Capacity Scaling Laws in MIMO Wireless Networks

Rohit U. Nabar[1)], Özgür Oyman[2)], Helmut Bölcskei[1)], and Arogyaswami J. Paulraj[2)]

[1)] Communication Technology Laboratory
Swiss Federal Institute of Technology (ETH) Zürich
ETF E119, Sternwartstrasse 7, 8092 Zürich, Switzerland
Email: {nabar, boelcskei}@nari.ee.ethz.ch

[2)] Information Systems Laboratory, Stanford University
228 Packard, 350 Serra Mall, Stanford, CA 94305
Email: {oyman, apaulraj}@stanford.edu

Abstract

The use of multiple antennas at both ends of a wireless link, popularly known as multiple-input multiple-output (MIMO) wireless, has been shown to offer significant improvements in *spectral efficiency* and *link reliability* through *spatial multiplexing* and *space-time coding*, respectively. This paper demonstrates that similar performance gains can be obtained in wireless relay and adhoc networks employing terminals with MIMO capability. In the relay case a source terminal communicates with a destination terminal assisted by multiple relay terminals. For this scenario, assuming that transmitter and receiver employ M antennas and operate in spatial multiplexing mode, we show that the network capacity scales as $C = \frac{M}{2}\log(KN) + O(1)$ for a large number of relay terminals K and large number of antennas $N \geq M$ at each of the relay terminals. For finite $N \geq M$, we find that capacity scales as $C = \frac{M}{2}\log(K) + O(1)$. Furthermore, we propose an asymptotically optimal architecture which requires that each of the relay terminals knows its backward and forward channels. Our results are extended to the adhoc case where L source-destination pairs communicate concurrently in spatial multiplexing mode through the same set of relay terminals and sum-capacity is shown to scale as $C = \frac{LM}{2}\log(KN) + O(1)$ for large K and N and $C = \frac{LM}{2}\log(K) + O(1)$ for finite $N \geq LM$. Finally, we establish the importance of stream separation and coherent combining at the relay terminals, in the absence of which we show that for any N, asymptotically in K, $C = \frac{M}{2}\log(\text{SNR}) + O(1)$, demonstrating that the number of relays does not enter the scaling law.

1 Introduction and Outline

Wireless networks can broadly be divided into two different categories based on their architecture: In a *cellular* network, all terminals communicate directly with a base-station which controls all transmissions and forwards data to the intended users. In an *adhoc* network, all terminals are typically on an equal footing, having the same capabilities and responsibilities and communication takes place in a *decentralized* fashion. In this paper, we focus on the latter case, where two terminals communicate by routing their data through other (relay) terminals. Consequently, every terminal can act both as a sender/receiver of data and as a *relay* for other transmissions. Any kind of cooperation between the terminals is in general permitted.

Wireless channels are impaired by random fluctuations in signal level known as *fading* and by *co-channel interference*. The use of multiple antennas at both ends of a point-to-point wireless link, known as *multiple-input multiple-output* (MIMO) wireless, is a powerful performance enhancing technology [1]-[4]. First results on MIMO in broadcast and multiple-access channels have been reported in [5]-[11]. In more complex systems such as relay and adhoc networks, it is not yet clear how to exploit the spatial degrees of freedom offered by multiple antenna terminals and what the possible performance improvements would be over single-input single-output (SISO) networks. This paper is an attempt to quantify the impact of MIMO technology on wireless relay and adhoc networks.

MIMO gains in point-to-point links. In point-to-point wireless links MIMO systems improve spectral efficiency and link reliability through *spatial multiplexing gain* and *diversity gain*, respectively. We shall briefly describe these gains in the following:

- *Spatial multiplexing* in MIMO systems yields a *linear* (in the minimum of the number of transmit and receive antennas) increase in capacity for no additional power or bandwidth expenditure [1]-[4]. The corresponding gain is realized by simultaneously transmitting independent data streams in the same frequency band. Under conducive channel conditions (such as rich scattering), the receiver exploits differences in the spatial signatures of the multiplexed streams to separate the different signals, thereby realizing a capacity gain.

- *Diversity* [12] is a powerful technique to mitigate fading and increase robustness to interference. Diversity techniques rely on transmitting the data signal over multiple (ideally) *independently* fading paths (time/frequency/space). Spatial (i.e., antenna) diversity is particularly attractive when compared to time/frequency diversity since it does not incur an expenditure in transmission time/bandwidth. Space-time coding to exploit spatial diversity gain in point-to-point MIMO channels has been studied extensively [13, 14, 15].

Contributions. The aim of this paper is to investigate and quantify the impact of multiplexing and diversity gains on the capacity of wireless relay and adhoc networks. We attack this problem by deriving the asymptotic (in the number of network terminals and possibly in the number of terminal antennas) network capacity, a framework pioneered for single antenna systems and AWGN channels in [16, 17]. Results of the Gupta-Kumar [16, 18] and Gastpar-Vetterli [17] type have recently been reported for single antenna fading channels in [19] and for the case of node mobility in [20]. Throughout this paper, we focus on the case where the transmitters operate in spatial multiplexing mode, i.e., statistically independent signals are transmitted from different antennas. In the relay case a single source terminal equipped with M antennas spatially multiplexes data to a single destination terminal with M antennas through K relay terminals, with N antennas each, using a one-hop relaying protocol. In the adhoc case L source-destination terminal pairs communicate simultaneously using one-hop relaying over the same set of relay terminals. For both scenarios we derive upper and lower bounds on network capacity and show that employing power control these bounds can be made to meet asymptotically. Our results indicate that throughput as well as link reliability are substantially improved through the use of MIMO terminals. The detailed contributions reported in this paper can be summarized as follows:

- We derive an *upper bound* on the capacity of the MIMO relay network in the relay case (assuming one-hop relaying) based on the "*cut-set bound*" in [21]. Assuming each relay terminal knows its backward and forward channels, we then propose a *zero-forcing amplify-and-forward relaying architecture* which leads to a *lower bound on network capacity.* We then conclude that for $N \geq M$, asymptotically in K and N capacity grows *linearly in the number of multiplexed streams* M and *logarithmically* in KN, showing that *multiplexing gain* and *distributed array gain*, respectively, are realized. Our upper and lower bounds differ in an $O(1)$ term and can be shown to meet if power control is employed.

