Power-Bandwidth Tradeoff in Linear MIMO Interference Relay Networks

Özgür Oyman, *Student Member, IEEE,* and
Arogyaswami J. Paulraj, *Fellow, IEEE*

Abstract

We consider a dense fading interference network with multiple active multi-antenna source-destination pair terminals communicating simultaneously through a large common set of K multi-antenna relay terminals. We use Shannon-theoretic tools to analyze the tradeoff between energy efficiency and spectral efficiency (known as the power-bandwidth tradeoff) in meaningful asymptotic regimes of signal-to-noise ratio (SNR) and network size. We propose low-complexity distributed coordination schemes based on linear relay forwarding and characterize their power-bandwidth tradeoff under a system-wide power constraint on source and relay transmissions. The impact of multiple users, multiple relays and multiple antennas on the key performance measures of the high and low SNR regimes is investigated in order to shed new light on the possible reduction in power and bandwidth requirements through the usage of distributed relaying techniques. Our results indicate that point-to-point coded interference networks supported by linear-beamformer relays yield enhanced energy and spectral efficiency, and achieve asymptotically optimal power-bandwidth tradeoff with energy scaling of K^{-1} at any SNR for large number of relay terminals.

Index Terms

Shannon theory, relay networks, MIMO, distributed beamforming, power-bandwidth tradeoff, energy efficiency, spectral efficiency, dense networks, scaling laws, fading channels

This paper was presented in part at the $39^{\rm th}$ Conference on Information Sciences and Systems, Baltimore, MD, March 2005, the $43^{\rm rd}$ Allerton Conference on Communications, Control, and Computing, Monticello, IL, Sep. 2005, and the $1^{\rm st}$ IEEE Workshop on Wireless Mesh Networks, Santa Clara, CA, Sep. 2005.

This work was supported by the NSF grant DMS-0354674-001 and ONR grant N00014-02-0088.

The authors are with the Information Systems Laboratory, Stanford University, Stanford, CA 94305, USA.

I. INTRODUCTION

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]. In point-to-point wireless links, MIMO systems improve spectral efficiency, link reliability and power efficiency through *spatial multiplexing gain*, *diversity gain* and *array 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. In rich scattering environments, the receiver exploits differences in the spatial signatures of the multiplexed streams to separate the different signals, thereby realizing a capacity gain.
- *Diversity* [5] 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 [6], [7], [8].
- *Array gain* [9] is achieved in MIMO systems through the enhancement of average signal-to-noise ratio (SNR) due to the transmission and reception by multiple antennas. Availability of channel state information (CSI) at the transmitter/receiver is necessary to realize transmit/receive array gains.

The very high data rates envisioned for fourth-generation (4G) wireless systems in reasonably large or densely populated areas do not appear to be feasible with the conventional cellular architectures. Multiple antenna techniques [1]-[4] and advanced signal processing techniques (such as interference cancelation algorithms) [9]-[11] by themselves cannot provide enough leverage to achieve the desired level of performance. In this context, distributed communication techniques, *relaying* in particular [12]-[14], could provide an additional leverage without requiring significant infrastructure deployment costs. With this motivation, there has been recently a growing interest both in academia and industry in the concept of relaying in infrastructure-based wireless networks such as next generation cellular (B3G, 4G), wireless local area networks (WLANs) (802.11, WiFi, HyperLAN) and broadband fixed wireless (802.16, WiMax, HyperMAN) networks [15].

The design of large scale distributed (adhoc) networks poses a set of new challenges to information theory, communication theory and network theory. Such networks are characterized by the large size of the network both in terms of the number of nodes and in terms of the geographical area the network covers. Furthermore, each terminal could be severely constrained by its computational and transmission/receiving power and/or scarcity of bandwidth resources. These constraints require an understanding of the performance limits of such networks. In this paper, we shall take an information-theoretic approach to quantify the potential *power and bandwidth efficiency gains of distributed MIMO relaying* in wireless networks. We first recall Shannon's famous capacity theorem which serves as a key starting point in our analysis. This theorem suggests that there exists a tradeoff between power, bandwidth and coding complexity in achieving a certain target data rate R. To illustrate this tradeoff, let us consider a simple example; the additive white Gaussian noise (AWGN) channel. Any achievable data rate R over the AWGN channel is upper bounded by

$$R < B \log_2 \left(1 + \frac{P}{N_0 B}\right), \tag{1}$$

as a function of the signal power P, channel bandwidth B and single-sided noise power spectral density N_0. Two measures determine the level of how efficiently the power and bandwidth resources of the system are utilized: i) *Energy efficiency*, quantified by the energy per information bit $E_b = \frac{P}{R}$ (in Joules), and ii) *Spectral efficiency*, quantified by $\mathsf{C} = \frac{R}{B}$ (in bits per second per Hertz (b/s/Hz)). Re-expressing the terms in (1) based on these definitions, we find that the achievable set of energy and spectral efficiencies need to satisfy the condition

