], [3] which are only effective at high-SNR.

II. Model

There are n wireless nodes located in a rectangle of area of $\sqrt{A} \times \sqrt{A}$. We consider two different spatial models for the distribution of the nodes in the network.

(a) The n nodes are distributed uniformly and independently over the area A.

(b) The n nodes are clustered into n/M clusters each of size M nodes and area A_c. Each cluster of M nodes is uniformly distributed over the cluster area, while the clusters are distributed uniformly over the network area A. We assume that $A_c \leq \frac{MA}{n}$. Note that if the n nodes were uniformly distributed over the area A, the area occupied by M nodes would be $\frac{MA}{n}$ on the average. $A_c \leq \frac{MA}{n}$ corresponds to assuming that nodes are indeed clustered together.

The scaling law studies in [6], [7], [8] reveal that when n is large the capacity of such a random network is the same as the capacity of a network where nodes are placed on a regular grid. See Figure 2. In other words, typical configurations of the large random network are close to a grid, and typical deviations from the grid do not significantly impact capacity while atypical deviations are less and less likely as n increases. For simplicity in derivations, we will consider the regular networks corresponding to the random distributions (a) and (b), as given in Figure 2. A randomly chosen node s among the n nodes wants to communicate to a randomly chosen destination d at rate R bits/s/Hz.

We assume that communication takes place over a flat channel of bandwidth W Hz around a carrier frequency of

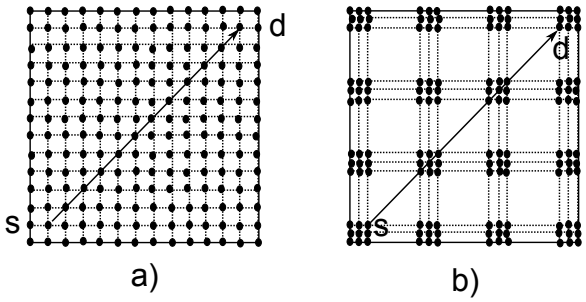

Fig. 2. (a) Uniform and (b) clustered network.

f_c, $f_c \gg W$. The complex baseband-equivalent channel gain between node i and node k at time m is given by:

$$H_{ik}[m] = \sqrt{G}\, r_{ik}^{-\alpha/2} \exp(j\theta_{ik}[m]) \tag{1}$$

where r_{ik} is the distance between the nodes, $\theta_{ik}[m]$ is the random phase at time m, uniformly distributed in $[0, 2\pi]$ and $\{\theta_{ik}[m]; 1 \leq i \leq 2n, 1 \leq k \leq 2n\}$ is a collection of independent identically distributed random processes. The $\theta_{ik}[m]$'s and the r_{ik}'s are also assumed to be independent. The parameters G and $\alpha > 2$ are assumed to be constants; α is called the power path loss exponent.

Note that the channel is random, depending on the location of the users and the phases. The locations are assumed to be fixed over the duration of the communication. The phases are assumed to vary in a stationary ergodic manner (fast fading). We assume that each node has power P and the network is allocated a total bandwidth of W. While deriving upper bounds on the best achievable rate between s and d, we assume that the phases $\{\theta_{ik}[m], \forall i, k\}$ are known in a casual manner to all the nodes in the network. However, the strategies we propose only use receive channel state information, i.e. it is sufficient for each node k in the network to know only the phases of its incoming channels $\{\theta_{ik}[m], \forall i\}$ in a casual manner. The signal received by node i at time m is given by

$$Y_i[m] = \sum_{k \neq i} H_{ik}[m] X_k[m] + Z_i[m]$$

where $X_k[m]$ is the signal sent by node k at time m and $Z_i[m]$ is white circularly symmetric Gaussian noise of power spectral density $N_0 W$ per symbol.