- We establish the importance of *stream separation* and *coherent combining* (achieved through transmit and receive zero-forcing) at the relay terminal, in the absence of which *spatial multiplexing gain is realized* but *distributed array gain is not obtained.*

- Finally, we consider the adhoc case with $N \geq LM$ and show that the sum-capacity grows *linearly* in LM and *logarithmically* in KN, establishing that *multi-user multiplexing gain* and *distributed array gain* are realized.

Notation. The superscripts ^{T} and ^{H} stand for transposition and conjugate transpose, respectively. $\mathcal{E}$ denotes the expectation operator. $\mathrm{Tr}(\mathbf{A})$, $\det(\mathbf{A})$ and $\|\mathbf{A}\|_F$ stand for the trace, determinant and Frobenius norm of $\mathbf{A}$ respectively. $\|\mathbf{a}\|$ stands for the Euclidean norm of the vector $\mathbf{a}$. $\mathbf{I}_m$ denotes the $m \times m$ identity matrix. A circularly symmetric complex Gaussian random variable is a random variable $Z = X + j\,Y \sim \mathcal{CN}(0, \sigma^2)$, where X and Y are i.i.d. $\mathcal{N}(0, \sigma^2/2)$. The chi-squared distribution with N degrees of freedom, denoted as χ^2_N, is defined as the distribution of the sum of the squared magnitudes of N i.i.d. $\mathcal{CN}(0, 1)$ random variables. Throughout the paper all logarithms are to the base 2.

Outline. The rest of this paper is organized as follows. Section 2 describes the channel and signal models for the MIMO relay network. In Section 3, we derive a cut-set upper bound on the capacity of the MIMO relay network. Section 4 describes a practical relaying protocol, based on which in Section 5, we derive a lower bound on the capacity of the MIMO relay network. Section 6 describes extensions to the adhoc setting. In Section 7, we discuss the impact of stream separation and coherent combining on capacity of the MIMO relay network. Finally, we conclude in Section 8.

2 Channel and Signal Model for the Relay Case

General assumptions. We consider a wireless network consisting of $K + 2$ multi-antenna terminals, with one active source-destination pair and K relay terminals. We denote the source terminal by $\mathcal{S}$, the destination terminal by $\mathcal{D}$, and the k-th relay terminal by $\mathcal{R}_k$ $(k = 1, 2, ..., K)$. The source and destination terminals are equipped with M antennas each, while each of the relay terminals employs N transmit/receive antennas. We assume that no direct link between the source and destination terminals exists and transmission takes place in two hops (a.k.a. one-hop relaying) over two separate time slots. In the first time slot, the relay terminals receive the signal transmitted from the source terminal. After processing the received signals, the relay terminals simultaneously transmit the processed data to the destination terminal during the second time slot. We assume that the relay terminals are located randomly and independently in a domain of fixed area.

Channel and signal model. Throughout the paper, frequency-flat fading over the bandwidth of interest and perfectly synchronized transmission/reception between the terminals is assumed. The input-output relation for the $\mathcal{S} \rightarrow \mathcal{R}_k$ link[1] is given by

$$\mathbf{r}_k = \sqrt{\frac{E_k}{M}}\,\mathbf{H}_k\mathbf{s} + \mathbf{n}_k, \quad k = 1, 2, ..., K,$$

where $\mathbf{r}_k$ is the $N \times 1$ received vector signal, E_k is the average energy available at the source terminal over a symbol period (having accounted for path loss and shadowing in the $\mathcal{S} \rightarrow \mathcal{R}_k$ link), $\mathbf{H}_k$ is the $N \times M$ corresponding channel matrix consisting of i.i.d. $\mathcal{CN}(0,1)$ entries, $\mathbf{s}$ is the zero-mean $M \times 1$ circularly symmetric complex Gaussian transmit signal vector satisfying $\mathcal{E}\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}_M$ (recall that the signaling mode is spatial multiplexing), and $\mathbf{n}_k$ is the $N \times 1$ spatio-temporally white circularly symmetric complex Gaussian noise vector at the relay terminal with covariance matrix $\mathcal{E}\{\mathbf{n}_k\mathbf{n}_k^H\} = \sigma_n^2\mathbf{I}_N$.

Each relay terminal processes its received vector signal $\mathbf{r}_k$ to produce an $N \times 1$ vector signal denoted as $\mathbf{t}_k$, which is then transmitted to the destination terminal over the second time slot. The destination terminal receives the $M \times 1$ vector signal

$$\mathbf{y} = \sum_{k=1}^{K} \sqrt{\frac{P_k}{N}}\,\mathbf{G}_k\mathbf{t}_k + \mathbf{z}, \tag{1}$$

where P_k is the average energy available at the k-th relay terminal over a symbol period (having accounted for path loss and shadowing in the $\mathcal{R}_k \rightarrow \mathcal{D}$ link), $\mathbf{G}_k$ is the corresponding $M \times N$ i.i.d. $\mathcal{CN}(0,1)$ channel matrix and $\mathbf{z} = [z_1\ z_2\ ...\ z_M]^T$ is the $M \times 1$ spatio-temporally white circularly symmetric complex Gaussian noise vector at the destination terminal satisfying $\mathcal{E}\{\mathbf{z}\mathbf{z}^H\} = \sigma_n^2\mathbf{I}_M$. The transmit signal vectors $\mathbf{t}_k$ will in general depend both on the forward and backward channels, $\mathbf{G}_k$ and $\mathbf{H}_k$, respectively, and are chosen to satisfy the average (conditioned on the backward and forward channels) power constraint $\mathcal{E}\{\|\mathbf{t}_k\|^2|\mathbf{H}_k, \mathbf{G}_k\} \leq N$.