$$\frac{E_b}{N_0} > \frac{2^{\mathsf{C}} - 1}{\mathsf{C}}.$$

We plot this relationship in Fig. 1. First, we note that there exists a tradeoff between the efficiency measures $\frac{E_b}{N_0}$ and C (known as the power-bandwidth tradeoff) in achieving a given target data rate. All points above the power-bandwidth tradeoff curve are feasible with a certain amount of coding complexity. On the other hand, in the region below the curve, reliable communication is not possible. We observe that $\frac{E_b}{N_0} = \ln 2 \approx -1.6\,\text{dB}$ is the minimum required level of energy efficiency for reliable communication. When $\mathsf{C} \ll 1$, the system operates in the *power-limited* *regime*; i.e., the bandwidth is large and the main concern is the limitation on power. Similarly, the case of $\mathsf{C} \gg 1$ corresponds to the *bandwidth-limited regime*. The interplay of power, bandwidth and rate in the power-limited regime has been previously analyzed in the context of single-user [16]-[17], multi-user [18]-[21], single-relay [22]-[23] and relay network [24] scenarios. In the bandwidth-limited regime, the necessary tools to perform the power-bandwidth tradeoff analysis were developed by [18] in the context of code-division multiple access (CDMA) systems and were later used by [25] to characterize fundamental limits over correlated multi-antenna channels. We emphasize that the power-bandwidth tradeoff formulation is a beneficial analytical tool for evaluating the performance of relay networks considering *jointly gains in terms of energy and spectral efficiencies*, while previous works have mostly focused either only on energy efficiency [26] or only on spectral efficiency [27]-[29]. Specifically, we shall focus on the power-bandwidth tradeoff in the low and high SNR regimes, as depicted in Fig. 2. Finally, the ergodicity assumption on the channel statistics allows us to ignore the temporal dimension; which enters the three-fold tradeoff among power, bandwidth and delay for slow fading channels [30].

We consider a fading MIMO interference relay network (MIRN), where multiple active multi-antenna source-destination pair terminals communicate with each other by routing their data through other (relay) terminals. Every terminal can act both as a sender/receiver of data and as a relay for other transmissions. No cooperation among the terminals is allowed. All users transmit and receive in a point-to-point fashion; in this sense, we emphasize that the assistance of relay terminals does not significantly alter the transceiver design at the source-destination pairs. In order to reduce the relay transceiver complexity, we constrain the processing at the relay terminals to be linear; in other words, no encoding or decoding operations are performed at the relay terminals. We design low-complexity linear distributed multi-antenna relaying schemes that take advantage of local channel state information (CSI) to beamform simultaneously multiple users' signals to their intended destinations. We analyze $\frac{E_b}{N_0}$ as a function of C (from now on, we are only interested in the optimal $(\mathsf{C}, \frac{E_b}{N_0})$ points on the power-bandwidth tradeoff curve, while it is clear that all points above this curve are achievable by suboptimal coding techniques) for the linear MIRN (L-MIRN) and investigate the impact of multiple users, multiple relays and multiple antennas on the key performance measures of the low and high SNR regimes in the limit of large number of relay terminals. We prove the *asymptotic optimality* of linear relaying for any SNR. In particular, we show that with bursty signaling, much better energy scaling (K^{-1} rather than $K^{-1/2}$) is achievable with linear relaying compared to previous work in [26]. Finally, our work demonstrates that the region of achievable $(\mathsf{C}, \frac{E_b}{N_0})$ pair points can be enlarged significantly by low-complexity distributed MIMO relaying techniques and provides insights on power and bandwidth efficient design of adhoc wireless networks.

The rest of this paper is organized as follows. Section II describes the MIRN model and the power-bandwidth tradeoff problem formulation. In Section III, we derive an upper-limit on the achievable power-bandwidth tradeoff based on the cut-set theorem [31]. We analyze the performance of L-MIRN for various low-complexity distributed MIMO relaying schemes in Section IV. Finally, we conclude in Section V.

II. Network Model and Definitions

General Assumptions. We assume that the MIRN network consists of $K+2L$ terminals, with L active source-destination pairs and K relay terminals located randomly and independently in a domain of fixed area. We denote the l-th source terminal by $\mathcal{S}_l$, the l-th destination terminal by $\mathcal{D}_l$, where $l = 1, ..., L$, and the k-th relay terminal by $\mathcal{R}_k$, $k = 1, 2, ..., K$. The source and destination terminals $\{\mathcal{S}_l\}$ and $\{\mathcal{D}_l\}$ are equipped with M antennas each, while each of the relay terminals $\mathcal{R}_k$ employs N transmit/receive antennas. We assume that there is a "dead zone" of non-zero radius around $\{\mathcal{S}_l\}$ and $\{\mathcal{D}_l\}$ [28], which is free of relay terminals and that no direct link exists between the source-destination pairs. The source terminal $\mathcal{S}_l$ is only interested in sending data to the destination terminal $\mathcal{D}_l$ and the communication of all L source-destination pairs is supported through the same set of K relay terminals. As terminals can often not transmit and receive at the same time, we consider time-division based (half duplex) relaying schemes for which transmissions take place in two hops over two separate time slots. In the first time slot, the relay terminals receive the signals transmitted from the source terminals. After processing the received signals, the relay terminals simultaneously transmit their data to the destination terminals during the second time slot.