We define SNR_s to be the typical SNR between nearest neighbor pairs. In the case of a uniform network

$$\text{SNR}_s = \frac{GP}{N_0 W (\sqrt{A/n})^{\alpha}}, \tag{2}$$

since $\sqrt{A/n}$ is the nearest neighbor distance in a regular network of the form in Figure 2-(a) (and the typical nearest neighbor distance in a network where nodes are uniformly distributed over the network area). We similarly define the long-distance SNR in the network to be the SNR of a point-to-point transmission over the largest scale in the network, the

diameter $\sqrt{A}$,

$$\text{SNR}_l = \frac{GP}{N_0 W(\sqrt{A})^\alpha}. \quad (3)$$

Note that SNR_s and SNR_l are related as $\text{SNR}_s = n^{\alpha/2}\text{SNR}_l$.

In the case of a clustered network with clusters of M nodes distributed over an area A_c, the nearest neighbor is instead given by

$$\text{SNR}_s = \frac{GP}{N_0 W(\sqrt{A_c/M})^\alpha}, \quad (4)$$

which we again denote by SNR_s by slightly abusing notation. A_c/M is the separation between nearest neighbors in each cluster in this case. We also define the inter-cluster SNR for clustered networks which corresponds to the SNR of a point-to-point transmission between two nodes located in neighboring clusters,

$$\text{SNR}_c = \frac{GP}{N_0 W(\sqrt{MA/n})^\alpha}. \quad (5)$$

Note that $\sqrt{MA/n}$ is the separation between the midpoints of two neighboring clusters in Figure 2-(b). Finally, the long distance SNR in the network, SNR_l is again given by (3). Since $A_c \leq \frac{MA}{n} \leq \sqrt{A}$, we have

$$\text{SNR}_s \geq M^{\alpha/2}\text{SNR}_c = n^{\alpha/2}\text{SNR}_l. \quad (6)$$

A strategy is called scaling optimal in a certain regime $\mathcal{R}$, if for a certain coupling of the system parameters $A = n^{\beta_1}$, $P = n^{\beta_2}$, $W = n^{\beta_3}$ such that $(\beta_1, \beta_2, \beta_3) \in \mathcal{R}$, it achieves the scaling exponent of the capacity of the network

$$e(\beta_1, \beta_2, \beta_3) := \lim_{n\to\infty} \frac{\log C(n, \beta_1, \beta_2, \beta_3)}{\log n} \quad (7)$$

where C is the largest reliable communication rate R achievable between s and d. n, A, P and W are all independent parameters of a network that can take on a wide range of values. By considering all possible couplings between these parameters, we aim to investigate all cases where they can be large or small with respect to each other. (In the case of a clustered network, we have the additional parameters A_c and M.) Most often, we will see that the scaling of the capacity and the performance of a strategy depend on β_1, β_2, β_3 not separately but only through a single SNR parameter. Note that if a strategy is scaling optimal, within logarithmic factors, its performance exhibits the right dependence to major system parameters, same as the capacity itself.

III. Main Result

The main conclusions of this paper are summarized in the following theorems. The first theorem establishes an upper bound on the best achievable rate between s and d in a uniform network. The second theorem establishes a lower bound on the rate achieved by multi-hopping in the same network. The proposition gives the rate achieved with direct transmission from s to d.

Theorem 3.1: When the nodes are uniformly distributed over the network area, the largest end-to-end achievable communication rate between s and d is bounded by

$$R \leq \log(1 + K_1 \text{SNR}_s)$$

for a positive constant[1] $K_1 \leq 4(2 + \pi/2 + \pi/(2(\alpha - 2)))$.

Theorem 3.2: When the nodes are uniformly distributed over the network area, multi-hop achieves a rate given by

$$R \geq \log\left(1 + \frac{\text{SNR}_s}{1 + K_2 \text{SNR}_s}\right)$$

between s and d for a positive constant $K_2 \leq 4\alpha/(\alpha - 1)$.