As already mentioned above, throughout the paper, the path-loss and the shadowing statistics are captured by $\{E_k\}_{k=1}^K$ (for the first hop) and $\{P_k\}_{k=1}^K$ (for the second hop). We assume that these parameters are random, independently drawn from the same distribution, strictly positive, bounded above, and remain constant over the entire time period of interest. The exact statistics will in general depend on the cell topology. Additionally, we assume an ergodic block fading channel model [22] such that the channel matrices $\mathbf{H}_k$ and $\mathbf{G}_k$ remain constant over the entire duration of a time slot and change in an independent fashion across time slots. Finally, we assume that each relay terminal $\mathcal{R}_k$ has perfect knowledge of its forward and backward channels, $\mathbf{G}_k$ and $\mathbf{H}_k$, respectively, and the destination terminal $\mathcal{D}$ has perfect knowledge of all channel matrices (asymptotically in K, knowledge of the composite $M \times M$ channel at the destination terminal will suffice).

3 Upper Bound on MIMO Relay Network Capacity

In this section, we present an asymptotic (in the number of relay terminals K) upper bound on the MIMO relay network capacity. In Section 4, we introduce a relaying architecture which leads to a corresponding lower bound derived in Section 5.

Theorem 1. *In the large relay limit $K \rightarrow \infty$, the capacity of the MIMO relay network is upper bounded by*

$$C_{upper}^{\infty} = \frac{M}{2}\log\left(\frac{KN\mathcal{E}\{E_k\}}{M\sigma_n^2}\right) \quad \text{bps/Hz}.$$

[1] $\mathcal{A} \rightarrow \mathcal{B}$ signifies communication from terminal $\mathcal{A}$ to terminal $\mathcal{B}$.

Proof: Separating the source terminal $\mathcal{S}$ from the rest of the network, using a broadcast cut [21], and applying the cut-set theorem [21, Th. 14.10.1] it follows that the capacity of the MIMO relay network is upper bounded by

$$C_u = \mathcal{E}_{\{\mathbf{H}_k, \mathbf{G}_k\}_{k=1}^K} \left\{ \frac{1}{2} I(\mathbf{s}; \mathbf{r}_1, ..., \mathbf{r}_K, \mathbf{y} | \mathbf{t}_1, ..., \mathbf{t}_K) \right\},$$

where the factor 1/2 results from the fact that data is transmitted over two time slots (corresponding to two hops). Recalling that $\mathbf{s}$ is circularly symmetric complex Gaussian with $\mathcal{E}\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}_M$, the upper bound on capacity is given by [23]

$$C_u = \mathcal{E}_{\{\mathbf{H}_k\}_{k=1}^K} \left\{ \frac{1}{2} \log \left(\det \left(\mathbf{I}_M + \frac{1}{\sigma_n^2} \sum_{k=1}^K \frac{E_k}{M} \mathbf{H}_k^H \mathbf{H}_k \right) \right) \right\}. \tag{2}$$

Now, applying Jensen's inequality to (2) it follows that

$$C_u \leq \underbrace{\frac{M}{2} \log \left(1 + \frac{N}{M\sigma_n^2} \sum_{k=1}^K E_k \right)}_{C_{upper}}.$$

Finally, noting that for $K \to \infty$, applying the strong law of large numbers [24]

$$\frac{1}{K} \sum_{k=1}^K E_k \xrightarrow{\text{wp1}} \mathcal{E}\{E_k\}$$

and hence by [25, Th. 1.7]

$$C_{upper} \xrightarrow{\text{wp1}} \frac{M}{2} \log \left(1 + \frac{KN\mathcal{E}\{E_k\}}{M\sigma_n^2} \right) \approx \frac{M}{2} \log \left(\frac{KN\mathcal{E}\{E_k\}}{M\sigma_n^2} \right) = C_{upper}^{\infty}, \tag{3}$$

which explicitly shows the logarithmic scaling in KN and linear scaling in the number of source and destination terminal antennas M. □

Discussion. It is straightforward to see that physically the right-hand-side (RHS) of (2) results when all the relay terminals can fully cooperate and know all the $\mathbf{H}_k$ perfectly so as to effectively form a point-to-point coherent MIMO channel with M transmit antennas and KN receive antennas, and can further convey information in a lossless fashion to the destination terminal. Even though our upper bound holds for finite N, in the following we shall be interested in its asymptotic behavior in both K and N which follows from (3) as

$$C_{upper}^{\infty,\infty} = \lim_{N \to \infty} C_{upper}^{\infty} = \frac{M}{2} \log(KN) + O(1). \tag{4}$$