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. In frequency-selective environments, the channel can be decomposed into parallel non-interacting subchannels each experiencing frequency-flat fading and having the same capacity as the overall channel. The channel model is depicted in Fig. 3. The discrete-time complex baseband input-output relation for the $\mathcal{S}_l \to \mathcal{R}_k$ link[1] is given by

$$\mathbf{r}_k = \sum_{l=1}^{L} \sqrt{E_{k,l}}\, \mathbf{H}_{k,l}\mathbf{s}_l + \mathbf{n}_k, \quad k = 1, 2, ..., K,$$

where $\mathbf{r}_k \in \mathbb{C}^N$ is the received vector signal at $\mathcal{R}_k$, $E_{k,l} \in \mathbb{R}$ is an energy normalization factor to account for path loss and shadowing in the $\mathcal{S}_l \to \mathcal{R}_k$ link, $\mathbf{H}_{k,l} \in \mathbb{C}^{N\times M}$ is the corresponding channel matrix independent across source and relay terminals (i.e., independent across k and l) and consisting of i.i.d. $\mathcal{CN}(0,1)$ entries, $\mathbf{s}_l \in \mathbb{C}^M$ is the temporally i.i.d. zero-mean circularly symmetric complex Gaussian transmit signal vector for $\mathcal{S}_l$ satisfying $\mathbb{E}\left[\mathbf{s}_l\mathbf{s}_l^H\right] = \frac{P_{\mathcal{S}_l}}{M}\mathbf{I}_M$ (i.e. $P_{\mathcal{S}_l} = \mathbb{E}\left[\|\mathbf{s}_l\|^2\right]$ is the average transmit power for source $\mathcal{S}_l$), and $\mathbf{n}_k \in \mathbb{C}^N$ is the spatio-temporally white zero-mean circularly symmetric complex Gaussian noise vector at $\mathcal{R}_k$, independent across k, with single-sided noise spectral density N_0.

Each relay terminal $\mathcal{R}_k$ *linearly* processes its received vector signal $\mathbf{r}_k$ to produce the vector signal $\mathbf{t}_k \in \mathbb{C}^N$ (i.e., $\mathbf{t}_k = \mathbf{A}_k\mathbf{r}_k$ with linear-beamformer matrices $\mathbf{A}_k \in \mathbb{C}^{N\times N}$), which is then transmitted to the destination terminals over the second time slot. The destination terminal $\mathcal{D}_l$ receives the signal vector $\mathbf{y}_l \in \mathbb{C}^M$ expressed as

$$\mathbf{y}_l = \sum_{k=1}^{K} \sqrt{F_{k,l}}\, \mathbf{G}_{k,l}\mathbf{t}_k + \mathbf{z}_l, \quad l = 1, ..., L,$$

where $F_{k,l} \in \mathbb{R}$ is an energy normalization factor to account for path loss and shadowing in the $\mathcal{R}_k \to \mathcal{D}_l$ link, $\mathbf{G}_{k,l} \in \mathbb{C}^{M\times N}$ is the corresponding i.i.d. $\mathcal{CN}(0,1)$ channel matrix, independent across k and l, and $\mathbf{z}_l \in \mathbb{C}^M$ is the spatio-temporally white circularly symmetric complex Gaussian noise vector at $\mathcal{D}_l$ with single-sided noise spectral density N_0. The transmit signal vectors $\mathbf{t}_k$ satisfy the average power constraint $\mathbb{E}\left[\|\mathbf{t}_k\|^2\right] \leq P_{\mathcal{R}_k}$ ($P_{\mathcal{R}_k}$ is the average transmit power for relay $\mathcal{R}_k$).