Proposition 3.3: Direct transmission from s to d achieves a rate

$$R \geq \log\left(1 + (2)^{-\alpha/2}\text{SNR}_l\right).$$

Comparing these results, we observe that:

- when $\text{SNR}_s \ll 0$ dB (i.e. in the scaling law sense $\text{SNR}_s = n^\gamma$ for $\gamma \leq 0$) the upper and the lower bounds in Theorem 3.1 and Theorem 3.2 respectively are of the order of SNR_s, therefore multi-hop is scaling optimal. Note that this is a low-SNR regime; even the nearest neighbor transmissions in the network are at low-SNR. For example, this is the case in a network with growing number of users and fixed density, known as extended scaling [7]. In this regime, more sophisticated relaying strategies (such as amplify-and-forward) can not provide significant gain over multi-hop.
- when $\text{SNR}_s \gg 0$ dB the upper bound in Theorem 3.1 is of the order of $\log n$ and the lower bound in Theorem 3.2 is constant, the two differing only by a factor of $\log n$. Therefore, multi-hop is still (approximately) scaling optimal. More sophisticated relaying strategies can only provide marginal gain over multi-hop. Note that this can be interpreted as a moderate to high SNR regime for the network; nearest neighbor transmissions are at low SNR, while transmissions over larger distances can be in either high or low SNR. When $\text{SNR}_l \gg 0$ dB, i.e., when all the channels in the network are at high-SNR, including the direct channel between s and d, direct transmission from s to d becomes scaling optimal. This is, for example, the case when there are an increasing number of nodes on a fixed area, known as dense scaling [6].
- although it is not part of our scaling law formulation in (7), it is instructive to look at the case when $\text{SNR}_s = e^{n^\gamma}$ for $\gamma > 0$, i.e. when the network is in a very high-SNR regime, the upper bound is now of the order of n^γ while the rate achieved by multi-hop is still constant. At very high-SNR, spatial reuse is not desirable since interference from simultaneous transmissions in the network, no matter how far they are, significantly degrades performance. Since multi-hop is based on spatial reuse, it is not anymore scaling optimal. Simple direct transmission from s to d achieves the optimal scaling.

[1]A constant is independent of the parameters of the network, n,A, P, W, N_0 etc.

The above discussion shows that in all SNR-regimes, simple strategies such as multi-hop and direct transmission are sufficient to approximately achieve the capacity of the network. More sophisticated relaying strategies, such as amplify-and-forward, quantize-map-forward or noisy network coding, can not provide significant gains over these simple strategies in uniform networks.

We next turn to networks where nodes are clustered as in Figure 2-(b). In Theorem 3.4, we establish an upper bound on the best achievable rate between s and d. Proposition 3.5 gives the rate achieved by multi-hop. Theorem 3.6 gives the rate achieved with a cluster decode-and-forward strategy.

Theorem 3.4: When the nodes are clustered as described in Section II, the largest end-to-end achievable communication rate between s and d is bounded by

$$R \leq \min\left(\log(1+K_1\text{SNR}_s),\ K_3 M^{\epsilon}\, M^2\text{SNR}_c\right)$$

for the constant K_1 in Theorem 3.1, and a constant $K_3 > 0$.

Proposition 3.5: In a clustered network, multi-hop from s to d achieves a rate

$$R \geq \log\left(1+\frac{\text{SNR}_c}{1+K_2\text{SNR}_c}\right) \tag{8}$$

where K_2 is the positive constant in Theorem 3.2. Direct transmission from s to d achieves a rate

$$R \geq \log\left(1+(2)^{-\alpha/2}\text{SNR}_l\right). \tag{9}$$

Theorem 3.6: When the nodes are clustered as described in Section II, the cluster decode-and-forward strategy described in the Section VI achieves a rate

$$\min\left(\log\left(1+\frac{\text{SNR}_s}{1+K_3\text{SNR}_s}\right),\ K_4 M\log\left(1+\frac{M\text{SNR}_c}{1+M\text{SNR}_c}\right)\right)$$

for positive constants K_3 and K_4.