4 A Zero-Forcing Relaying Architecture

In this section, we introduce a relaying architecture based on which a lower bound on the MIMO relay network capacity is derived in Section 5. In the following, we shall assume that the relay terminals employ at least as many antennas as the source and destination terminals, i.e., $N \geq M$. Each relay terminal performs zero-forcing (ZF) on its received signal $\mathbf{r}_k$ and normalizes the average energy (conditioned on $\mathbf{H}_k$) in each of the multiplexed streams to unity resulting in

$$\mathbf{u}_k = \left(\frac{E_k}{M} \right)^{-1/2} \operatorname{diag} \left\{ \frac{1}{\sqrt{1 + \left(\frac{E_k}{M\sigma_n^2} \right)^{-1} \|\mathbf{b}_{k,i}\|^2}} \right\}_{i=1}^M \mathbf{H}_k^{\dagger} \mathbf{r}_k,$$

where $\mathbf{H}_k^\dagger = (\mathbf{H}_k^H \mathbf{H}_k)^{-1}\mathbf{H}_k^H$, $\mathbf{u}_k$ is the $M \times 1$ processed received signal vector at terminal $\mathcal{R}_k$ and $\mathbf{b}_{k,i}$ denotes the i-th $(i = 1, 2, ..., M)$ row of $\mathbf{H}_k^\dagger$. The resulting processed streams are spatially pre-filtered so that $\mathbf{t}_k = \mathbf{W}_k \mathbf{u}_k$ and relayed to the destination terminal to finally give

$$\mathbf{y} = \sum_{k=1}^{K} \sqrt{\frac{P_k}{N}}\, \mathbf{G}_k \mathbf{W}_k \mathbf{u}_k + \mathbf{z}.$$

Recall that we imposed a power constraint on $\mathbf{t}_k$ requiring that

$$\mathcal{E}\left\{\|\mathbf{W}_k \mathbf{u}_k\|^2 \,|\mathbf{H}_k, \mathbf{G}_k\right\} \leq N. \tag{5}$$

Using $\|\mathbf{W}_k \mathbf{u}_k\|^2 \leq \|\mathbf{W}_k\|_F^2 \,\|\mathbf{u}_k\|^2$ [26] with $\mathcal{E}\{\|\mathbf{u}_k\|^2\} = M$ it follows immediately that normalizing $\mathbf{W}_k$ such that $\|\mathbf{W}_k\|_F^2 = N/M$ ensures that (5) is met for the $\mathcal{R}_k \rightarrow \mathcal{D}$ link irrespectively of the channel realizations $\mathbf{H}_k$ and $\mathbf{G}_k$. We choose $\mathbf{W}_k$ such that the forward channel $\mathbf{G}_k$ is pre-equalized which combined with the normalization $\|\mathbf{W}_k\|_F^2 = \frac{N}{M}$ yields

$$\mathbf{W}_k = \frac{\sqrt{N}}{M}\begin{bmatrix} \frac{\mathbf{c}_{k,1}}{\|\mathbf{c}_{k,1}\|} & \frac{\mathbf{c}_{k,2}}{\|\mathbf{c}_{k,2}\|} & \cdots & \frac{\mathbf{c}_{k,M}}{\|\mathbf{c}_{k,M}\|} \end{bmatrix},$$

where $\mathbf{c}_{k,i}$ $(i = 1, 2, ..., M)$ is the i-th column of $\mathbf{G}_k^\dagger = \mathbf{G}_k^H(\mathbf{G}_k \mathbf{G}_k^H)^{-1}$. It now follows that the signal received at the destination terminal corresponding to the i-th multiplexed stream s_i is given by

$$y_i = \left(\sum_{k=1}^{K} d_{k,i}\right) s_i + \sum_{k=1}^{K} d_{k,i}\, \tilde{n}_{k,i} + z_i, \quad i = 1, 2, ..., M,$$

where

$$d_{k,i} = \sqrt{\frac{P_k X_{k,i}}{M^2\left(1 + \left(\frac{E_k}{M\sigma_n^2}\right)^{-1}\frac{1}{Y_{k,i}}\right)}} \tag{6}$$

with $X_{k,i} = 1/\|\mathbf{c}_{k,i}\|^2$, $Y_{k,i} = 1/\|\mathbf{b}_{k,i}\|^2$ and $\tilde{n}_{k,i}$ denoting the i-th element of the $M \times 1$ vector $\tilde{\mathbf{n}}_k = \left(\frac{E_k}{M}\right)^{-1/2}\mathbf{H}_k^\dagger \mathbf{n}_k$. Note that $\tilde{n}_{k,i}|\mathbf{H}_k \sim \mathcal{CN}\left(0, \left(\frac{E_k}{M\sigma_n^2}\right)^{-1}\frac{1}{Y_{k,i}}\right)$.

Interpretation as distributed beamforming. From [27] it is known that $X_{k,i} \sim \chi^2_{N-M+1}$ and $Y_{k,i} \sim \chi^2_{N-M+1}$. Next, noting

$$y_i = \sqrt{\frac{P_k}{M^2}} \sum_{k=1}^{K} \sqrt{X_{k,i}}\, u_{k,i} + z_i, \quad i = 1, 2, ..., M,$$

where $u_{k,i}$ is the i-th element of $\mathbf{u}_k$, it follows that the individual (data) streams forwarded from the relay terminals combine coherently at the receiver. Our ZF relaying architecture can therefore be viewed as performing distributed transmit maximum ratio combining (MRC).

5 Lower Bound on MIMO Relay Network Capacity

Based on the ZF relaying architecture described in the previous section, we are now ready to derive an asymptotic lower bound on the capacity of the MIMO relay network.