For any block length Q, a $((2^{QR_1}, ..., 2^{QR_L}), Q)$ code $\mathcal{C}_Q$ is defined by a codebook of $\sum_{l=1}^{L} 2^{QR_l}$ codewords such that R_l is the rate of communication for the l-th source-destination pair. The source codebook is determined by the encoding functions ϕ_l , $l = 1, ..., L$, that map each message $w_l \in \mathcal{W}_l = \{1, ..., 2^{QR_l}\}$ of user l to a transmit codeword $\mathbf{S}_l = \left[\mathbf{s}_{l,1}, \ldots, \mathbf{s}_{l,Q}\right] \in \mathbb{C}^{M\times Q}$, where $\mathbf{s}_{l,q}$ is the transmit codeword of user l at time q. Furthermore, each destination terminal employs a decoding function ψ_l , $l = 1, ..., L$ to perform the mapping $\mathbb{C}^{M\times Q} \to \hat{w}_l \in \mathcal{W}_l$ based

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

on its received signal $\mathbf{Y}_l = [\mathbf{y}_{l,1}, \ldots, \mathbf{y}_{l,Q}]$. The error probability for the l-th user given by $\epsilon_l = \mathbb{P}(\psi_l(\mathbf{Y}_l) \neq w_l)$. An L-tuple of rates $(R_1, \ldots, R_L)$ is achievable if there exists a sequence of $((2^{QR_1}, \ldots, 2^{QR_L}), Q)$ codes $\{\mathcal{C}_Q : Q = 1, 2, \ldots\}$ with vanishing ϵ_l, $\forall l$. Based on independent encoding/decoding for each user and linear-beamforming at the relay terminals, the $\mathcal{S}_l \rightarrow \mathcal{D}_l$, $l = 1, \ldots, L$ source-destination links can be viewed as a composite *interference channel* where the properties of the resulting channel distribution $p(\mathbf{y}_1, \ldots, \mathbf{y}_L \,|\, \mathbf{s}_1, \ldots, \mathbf{s}_L)$ heavily rely upon the choice of the linear-beamformer matrices $\{\mathbf{A}_k\}_{k=1}^K$.

As already mentioned above, throughout the paper, the path-loss and shadowing statistics are captured by $\{E_{k,l}\}$ (for the first hop) and $\{F_{k,l}\}$ (for the second hop). We assume that these parameters are random, i.i.d., strictly positive (due to the fact that the domain of interest has a fixed area, i.e. dense network), bounded above (due to the dead zone requirement), and remain constant over the entire time period of interest. Additionally, we assume an ergodic block fading channel model such that the channel matrices $\{\mathbf{H}_{k,l}\}$ and $\{\mathbf{G}_{k,l}\}$ remain constant over the entire duration of a time slot and change in an independent fashion across time slots. Finally, we assume that there is no CSI at the source terminals $\{\mathcal{S}_l\}$, each relay terminal $\mathcal{R}_k$ has perfect knowledge of its local forward and backward channels, $\{F_{k,l}, \mathbf{G}_{k,l}\}_{l=1}^L$ and $\{E_{k,l}, \mathbf{H}_{k,l}\}_{l=1}^L$, respectively, and the destination terminals $\{\mathcal{D}_l\}$ have perfect knowledge of all channel variables.[2]

Spectral Efficiency vs. $\frac{\mathbf{E_b}}{\mathbf{N_0}}$. We assume that the network is supplied with fixed finite total power P (W) over unconstrained bandwidth B (Hz). We define the network signal-to-noise ratio (SNR) for the $\mathcal{S}_l \rightarrow \mathcal{D}_l$, $l = 1, \ldots, L$ links as

$$\mathsf{SNR}_{\,\text{network}} \doteq \frac{P}{N_0 B} = \frac{\sum_{l=1}^L P_{\mathcal{S}_l} + \sum_{k=1}^K P_{\mathcal{R}_k}}{2 N_0 B},$$

where the factor of $1/2$ comes from the half duplex nature of source and relay transmissions. Note that our definition of network SNR captures power consumption at the relay as well as source terminals ensuring a fair performance comparison between distributed relaying and direct transmissions. To simplify notation, from now on we refer to $\mathsf{SNR}_{\text{network}}$ as SNR. Due to the statistical symmetry of their channel distributions, we allow for equal power allocation among the source and relay terminals and set $P_{\mathcal{S}_l} = P_{\mathcal{S}}$, $\forall l$ and $P_{\mathcal{R}_k} = P_{\mathcal{R}}$, $\forall k$. The network with desired sum rate $R = \sum_{l=1}^L R_l$ (the union of the set of achievable rate tuples $(R_1, R_2, \ldots, R_L)$

[2]As we shall show in Section IV, the CSI knowledge at the destination terminals is not required for our results to hold in the asymptotic regime where the number of relays tends to infinity.

defines the capacity region of the interference relay network) and available bandwidth B must respect the fundamental limit

$$\frac{R}{B} \leq \mathsf{C}\left(\frac{E_b}{N_0}\right).$$

Tightly framing achievable performance, particular emphasis throughout our power-bandwidth tradeoff analysis is placed on the regions of low and high $\frac{E_b}{N_0}$.