Comparing these three results we identify the following regimes:

- when $\text{SNR}_l \gg 0$ dB, the upper bound in Theorem 3.4 is of the order of $\log n$ and the lower bound in (9) is constant. Direct transmission from s to d is approximately scaling optimal. The network is in a high-SNR regime; even the long-range direct channel from s to d is at high SNR.
- when $\text{SNR}_l \ll 0$ dB but $\text{SNR}_c \gg 0$ dB, the upper bound in Theorem 3.4 is of the order of $\log n$, and multi-hop in (8) achieves a constant rate. Therefore, multi-hop is scaling optimal. This is a moderate SNR regime; while the long-range direct channel between s and d is at low-SNR, the channels between nodes in neighboring clusters are at high-SNR. Both in the earlier and the current regime, more sophisticated relaying can not provide significant gains over direct transmission and multi-hop relaying respectively.
- when $\text{SNR}_c \ll 0$ dB, the rate achieved by multi-hop is of the order of SNR_c while the upper bound in Theorem 3.4 is either still of the order of $\log n$ or $\min(SNR_s, M^2\text{SNR}_c)$. In all cases, the cluster decode-and-forward strategy of Theorem 3.6 is needed to achieve the optimal scaling. Note that this regime comprises many sub-regimes: although $\text{SNR}_c \ll 0$ dB, SNR_s as well as $M\text{SNR}_c$, the SNR of a MIMO transmission between two neighboring clusters, can still be larger than 0 dB. This can be interpreted as a moderate SNR regime. On the other hand, both SNR_s and $M\text{SNR}_c$ can be smaller than 0 dB, in which case the network is at low SNR. We can also have one of these quantities at high and the other one at low-SNR. The cluster decode-and-forward strategy achieves the optimal scaling in all cases.

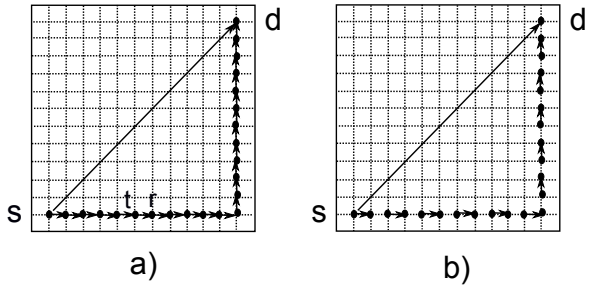

Fig. 3. (a) The route from s to d. (b) Spatial Reuse.

IV. Uniform Networks

In this section, we prove Theorems 3.1 and 3.2 and Proposition 3.3.

Proof of Theorem 3.1: The largest communication rate between s and d can be bounded by the total mutual information that can be conveyed from s to the rest of the network. The capacity of the single-input-multiple-output (SIMO) channel between s and the remaining nodes in the network is given by [11]

$$\begin{aligned} R &\leq \log\left(1+\sum_{i,k}|h_{ik}|^2\frac{P}{N_0W}\right) \\ &\leq \log\left(1+4\sum_{i=0,k=1}^{\sqrt{n}/2}\frac{1}{(i^2+k^2)^{\alpha/2}}\frac{GP}{N_0W(\sqrt{A/n})^{\alpha}}\right) \\ &= \log\left(1+4B\,\text{SNR}_s\right), \end{aligned}$$

where the second inequality follows by observing that the sum $\sum_{i,k}|h_{ik}|^2$ is largest if s is located in the middle of the grid

in Figure 2-(a). A can be upper bounded as

$$B = \sum_{i=0,k=1}^{\sqrt{n}/2} \frac{1}{(i^2+k^2)^{\alpha/2}} \tag{10}$$

$$\overset{(a)}{\leq} \sum_{i=0}^{\sqrt{n}/2} \left(\frac{1}{(1+i^2)^{\alpha/2}} + \int_1^{\sqrt{n}/2} \frac{1}{(x^2+i^2)^{\alpha/2}} dx \right)$$