Theorem 2. *In the large relay and large antenna limit $K \to \infty$, $N \to \infty$ the capacity of the MIMO relay network is lower bounded by*

$$C_{lower}^{\infty,\infty} = \frac{M}{2} \log \left(\frac{K(N-M+1)(\mathcal{E}\{\sqrt{P_k}\})^2}{M\sigma_n^2 \mathcal{E}\{\frac{P_k}{E_k}\}} \right) \quad \text{bps/Hz.}$$

Proof: The relaying architecture proposed in Section 4 effectively decouples the MIMO channel into M parallel SISO channels. With independent encoding/decoding on each stream, we have

$$C_{lower} = \sum_{i=1}^{M} \mathcal{E}_{\{\mathbf{H}_k, \mathbf{G}_k\}_{k=1}^K} \{I_i\},$$

where

$$I_i = \frac{1}{2} \log \left(1 + \frac{\left(\sum_{k=1}^K d_{k,i}\right)^2}{\sigma_n^2 \left(1 + \sum_{k=1}^K f_{k,i}^2\right)} \right), \qquad i = 1, 2, ..., M,$$

denotes the mutual information of the i-th multiplexed stream, $d_{k,i}$ was defined in (6) and

$$f_{k,i} = \sqrt{\frac{P_k X_{k,i}}{M^2 \left(\sigma_n^2 + \frac{E_k Y_{k,i}}{M}\right)}}.$$

Here, we assumed that the receiver has perfect knowledge of the individual[2] $d_{k,i}$, knowledge of $\mathbf{G}_k$ and $\mathbf{H}_k$ is not required. Next, we rewrite

$$I_i = \frac{1}{2} \log \left(1 + \frac{K^2 \left(\frac{1}{K} \sum_{k=1}^K d_{k,i}\right)^2}{\sigma_n^2 \left(1 + K \left(\frac{1}{K} \sum_{k=1}^K f_{k,i}^2\right)\right)} \right),$$

and the variables $d_{k,i}$ and $f_{k,i}$ as

$$d_{k,i} = \sqrt{\frac{P_k (N-M+1) \left(\frac{X_{k,i}}{N-M+1}\right)}{M^2 \left(1 + \frac{1}{N-M+1} \left(\frac{E_k}{M\sigma_n^2}\right)^{-1} \left(\frac{Y_{k,i}}{N-M+1}\right)^{-1}\right)}},$$

$$f_{k,i} = \sqrt{\frac{P_k (N-M+1) \left(\frac{X_{k,i}}{N-M+1}\right)}{M^2 \left(\sigma_n^2 + (N-M+1) \left(\frac{E_k}{M}\right) \left(\frac{Y_{k,i}}{N-M+1}\right)\right)}}.$$

By the strong law of large numbers [24] we have $X_{k,i}/(N-M+1) \xrightarrow{\text{wp1}} 1$ and $Y_{k,i}/(N-M+1) \xrightarrow{\text{wp1}} 1$ for $N \to \infty$. Using the fact that P_k and E_k are i.i.d. across k and uniformly bounded, we have

$$\frac{1}{K} \sum_{k=1}^K \sqrt{P_k} \xrightarrow{\text{wp1}} \mathcal{E}\left\{\sqrt{P_k}\right\}, \qquad \frac{1}{K} \sum_{k=1}^K \frac{P_k}{E_k} \xrightarrow{\text{wp1}} \mathcal{E}\left\{\frac{P_k}{E_k}\right\}.$$

[2]If only the composite channel $\sum_{k=1}^K d_{k,i} (i = 1, 2, ..., M)$ is known to the receiver, network capacity can be lower bounded through a Gaussian approximation of the noise term [28]. However, in the large relay limit $K \to \infty$, through the central limit theorem [24], the noise, suitably normalized, will in fact be Gaussian and knowledge of the composite channel will suffice.

Combining our results, and again using [25, Th. 1.7], we finally obtain[3]

$$C_{lower} \xrightarrow{\text{wp1}} \frac{M}{2} \log\left(\frac{K(N-M+1)\left(\mathcal{E}\{\sqrt{P_k}\}\right)^2}{M\sigma_n^2 \mathcal{E}\{\frac{P_k}{E_k}\}}\right) = C_{lower}^{\infty,\infty}. \quad (7)$$

□

Discussion. We can now conclude that asymptotically[4] in the number of relay terminals K and number of relay antennas N

$$C_{lower}^{\infty,\infty} = \frac{M}{2}\log(KN) + O(1). \quad (8)$$

Combining (4) and (8) we have thus established that asymptotically the MIMO relay network capacity scales linearly in the number of multiplexed streams M and logarithmically in KN. In MIMO terminology we can say that the network exhibits a multiplexing gain of $\frac{M}{2}$ (accounting for transmission over two time slots) and an array gain of KN. Hence, the loss in spectral efficiency incurred by one-hop relaying can be compensated by employing $M \geq 2$ antennas at the source and destination terminals. The total array gain results from a per-relay terminal array gain of N and distributed array gain of K. Furthermore, we can conclude that our analysis shows that the simple ZF relaying architecture proposed in Section 4 is asymptotically optimal. Finally, we note that employing power control so that $P_k = P$ and $E_k = E$ are constant for all k, the $O(1)$ terms in (4) and (8) are equal and we obtain the stronger result $C_{upper}^{\infty,\infty} = C_{lower}^{\infty,\infty}$.

The finite N case. So far, we considered asymptotics in both K and N. We note however that for finite $N \geq M$ it can be shown that in the large K limit [23]

$$C_{upper}^{\infty} = \frac{M}{2}\log(K) + O(1) \quad (9)$$

$$C_{lower}^{\infty} = \frac{M}{2}\log(K) + O(1). \quad (10)$$

We can thus conclude that for $K \to \infty$ we continue to have a multiplexing gain of $\frac{M}{2}$, but the array gain is K rather than KN, i.e., we have only distributed beamforming gain. Note however that the $O(1)$ terms in (9) and (10) will be functions of N.