Low $\frac{E_b}{N_0}$ regime. Defining $\frac{E_b}{N_0}_{\min}$ as the minimum system-wide $\frac{E_b}{N_0}$ required to convey any positive rate reliably, we have [3]

$$\frac{E_b}{N_0}_{\min} = \min_{\mathsf{SNR}} \frac{\mathsf{SNR}}{C(\mathsf{SNR})}.$$

In most of the scenarios we will consider, $\frac{E_b}{N_0}$ is minimized when SNR is low. This regime of operation is referred as the *wideband regime* in which the spectral efficiency C is near zero. We consider the first-order behavior of C as a function of $\frac{E_b}{N_0}$ in the wideband regime (i.e., $\mathsf{C} \rightarrow 0$) by analyzing the affine function (in decibels) [4]

$$10 \log_{10} \frac{E_b}{N_0}(\mathsf{C}) = 10 \log_{10} \frac{E_b}{N_0}_{\min} + \frac{\mathsf{C}}{S_0} 10 \log_{10} 2 + o(\mathsf{C}),$$

where S_0 denotes the "wideband" slope of spectral efficiency in b/s/Hz/(3 dB) at the point $\frac{E_b}{N_0}_{\min}$,

$$S_0 = \lim_{\frac{E_b}{N_0} \downarrow \frac{E_b}{N_0}_{\min}} \frac{\mathsf{C}(\frac{E_b}{N_0})}{10 \log_{10} \frac{E_b}{N_0} - 10 \log_{10} \frac{E_b}{N_0}_{\min}} 10 \log_{10} 2.$$

It can be shown that [16]

$$\frac{E_b}{N_0}_{\min} = \frac{\ln 2}{\dot{C}(0)}, \quad \text{and} \quad S_0 = \frac{2\left[\dot{C}(0)\right]^2}{-\ddot{C}(0)}, \tag{2}$$

where $\dot{C}$ and $\ddot{C}$ denote the first and second order derivatives of $C(\mathsf{SNR})$ (evaluated in nats/s/Hz) at $\mathsf{SNR} = 0$.

High $\frac{E_b}{N_0}$ regime. In the high spectral efficiency region (i.e., $\mathsf{C} \rightarrow \infty$), the dependence between $\frac{E_b}{N_0}$ and C can be characterized as [18]

$$10 \log_{10} \frac{E_b}{N_0}(\mathsf{C}) \;=\; \frac{\mathsf{C}}{S_\infty} 10 \log_{10} 2 \;-\; 10 \log_{10}(\mathsf{C}) \;+\; 10 \log_{10} \frac{E_b}{N_0}_{\text{imp}} + o(1),$$

[3] The use of C and C avoids assigning the same symbol to capacity functions of SNR and $\frac{E_b}{N_0}$.

[4] $u(x) = o(v(x)), x \rightarrow L$ stands for $\lim_{x \rightarrow L} \frac{u(x)}{v(x)} = 0$.

where S_∞ denotes the "high SNR" slope of the spectral efficiency in b/s/Hz/(3 dB)

$$\begin{aligned} S_\infty &= \lim_{\frac{E_b}{N_0}\to\infty} \frac{\mathsf{C}(\frac{E_b}{N_0})}{10\log_{10}\frac{E_b}{N_0}} 10\log_{10} 2 \\ &= \lim_{\mathsf{SNR}\to\infty} \mathsf{SNR}\,\dot{C}(\mathsf{SNR}) \end{aligned} \tag{3}$$

and $\frac{E_b}{N_0}_{\mathrm{imp}}$ is the $\frac{E_b}{N_0}$ improvement factor with respect to a single-user single-antenna unfaded AWGN reference channel[5] and it is expressed as

$$\frac{E_b}{N_0}_{\mathrm{imp}} = \lim_{\mathsf{SNR}\to\infty} \left[\mathsf{SNR} \exp\left(-\frac{C(\mathsf{SNR})}{S_\infty} \right) \right]. \tag{4}$$

III. Upper-Limit on Power-Bandwidth Tradeoff

Theorem 1. *In the limit of large K, $\frac{E_b}{N_0}$ can almost surely be lower bounded by*

$$\frac{E_b}{N_0}(\mathsf{C}) \geq \frac{2^{2\mathsf{C}/(LM)}-1}{2\mathsf{C}/(LM)} \frac{1}{KN\,\mathbb{E}\left[E_{k,l}\right]} + o(\frac{1}{K}),\ K\to\infty. \tag{5}$$

1- Best-case power-bandwidth tradeoff at low $\frac{E_b}{N_0}$:

$$\frac{E_b}{N_0}_{\min}^{\mathrm{best}} = \frac{\ln 2}{K\,N\,\mathbb{E}\left[E_{k,l}\right]} + o(\frac{1}{K}) \qquad \textit{and} \qquad S_0^{\mathrm{best}} = LM$$

2- Best-case power-bandwidth tradeoff at high $\frac{E_b}{N_0}$:

$$\frac{E_b}{N_0}_{\mathrm{imp}}^{\mathrm{best}} = \frac{L\,M}{2\,K\,N\,\mathbb{E}\left[E_{k,l}\right]} + o(\frac{1}{K}) \quad \textit{and} \quad S_\infty^{\mathrm{best}} = \frac{LM}{2}.$$