$$\overset{(b)}{\leq} 1 + \int_1^{\sqrt{n}/2} \frac{1}{x^\alpha} dx + \int_0^{\sqrt{n}/2} \frac{1}{(1+y^2)^{\alpha/2}} dy + \int_0^{\sqrt{n}/2} \int_1^{\sqrt{n}/2} \frac{1}{(x^2+y^2)^{\alpha/2}} dxdy$$

$$\overset{(c)}{\leq} (2+\pi/2) + \int_0^{\pi/2} \int_1^{\sqrt{n}/2} \frac{1}{r^\alpha} r dr d\theta, \tag{11}$$

where (a) follows by bounding the area given by the Riemann sum with the integral, (b) follows by applying the idea in (a) for the second sum, (c) follows by bounding the two integrals in (b) by assuming $\alpha = 2$ and a change of variables for the last integral. Bounding the last integral in (c), we obtain

$$B \leq 2 + \pi/2 + \pi/(2(\alpha-2)). \tag{12}$$

Proof of Theorem 3.2: In the multi-hop strategy, the packets between s and d are relayed by successive point-to-point transmissions between neighboring nodes. Each intermediate relay node decodes the packets from the previous node and forwards them to the next while the interference from simultaneous transmissions is treated as additional noise. We assume that the packets are first routed over a horizontal and then a vertical path as illustrated in Figure 3-(a). The rate achieved by this multi-hop strategy depends on two parameters: the SNR of the nearest neighbor transmissions, given by SNR_s, and the interference from simultaneous transmissions. The interference-to-noise power ratio (INR) at receiver node r is given by

$$\text{INR} = \sum_{k \in \mathcal{U}_\sqcup} \frac{GP}{N_0 W r_{rk}^\alpha},$$

where $\mathcal{U}_\sqcup$ is the set of all transmitters other than the intended transmitter t. The above sum can be bounded as

$$\text{INR} \leq 4 \sum_{k=1}^{\sqrt{n}/2} \frac{GP}{N_0 W (k\sqrt{A/n})^\alpha}$$

by assuming the worst case scenario that r is in the middle of the horizontal line and loosely upper bounding the interference from the transmitters on the vertical line by that from the horizontal line. This yields,

$$\text{INR} \leq 4 \sum_{k=1}^{\sqrt{n}/2} \frac{1}{k^\alpha} \text{SNR}_s \leq 4(1+1/(\alpha-1))\text{SNR}_s.$$

Therefore, the rate achieved by multi-hop can be lower bounded as

$$R \geq \log\left(1 + \frac{\text{SNR}_s}{1+(4\alpha/(\alpha-1))\text{SNR}_s}\right).$$

In order to decrease interference one can also consider a partial spatial reuse strategy where only alternate nodes on the route are allowed to transmit simultaneously, see Figure 3-(b).

Proof of Proposition 3.3: The proof of the proposition simply follows from the fact that the separation between s and d can be at most $\sqrt{2A}$.

V. Clustered Networks

In this section, we prove Theorems 3.4 and 3.6. The proof idea for Proposition 3.5 is given in the next section.

Proof of Theorem 3.4: The first upper bound in the theorem follows from the SIMO bound in Theorem 3.1. Note that the SIMO bound only depends on the SNR of s to its nearest neighbors, which is SNR_s in the case of a uniform network and SNR_s^c in the current case. The number of nodes in the network is irrelevant to the upper bound as long as nodes are located on a regular grid of the corresponding minimal distance. Note that if not every location on the grid is occupied, this can only decrease the upper bound.