Numerical example. This example serves to demonstrate the impact of power control on the $O(1)$ terms in (4) and (8). We consider a relay network with $M = 2$ and P_k/σ_n^2 and E_k/σ_n^2 are drawn independently from a truncated log-normal distribution (modeling macroscopic fading) with mean 10dB and standard deviation (of the underlying log-normal distribution) σdB. The truncation level reflects the quality of power control (the underlying log-normal distribution is renormalized accordingly). We denote the truncation margin by τdB ($\tau \geq 0$) so that P_k/σ_n^2 and E_k/σ_n^2 are bounded above and below by $(10 \pm \tau)$dB. Tight power control results in small τ and vice-versa. Fig. 1 plots the difference of the $O(1)$ terms in $C_{upper}^{\infty,\infty}$ and $C_{lower}^{\infty,\infty}$ as a function of τ for $\sigma = 4$ and $\sigma = 6$. Note that the $O(1)$ term of $C_{upper}^{\infty,\infty}$ in the above setup is independent of τ and σ. As expected, the capacity difference is 0 when $\tau = 0$ (corresponding to perfect power control) and increases with increasing τ (i.e., poorer power control). The capacity difference increases with increasing σ due to an increased spread in the individual power levels.

[3] We note that strictly speaking, C_{lower} converges to an expression that is well approximated by the RHS of (7), asymptotically.

[4] In [23] it is shown that the $O(1)$ term in $C_{lower}^{\infty,\infty}$ is upper bounded by the $O(1)$ term in $C_{upper}^{\infty,\infty}$.

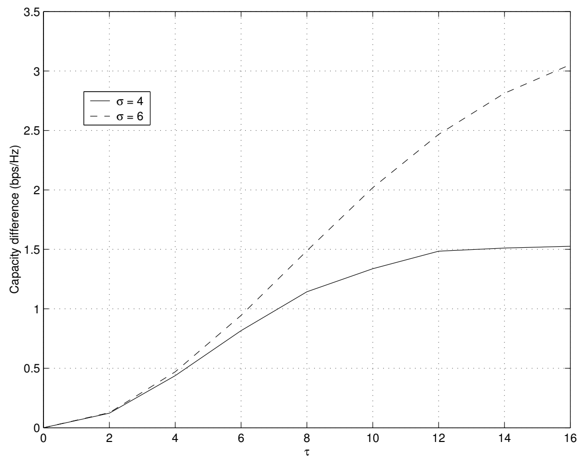

Figure 1: *Impact of power control on the difference between the* $O(1)$ *terms in* $C^{\infty,\infty}_{upper}$ *and* $C^{\infty,\infty}_{lower}$.

6 Extension to the Multi-User Case

The analysis presented in Sections 3-5 carries over to the adhoc case where L source-destination pairs communicate simultaneously with each other in spatial multiplexing mode through K (common) relay terminals using one-hop relaying. We assume that no direct links exist between the individual source-destination pairs. Each source and destination terminal is equipped with M antennas, while the relay terminals employ $N \geq LM$ antennas each. We note that the latter assumption guarantees that the relay terminals can perform stream and user separation. In the first time interval, the K relay terminals receive simultaneously the data signals from all the source terminals. In the second time slot, the received signals (after appropriate processing) are simultaneously relayed to the destination terminals.

Foregoing details, we briefly discuss a ZF relaying architecture for the adhoc case. Again the relay terminals employ ZF reception to decouple the LM multiplexed streams, thereby eliminating multi-stream as well as multi-user interference. As in the single user scenario described in Section 4 the average energy in each of the processed streams is normalized to unity. Transmit pre-processing consists of each relay terminal zero-forcing its forward channel, which essentially amounts to null-steering used in the context of Space Division Multiple Access (SDMA) [29] to spatially separate subscribers operating in the same time and frequency slots in the downlink of a cellular network.

Assuming that each relay terminal knows all its L pairs of forward and backward channels perfectly, the destination terminals know their corresponding individual $d_{k,i}$ (the same comments made in Footnote 2 apply here), perfectly the resultant channel from each relay terminal, and the path-loss/shadowing statistics are identical for all source-destination pairs, it can be shown that the *sum-capacity* satisfies

$$\begin{aligned} C^{\infty,\infty}_{MU,lower} &= \frac{LM}{2}\log(KN) + O(1) \\ C^{\infty,\infty}_{MU,upper} &= \frac{LM}{2}\log(KN) + O(1). \end{aligned}$$

This establishes that asymptotically the adhoc MIMO network capacity scales linearly in the total number of streams multiplexed across the network and logarithmically in KN. We note that asymptotically the information rate for each source-destination pair is $1/L$-th of the sum-capacity. Again, if perfect power control is employed, the $O(1)$

terms in the upper and lower bounds can be made equal and $C^{\infty,\infty}_{MU,upper} = C^{\infty,\infty}_{MU,lower}$. We conclude this section by noting that as in the case of the single user scenario, for finite $N \geq LM$ and $K \to \infty$, the sum-capacity can be shown to scale as $C = \frac{LM}{2}\log(K)+O(1)$.

7 A Simple Amplify-and-Forward Architecture

So far, we considered a relaying architecture requiring that each relay terminal knows its corresponding forward and backward channels perfectly. In the following, we relax this assumption and investigate a simple amplify-and-forward (AF) architecture where relay terminals simply forward a scaled version of the received signal without additional processing. Consequently, we are also able to relax the assumption $N \geq M$ to $N \geq 1$. For the sake of simplicity of exposition, we focus on the single user (i.e., relay) scenario, originally described in Sections 2-5.