Proof: Separating the source terminals $\{\mathcal{S}_l\}$ from the rest of the network using a broadcast cut (see Fig. 4), and applying the cut-set theorem [31, Th. 14.10.1] it follows that the sum capacity of the MIMO relay network is upper bounded as

$$\mathsf{C} \leq \mathbb{E}_{\{\mathbf{H}_{k,l},\mathbf{G}_{k,l}\}} \left[\frac{1}{2} I(\{\mathbf{s}_l\}_{l=1}^L; \{\mathbf{r}_k\}_{k=1}^K, \{\mathbf{y}_l\}_{l=1}^L | \{\mathbf{t}_k\}_{k=1}^K) \right],$$

where the factor $1/2$ results from the fact that data is transmitted over two time slots. Observing that in our network model $\{\mathbf{s}_l\} \to \{\mathbf{r}_k\} \to \{\mathbf{t}_k\} \to \{\mathbf{y}_l\}$ forms a Markov chain, applying the chain rule of mutual information [31] and using the fact that conditioning reduces entropy, we extend the upper bound to

$$\mathsf{C} \leq \mathbb{E}_{\{\mathbf{H}_{k,l}\}} \left[\frac{1}{2} I(\mathbf{s}_1,...,\mathbf{s}_L; \mathbf{r}_1,...,\mathbf{r}_K) \right].$$

[5]For the AWGN channel; $C(\mathsf{SNR}) = \ln(1+\mathsf{SNR})$ resulting in $S_0 = 2$, $\frac{E_b}{N_0}_{\min} = \ln 2$, $S_\infty = 1$ and $\frac{E_b}{N_0}_{\mathrm{imp}} = 1$.

Recalling that $\{\mathbf{s}_l\}$ are circularly symmetric complex Gaussian with $\mathbb{E}\left[\mathbf{s}_l\mathbf{s}_l^H\right] = \frac{P_{\mathcal{S}}}{M}\mathbf{I}_M$, we have

$$\mathsf{C} \leq \mathbb{E}_{\{\mathbf{H}_{k,l}\}}\left[\frac{1}{2}\log_2\left(\left|\mathbf{I}_{LM} + \frac{P_{\mathcal{S}}}{M\,N_0\,B}\mathbf{V}\right|\right)\right], \tag{6}$$

where $\mathbf{V}$ is an $LM \times LM$ matrix of the form

$$\mathbf{V} = \begin{bmatrix} \mathbf{Q}_{1,1} & \cdots & \mathbf{Q}_{1,L} \\ \vdots & & \vdots \\ \mathbf{Q}_{L,1} & \cdots & \mathbf{Q}_{L,L} \end{bmatrix},$$

with $M \times M$ matrices $\mathbf{Q}_{i,j}$ given by

$$\mathbf{Q}_{i,j} = \sum_{k=1}^{K}\sqrt{E_{k,i}E_{k,j}}\mathbf{H}_{k,i}^H\mathbf{H}_{k,j}, \;\; i = 1,...,L, \; j = 1,...,L$$

Now, applying Jensen's inequality to (6) it follows that

$$\begin{aligned} \mathsf{C} &\leq \frac{1}{2}\log_2\left(\left|\mathbf{I}_{LM} + \frac{P_{\mathcal{S}}}{M\,N_0\,B}\mathbb{E}_{\{\mathbf{H}_{k,l}\}}\left[\mathbf{V}\right]\right|\right) \\ &= \frac{M}{2}\sum_{l=1}^{L}\log_2\left(1 + \frac{P_{\mathcal{S}}\,N}{M\,N_0\,B}\sum_{k=1}^{K}E_{k,l}\right). \end{aligned}$$

By our assumption that $\{E_{k,l}\}$ are bounded, it follows that $\mathrm{var}(E_{k,l})$ is also bounded $\forall k,l$ and hence the Kolmogorov condition [32]

$$\sum_{k=1}^{\infty}\frac{\mathrm{var}(E_{k,l})}{k^2} < \infty$$

is satisfied. We can therefore use [32, Th. 1.8.D] to obtain [6]

$$\sum_{k=1}^{K}\frac{E_{k,l}}{K} - \sum_{k=1}^{K}\frac{\mathbb{E}\left[E_{k,l}\right]}{K} \xrightarrow{\text{w.p.1}} 0$$

resulting in [32, Th. 1.7]

$$\mathsf{C} \leq \frac{L\,M}{2}\log_2\left(1 + \frac{P_{\mathcal{S}}KN\mathbb{E}\left[E_{k,l}\right]}{M\,N_0\,B}\right) \tag{7}$$

as $K \to \infty$. Since our application of the cut-set theorem through the broadcast cut leads to perfect relay-destination (i.e. $\mathcal{R}_k \to \mathcal{D}_l$) links, relays do not consume any transmit power and hence we set $P_{\mathcal{R}} = 0$ yielding

$$\mathsf{SNR} = \mathsf{C}\,\frac{E_b}{N_0} = \frac{L\,P_{\mathcal{S}}}{2N_0B}. \tag{8}$$

[6] $\xrightarrow{\text{w.p.1}}$ denotes convergence with probability 1.