The second term in the upper bound follows by considering the cut between the source cluster and the rest of the network. The largest communication rate from s to the d is upper bounded by the mutual information that can be conveyed over this cut, i.e. the capacity of the M to $n-M$ MIMO channel. When the channels are ergodically fading, the capacity of this MIMO channel is given by [11]

$$R \leq \max_{\substack{Q(H)\geq 0 \\ \mathbb{E}(Q_{kk}(H)) \leq P/N_0 W, \forall k \in S}} \mathbb{E}\left(\log\det(I + HQ(H)H^*)\right), \tag{13}$$

where

$$H_{ik} = \frac{\sqrt{G}\, e^{j\,\theta_{ik}}}{r_{ik}^{\alpha/2}}, \quad k \in S, i \in \mathcal{N} \setminus S.$$

under the optimistic assumption that the realizations of the channels are know at both the transmitter and the receiver. This allows the transmission covariance matrix $Q(\cdot)$ to be a function of the channel matrix H. We use S to the denote the cluster of the source node s and $\mathcal{N} \setminus S$ denotes the remaining nodes.

The capacity of the MIMO channel in (13) can be upper bounded in two steps. First, we can show that choosing $Q(H) = \frac{P}{N_0 W} I$ is approximately optimal. It has been shown in [7][Lemma 5.2] that the increase in capacity for any other choice $Q(H)$ of the covariance matrix is bounded by M^ϵ for any $\epsilon > 0$. This says that independent signaling at the transmit nodes is sufficient to achieve the cut-set bound. Beamforming techniques can provide limited gains, essentially because the phases between different nodes are independent of each other. The remaining step is to bound the capacity of the MIMO channel under independent transmissions $Q(H) = \frac{P}{N_0 W} I$. Note that

$$\mathbb{E}\left(\log\det(I + \frac{P}{N_0 W} HH^*)\right) \leq \mathbb{E}\left(\frac{P}{N_0 W}\text{Tr}(HH^*)\right) = \frac{P}{N_0 W}\text{Tr}(HH^*)$$

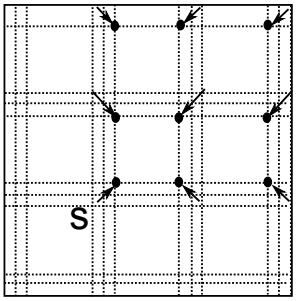

Fig. 4. Moving the nodes to upper bound (14).

Note that

$$\frac{P}{N_0 W}\mathrm{Tr}(HH^*) = \frac{P}{N_0 W}\sum_{k\in S, i\in\mathcal{N}\setminus S}\frac{G}{r_{ik}^{\alpha}}. \qquad (14)$$

The sum is largest if s is located in the middle of the network. Assuming s is located in the middle of the network, consider dividing this summation into 4 equal terms as was done in (10) in Theorem 3.1 by considering the nodes $\mathcal{N}\setminus S$ in each quadrant separately. One simple way to upper bound this summation is to observe that if we move all the M nodes in each cluster (including the source cluster) to the corner of the cluster as indicated by the arrows in Figure 4 we can only increase the summation since we decrease the inter-node distances $r_{ik}, k\in S, i\in\mathcal{N}\setminus S$. This gives us a regular grid where each grid point contains M nodes and the minimal distance of the grid is equal to $\sqrt{MA/n}$. The proof of Theorem 3.1 applies to the current case with SNR_s replaced with SNR_l^c and an additional factor of M^2 since there are M nodes on each grid point (we get a factor of M from the TX side and a factor of M from there RX side). We obtain

$$\frac{P}{N_0 W}\sum_{k\in S, i\in\mathcal{N}\setminus S}\frac{G}{r_{ik}^{\alpha}} \le 4\,B\,M^2\mathrm{SNR}_l^c,$$

where B is bounded in (12). Combined with the fact that independent signalling is optimal within a factor of M^{ϵ} from [7][Lemma 5.2], this completes the proof of the theorem.