AF architecture. The signal transmitted by the k-th relay terminal is given by $\mathbf{t}_k = (E_k + \sigma_n^2)^{-1/2}\mathbf{r}_k$, where $(E_k + \sigma_n^2)$ is the average (over the channel) energy received at each antenna of the relay terminal on the $\mathcal{S} \to \mathcal{R}_k$ link and $\mathbf{r}_k$ is the corresponding received signal[5]. It follows from (1) that the signal received at the destination terminal is given by

$$\mathbf{y} = \sum_{k=1}^{K}\sqrt{\frac{P_k}{N(E_k+\sigma_n^2)}}\mathbf{G}_k\mathbf{r}_k + \mathbf{z}. \tag{11}$$

Dividing the RHS and left-hand-side (LHS) of (11) by $\sqrt{K}$ and simplifying we get

$$\frac{\mathbf{y}}{\sqrt{K}} = \underbrace{\frac{1}{\sqrt{K}}\sum_{k=1}^{K}\left(\sqrt{\frac{P_kE_k}{NM(E_k+\sigma_n^2)}}\mathbf{G}_k\mathbf{H}_k\right)}_{\mathbf{A}}\mathbf{s} + \underbrace{\frac{1}{\sqrt{K}}\sum_{k=1}^{K}\left(\sqrt{\frac{P_k}{N(E_k+\sigma_n^2)}}\mathbf{G}_k\mathbf{n}_k\right) + \frac{\mathbf{z}}{\sqrt{K}}}_{\mathbf{b}},$$

or equivalently $\mathbf{y}/\sqrt{K} = \mathbf{A}\mathbf{s} + \mathbf{b}$. In the large relay limit $K \to \infty$, applying the central limit theorem [24] it follows that asymptotically in K

$$\begin{aligned}
[\mathbf{A}]_{i,j}\,(i,j=1,2,...,M) &\sim \text{ i.i.d. } \mathcal{CN}\left(0, \frac{1}{M}\mathcal{E}\left\{\frac{P_kE_k}{E_k+\sigma_n^2}\right\}\right), \\
b_i\,(i=1,2,...,M) &\sim \text{ i.i.d. } \mathcal{CN}\left(0, \sigma_n^2\mathcal{E}\left\{\frac{P_k}{E_k+\sigma_n^2}\right\}\right),
\end{aligned}$$

where $[\mathbf{A}]_{i,j}$ denotes the element in the i-th row and j-th column of $\mathbf{A}$ and b_i is the i-th element of $\mathbf{b}$. Combining these results with the fact that $I(\mathbf{y};\mathbf{s}) = I(\mathbf{y}/\sqrt{K};\mathbf{s})$, and assuming that the receiver knows the compound channel $\mathbf{A}$ we obtain the capacity of the AF architecture in the large relay limit as

$$C^{\infty}_{AF} = \frac{1}{2}\,\mathcal{E}_{\mathbf{H}_w}\left\{\log\left(\mathbf{I}_M + \frac{\rho}{M}\mathbf{H}_w\mathbf{H}_w^H\right)\right\}, \tag{12}$$

where the elements of $\mathbf{H}_w$ are i.i.d. $\mathcal{CN}(0,1)$ and $\rho = \mathcal{E}\left\{\frac{E_k}{E_k+\sigma_n^2}\right\}/\left(\sigma_n^2\mathcal{E}\{\frac{1}{E_k+\sigma_n^2}\}\right)$ can be interpreted as an effective SNR. It is interesting to observe that $C^{\infty}_{AF} = \frac{1}{2}C_{M\times M}$, where $C_{M\times M}$ denotes the capacity of an $M\times M$ i.i.d. Gaussian MIMO channel with no transmit

[5]Note that we are employing an average (over the channel) energy normalization which is more relaxed than the power constraint in (5).

channel state information (CSI) and perfect receive CSI. In the high SNR regime $\rho \gg 1$, we have from [30]

$$C_{AF}^{\infty} \approx \frac{M}{2} \log\left(\frac{\rho}{M}\right) + \frac{\log e}{2} \left(\sum_{j=1}^{M} \sum_{p=1}^{M-j} \frac{1}{p} - \gamma M \right),$$

where $\gamma \approx 0.57721$ is Euler's constant. We can now draw a number of interesting conclusions.

- The simple AF architecture realizes the same multiplexing gain ($\frac{M}{2}$) as the ZF-based relaying architecture. It is remarkable that multiplexing gain can be obtained even if the relays do not have channel knowledge and do not employ multiple antennas (recall that $N \geq 1$). Numerical evidence of this fact has been presented in [31].
- The absence of channel knowledge at the relay terminals results in a lack of distributed beamforming gain reflected by the fact that C_{AF}^{∞} is independent of K. Note however, that (12) implies receive array gain (independent of K) resulting from the fact that the receiver knows the compound MIMO channel.

Numerical example. Fig. 2 shows the capacity (obtained through Monte Carlo simulation) of the simple AF architecture as a function of the number of relays K for $N = 1$ and $M = 2, 4$ along with the large K capacity limit as predicted by (12). For finite K, knowledge of the individual $\mathbf{G}_k$ is assumed at the destination terminal, rendering the noise conditionally Gaussian. We assume perfect power control so that $E_k/\sigma_n^2 = P_k/\sigma_n^2 =$ 20dB. We can see that the capacity of the AF architecture converges very quickly to C_{AF}^{∞}.