We combine (7) and (8) to show (5). Expressing the upper bound on C given in (7) in terms of SNR and applying (2)-(4), we complete the proof. □

Discussion. Since the cut-set bound is a strict upper bound on the capacity of MIRN, the result of Theorem 1 gives a "best-case" picture in terms of how the MIRN parameters, namely the number of users L, number of relays K and number of antennas at the source-destination pair and relay terminals, M and N, respectively, impact energy and spectral efficiency. We observe that the minimum required transmit power to support reliable communication for fixed rate and bandwidth reduces (at best) by a factor of KN through distributed multi-antenna relaying. Furthermore, $\frac{E_b}{N_0}{}^{\text{best}}_{\text{imp}}$ also improves by a factor of KN, indicating significant power savings with respect to direct transmission in the high $\frac{E_b}{N_0}$ regime. In addition, larger L and larger M reduce the required bandwidth for fixed power and rate by a factor of $\frac{LM}{2}$ as seen through the improvements in the wideband slope S_0^{best} and high-SNR slope S_∞^{best} justifying the value of *superposition and antenna cooperation* techniques in terms of spectral efficiency over relay networks despite the factor of $1/2$ loss due to half duplex transmission.

IV. L-MIRN Power-Bandwidth Tradeoff

In this section, we present low-complexity linear (but suboptimal) relay forwarding schemes such that each relay transmit vector $\mathbf{t}_k \in \mathbb{C}^N$ is a linear transformation of the corresponding received vector $\mathbf{r}_k \in \mathbb{C}^N$. The beamforming schemes we consider at the relay terminals differ in the way they fight interference (arising due to simultaneous transmission of multiple users and multiple streams per user) and background Gaussian noise: *i) The matched filter (MF)* mitigates noise but ignores interference. *ii) The zero-forcing (ZF) filter* cancels interference completely (requiring $N \geq LM$), but amplifies noise. *iii) The linear minimum mean-square error (L-MMSE) filter* is the best tradeoff for interference and noise mitigation [9].

Relaying Scheme. Each relay terminal exploits its knowledge of the local backward CSI $\{E_{k,l}, \mathbf{H}_{k,l}\}_{l=1}^{L}$ to perform input linear-beamforming operations on its received signal vector to obtain estimates for each of the transmitted user signals. Accordingly, terminal $\mathcal{R}_k$ correlates its received signal vector $\mathbf{r}_k$ with each of the beamforming (row) vectors $\mathbf{u}_{k,l,i} \in \mathbb{C}^N$ to yield

$$\begin{aligned}\hat{s}_{k,l,i} &= \mathbf{u}_{k,l,i}, \mathbf{r}_k \\ &= \sqrt{E_{k,l}}\,\mathbf{u}_{k,l,i}\mathbf{h}_{k,l,i}s_{l,i} + \sum_{(p,q)\neq(l,i)} \sqrt{E_{k,p}}\,\mathbf{u}_{k,l,i}\mathbf{h}_{k,p,q}\,s_{p,q} + \mathbf{u}_{k,l,i}\,\mathbf{n}_k,\end{aligned}$$

as the estimate for $s_{l,i}$, where $s_{p,q}$ denotes the transmitted signal from the q-th antenna of source $\mathcal{S}_p$, $p = 1, 2, ..., L$, $q = 1, 2, ..., M$, and $\mathbf{h}_{k,p,q}$ is the q-th column of $\mathbf{H}_{k,p}$. Next, $\mathcal{R}_k$ sets the average energy (conditional on the channel realizations $\{E_{k,l}, \mathbf{H}_{k,l}\}_{l=1}^{L}$) of each estimate to unity and obtains the normalized estimates $\hat{s}^{\mathrm{U}}_{k,l,i}$. Finally, $\mathcal{R}_k$ passes the normalized estimates through output linear-beamformer (column) vectors $\mathbf{v}_{k,l,i} \in \mathbb{C}^N$ (which are designed to exploit the knowledge of the forward CSI $\{F_{k,l}, \mathbf{G}_{k,l}\}_{l=1}^{L}$) to produce its transmit signal vector

$$\mathbf{t}_k = \frac{\sqrt{P_{\mathcal{R}}}}{LM} \sum_{p=1}^{L} \sum_{q=1}^{M} \frac{\mathbf{v}_{k,p,q}}{\|\mathbf{v}_{k,p,q}\|}\, \hat{s}^{\mathrm{U}}_{k,p,q},$$