Proof of Proposition 3.5: The proof of the proposition follows similarly to the proofs of Theorem 3.2 and Proposition 3.3. When the network is clustered, the performance of multi-hop is limited by the rate achieved over the long hops across the clusters. Therefore, instead hopping over nearest-neighbors we can simplify the strategy by hoping over nearest neighbor clusters since the performance of the strategy is anyway limited by such hops. For example, s can transmit directly to a node in its nearest cluster, this node can then transmit to another node in a neighboring cluster etc. This yields a multi-hop strategy with the hop distance increased to $\sqrt{MA/n}$ instead of $\sqrt{A/n}$ in Theorem 3.2. Since the SNR over the a distance $\sqrt{MA/n}$ is given by SNR_c, this gives the result in (8). The lower bound for the rate achieved by direct transmission simply follows by noting that the largest separation between s and d can be at most $\sqrt{A}$.

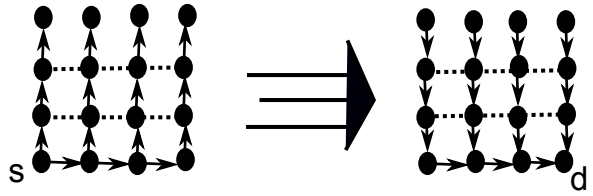

Fig. 5. The three phases of the cluster decode-and-forward strategy.

VI. Cluster Decode-and-Forward

In this section, we describe the cluster decode-and-forward strategy that achieves the performance in Theorem 3.6.

The building block of the strategy is the three phase scheme illustrated in Figure 5. Let us first focus on the case where s and d are located in neighboring clusters. Assume s has M bits to communicate to d. These bits can be communicated in three successive steps.

- s can first distribute its M bits among its M neighbors by using a multi-hop strategy, one bit for each node.
- These nodes together can then form a distributed transmit antenna array, sending the bits simultaneously to the neighboring cluster where d lies.
- Each node in the destination cluster obtained one observation from the MIMO transmission. It can quantize and ship the observation to the destination node d, which can then do joint MIMO processing of all the observations and decode the transmitted bits. The shipping of the quantized observations can be again handled by multi-hop.

We next give a back of the envelope calculation of the rate achieved by this strategy. Note that the bottleneck in the first stage is the output rate from the source node. Following the lines of the proof of Theorem 3.2, it can be shown that s can output bits at a rate

$$\log\left(1+\frac{\mathrm{SNR}_s}{1+K_3\mathrm{SNR}_s}\right)$$

for a constant K_3. Therefore, th phase can be completed in a total amount of time bounded by

$$T_1 = \frac{M}{\log\left(1+\frac{\mathrm{SNR}_s}{1+K_3\mathrm{SNR}_s}\right)}. \qquad (15)$$

Following the lines of [7, Lemma 4.3], it can be show that the MIMO transmissions in the second phase achieve a rate $K_4 M\log\left(1+M\mathrm{SNR}_c\right)$ for a constant K_4, i.e. the rate is of the order of M at high-SNR and of the order of $M^2\mathrm{SNR}_l^c$ at low-SNR. This gives a completion time

$$T_2 = \frac{M}{K_4 M\log\left(1+M\mathrm{SNR}_c\right)}$$

for the second phase. The traffic in the third phase is symmetrical to the first phase. If each observation is quantized into Q bits, it takes $T_3 = QT_1$ time. Assuming that d is able to decode the transmitted bits from its source node from the quantized

signals it gathers by the end of the third phase, the end-to-end communication rate achieved by the scheme is given by

$$\frac{M}{T_1+T_2+T_3}$$

which is of the order of

$$\min\left(\log\left(1+\frac{\text{SNR}_s}{1+K_3\text{SNR}_s}\right),\, K_4\, M\log\left(1+M\text{SNR}_c\right)\right).$$

When s and d are not neighboring clusters the bits of s can be relayed to d in multiple hops, where the above three phase scheme is repeated at each hop. We can designate a node in each intermediate relay cluster to serve as the destination node for the previous hop and the source node for the next hop. This relay node will collect the MIMO observations from the previous hop, process them to decode the transmitted block of M bits, then re-distribute them over the cluster so that they can be relayed by a new MIMO transmission to the next cluster. The operation at the network level can be again organized in three successive phases:

- In a first phase, the "source node" in each cluster on the relaying path from s to d distributes M bits among the M nodes in its cluster. The "source node" is s in the source cluster, and it is the designated relay node in each of the intermediate clusters. While s distributes a new set of M bits among its neighbors, each of the intermediate "source nodes" distribute the bits recovered from the previous hop. All clusters operate at the same time.
- In a second phase, simultaneous MIMO transmissions are performed between neighboring clusters. The picture is similar to the one in Figure 3-(a) and (b), but individual nodes are replaced by clusters of M nodes, and the point-to-point transmissions are replaced by MIMO transmissions between clusters.
- In a third phase, the MIMO observations from the previous hop are collected to the "destination node" in each cluster for joint decoding. The "destination node" is d in the actual destination cluster and it is the designated relay node in each intermediate cluster. All clusters again operate at the same time.

At the end of the these three phases, each of the M bit blocks from s proceed one hop to d. The completion time of the first phase is again given by (15) since all clusters operate at the same time and the bound $K_3\text{SNR}_s$ can be made to account for the total interference. The rate of the MIMO transmissions in the second phase is modified as

$$K_4 M\log\left(1+\frac{M\text{SNR}_c}{1+M\text{SNR}_c}\right)$$

due to the interference between simultaneous MIMO transmissions which as in Theorem 3.2 is of the order of the SNR between neighboring clusters. The details of the proof of Theorem 3.6 can be made precise by following similar lines to [7].

REFERENCES

[1] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, *Wireless Network Information Flow: A Deterministic Approach*, IEEE Trans. Info. Theory, vol. 57, no. 4, pp. 1872-1905, 2011.

[2] A. Özgür, S. Diggavi, *Approximately Achieving Gaussian Relay Network Capacity with Lattice Codes*, Proc. IEEE Int. Symposium on Information Theory, Austin, June 2010.

[3] S. H. Lim, Y.-H. Kim, A. El Gamal, S.-Y. Chung, *Noisy Network Coding*, IEEE Trans. Info. Theory, vol. 57, no. 5, pp. 3132-3152, May 2011.

[4] C. Nazaroglu, A. Özgür, and C. Fragouli, *Wireless Network Simplification: the Gaussian N-Relay Diamond Network* IEEE Int. Symposium on Information Theory (ISIT), St Petersburg, 2011.

[5] A. Raja and P. Viswanath, *Compress-and-Forward Scheme for a Relay Network: Approximate Optimality and Connection to Algebraic Flows*, IEEE Int. Symposium on Information Theory (ISIT) St Petersburg, 2011; e-print http://arxiv.org/abs/1012.0416.

[6] P. Gupta and P. R. Kumar, *The Capacity of Wireless Networks*, IEEE Trans. on Information Theory 42 (2), pp.388-404, 2000.

[7] A. Özgür, O. Lévêque, D. Tse, *Hierarchical Cooperation Achieves Optimal Capacity Scaling in Ad-Hoc Networks*, IEEE Trans. on Information Theory 53 (10), pp.3549-3572, 2007.

[8] A. Özgür, R. Johari, O. Lévêque, D. Tse, *Information Theoretic Operating Regimes of Large Wireless Networks*, IEEE Trans. on Information Theory 56 (1), pp.427-437, 2010.

[9] U. Niesen, P. Gupta, D. Shah, *On Capacity Scaling in Arbitrary Wireless Networks*, IEEE Trans. on Information Theory 55 (9), 3959–3982, September 2009.

[10] U. Niesen, P. Gupta and D. Shah, *The Balanced Unicast and Multicast Capacity Regions of Large Wireless Networks*, IEEE Transactions on Information Theory, 56(5), 2249 - 2271, May 2010.

[11] D. Tse and P. Viswanath. Fundamentals of Wireless Communication. Cambridge University Press, 2005.

[12] U. Niesen, S. Diggavi, *The Approximate Capacity of the Gaussian N-Relay Diamond Network*, IEEE Int. Symposium on Information Theory (ISIT), St Petersburg, 2011.