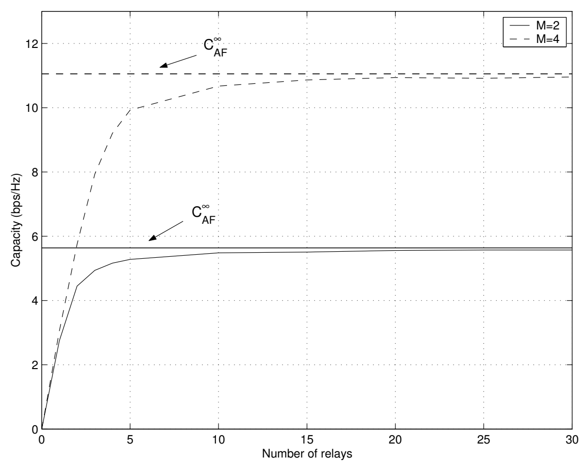

Figure 2: *Capacity vs. number of relays for the AF architecture.*

8 Conclusions

We studied capacity scaling laws in MIMO relay and adhoc networks where communication between the source and destination terminals occurs using one-hop relaying. For M antennas at the source and destination terminals and perfect channel state information in the relays we showed that asymptotically in the number of relay terminals K and number of antennas at each relay terminal N capacity scales as $C = \frac{M}{2}\log(KN) + O(1)$ in the relay case and as $C = \frac{LM}{2}\log(KN) + O(1)$ in the adhoc case with L source-destination pairs. For the finite $N \geq M$ relay case we found the large K capacity to be $C = \frac{M}{2}\log(K) + O(1)$ and correspondingly in the adhoc case for finite $N \geq LM$, $C = \frac{LM}{2}\log(K) + O(1)$. We then proposed a simple amplify-and-forward based architecture which does not require channel state information at the relays and showed that for $N \geq 1$, the large K capacity limit is given by $C = \frac{M}{2}\log(\text{SNR}) + O(1)$.

References

[1] A. J. Paulraj and T. Kailath, "Increasing capacity in wireless broadcast systems using distributed transmission/directional reception," *U. S. Patent,* no. *5,345,599*, 1994.

[2] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecomm.*, vol. 10, no. 6, pp. 585–595, Nov.-Dec. 1999.

[3] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, March 1998.

[4] H. Bölcskei, D. Gesbert, and A. J. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Trans. Comm.*, vol. 50, no. 2, pp. 225–234, Feb. 2002.

[5] P. Viswanath, D. N. C. Tse, and V. Anantharam, "Asymptotically optimal water-filling in vector multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 47, no. 1, pp. 241–267, Jan 2001.

[6] W. Yu and J. M. Cioffi, "Sum capacity of a Gaussian vector broadcast channel," *IEEE Trans. Inf. Theory*, Nov. 2001, submitted.

[7] S. Vishwanath, N. Jindal, and A. Goldsmith, "On the capacity of multiple-input multiple-output broadcast channels," in *Proc. IEEE ICC.*, New York, NY, Apr./May 2002, vol. 3, pp. 1444–1450.

[8] H. Boche and E. A. Jorswieck, "Sum capacity optimization of the MIMO Gaussian MAC," in *Proc. IEEE WPMC*, Honolulu, HI, Oct. 2002, vol. 1, pp. 130–134.

[9] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, July 2003.

[10] B. K. Ng and E. S. Sousa, "On bandwidth-efficient multiuser-space-time signal design and detection," *IEEE J. Sel. Areas Comm.*, vol. 20, no. 2, pp. 320–329, Feb. 2002.

[11] S. Visuri and H. Bölcskei, "MIMO-OFDM multiple access with variable amount of collision," in *IEEE ICC*, Paris, France, June 2004, submitted.

[12] W. C. Jakes, *Microwave Mobile Communications*, Wiley, New York, NY, 1974.

[13] J. Guey, M. Fitz, M. Bell, and W. Kuo, "Signal design for transmitter diversity wireless communication systems over Rayleigh fading channels," in *Proc. IEEE VTC*, Atlanta, GA, 1996, vol. 1, pp. 136–140.

[14] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Comm.*, vol. 16, pp. 1451–1468, Oct. 1998.

[15] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, March 1998.

[16] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, March 2000.

[17] M. Gastpar and M. Vetterli, "On the capacity of wireless networks: The relay case," in *Proc. IEEE INFOCOM*, New York, NY, June 2002, vol. 3, pp. 1577–1586.

[18] P. Gupta and P. R. Kumar, "Towards an information theory of large networks: An achievable rate region," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1877–1894, Aug. 2003.

[19] A. F. Dana and B. Hassibi, "On the power efficiency of sensory and ad-hoc wireless networks," *IEEE Trans. Inf. Theory*, 2003, submitted.

[20] M. Grossglauser and D. N. C. Tse, "Mobility increases the capacity of ad hoc wireless networks," *IEEE/ACM Trans. Networking*, vol. 10, no. 4, pp. 477–486, Aug. 2002.

[21] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, New York, NY, 1991.

[22] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[23] Ö. Oyman, R. U. Nabar, H. Bölcskei, and A. J. Paulraj, "Capacity scaling laws in MIMO relay networks," *IEEE Trans. Wireless Comm.*, to be submitted.

[24] W. Feller, *An introduction to probability theory and its applications*, vol. 2, Wiley, New York, NY, 2nd edition, 1971.

[25] R. J. Serfling, *Approximation theorems of mathematical statistics*, Wiley, New York, NY, 1980.

[26] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge, New York, NY, 1985.

[27] J. H. Winters, J. Salz, and R. D. Gitlin, "The impact of antenna diversity on the capacity of wireless communications systems," *IEEE Trans. Comm.*, vol. 42, no. 2, pp. 1740–1751, Feb. 1994.

[28] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.

[29] D. Gerlach and A. J. Paulraj, "Base station transmitter antenna arrays with mobile to base feedback," in *Proc. Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, Nov. 1993, vol. 2, pp. 1432–1436.

[30] Ö. Oyman, R. U. Nabar, H. Bölcskei, and A. J. Paulraj, "Characterizing the statistical properties of mutual information in MIMO channels," *IEEE Trans. Sig. Proc.*, Nov. 2003, to appear.

[31] A. Wittneben and B. Rankov, "Impact of cooperative relays on the capacity of rank-deficient MIMO channels," in *Proc. 12th IST Summit on Mobile Wireless Communications*, Aveiro, Portugal, June 2003, pp. 421–425.