concurrently ensuring that the transmit power constraint is satisfied. Under linear relay forwarding, it follows that the i-th element of the signal $\mathbf{y}_l$ received at $\mathcal{D}_l$ is given by

$$y_{l,i} = \sum_{k=1}^{K} \frac{\sqrt{F_{k,l} P_{\mathcal{R}}}}{LM} \sum_{p=1}^{L} \sum_{q=1}^{M} \frac{\mathbf{g}_{k,l,i}\, \mathbf{v}_{k,p,q}}{\|\mathbf{v}_{k,p,q}\|}\, \hat{s}^{\mathrm{U}}_{k,p,q} + z_{l,i},$$

where $\mathbf{g}_{k,p,q}$ is the q-th row of $\mathbf{G}_{k,p}$. We list the input and output linear-beamformer matrices $\{\mathbf{U}_k\}_{k=1}^{K}$ and $\{\mathbf{V}_k\}_{k=1}^{K}$ based on MF, ZF and L-MMSE filters in Table 1. [7]

Spectral Efficiency vs. $\frac{E_b}{N_0}$. In the following, we characterize the power-bandwidth tradeoff of L-MIRN for the MF and ZF schemes as the number of relay terminals grows asymptotically large. We summarize the low $\frac{E_b}{N_0}$ and high $\frac{E_b}{N_0}$ results separately.

Theorem 2:

Low $\frac{E_b}{N_0}$ analysis. *In the limit of large K, L-MIRN power-bandwidth tradeoff for MF and ZF schemes almost surely converges to the deterministic relationship*

$$\frac{E_b}{N_0}(\mathsf{C}) = \sqrt{\frac{L^3 M^3}{\epsilon_1^2 K} \frac{2^{2\mathsf{C}/(LM)} - 1}{\mathsf{C}^2}} + o\left(\frac{1}{\sqrt{K}}\right), \quad K \to \infty, \tag{9}$$

where the constant ϵ_1 is expressed as $\epsilon_1 = \mathbb{E}\left[\sqrt{E_{k,l} F_{k,l} X_{k,l,i} Y_{k,l,i}}\right]$, and fading-dependent random variables $X_{k,l,i}$ and $Y_{k,l,i}$ (independent across k) follow the $\Gamma(N)$ probability distribution $p(\gamma) = \frac{1}{(N-1)!}\gamma^{N-1} e^{-\gamma}$ for the MF scheme and $\Gamma(N - LM + 1)$ distribution for the ZF scheme. Both linear beamforming schemes achieve the minimum energy per bit at finite spectral efficiency

[7] The row vector $\mathbf{u}_{k,l,i} \in \mathbb{C}^N$ is the $((l-1)M + i)$-th row of $\mathbf{U}_k \in \mathbb{C}^{LM \times N}$. The column vector $\mathbf{v}_{k,l,i} \in \mathbb{C}^N$ is the $((l-1)M + i)$-th column of $\mathbf{V}_k \in \mathbb{C}^{N \times LM}$.

C^* *which is the positive solution of*

$$\left(\frac{\ln 2}{LM}\,\mathsf{C}-1\right)2^{2\mathsf{C}/(LM)}+1=0$$

expressed as $\mathsf{C}^* \approx 1.15\,LM$ *and consequently*

$$\frac{E_b}{N_0}^{\text{linear}}_{\min} \approx \sqrt{\frac{2.97LM}{\epsilon_1^2 K}} + o(\frac{1}{\sqrt{K}}), \quad K \to \infty. \tag{10}$$

High $\frac{E_b}{N_0}$ analysis. *In the limit of large K, L-MIRN power-bandwidth tradeoff for the ZF scheme almost surely converges to the deterministic relationship*

$$\frac{E_b}{N_0}(\mathsf{C}) = \frac{2^{2\mathsf{C}/(LM)}}{2\mathsf{C}/(LM)}\,\frac{\left(\sqrt{\epsilon_2}+\sqrt{L\,M}\right)^2}{K\,\epsilon_3^2} + o(\frac{1}{K}), \quad K \to \infty, \tag{11}$$

where constants ϵ_2 and ϵ_3 are expressed as $\epsilon_2 = \mathbb{E}\left[\frac{F_{k,l}X_{k,l,i}}{E_{k,l}Y_{k,l,i}}\right]$, and $\epsilon_3 = \mathbb{E}\left[\sqrt{F_{k,l}X_{k,l,i}}\right]$, and fading-dependent random variables $X_{k,l,i}$ and $Y_{k,l,i}$ (independent across k) follow the $\Gamma(N - LM + 1)$ probability distribution. In the high spectral efficiency region, we obtain

$$\frac{E_b}{N_0}^{\text{ZF}}_{\text{imp}} = \frac{L\,M}{2\,K\,\epsilon_3^2}\left(\sqrt{\epsilon_2}+\sqrt{L\,M}\right)^2 + o(\frac{1}{K}) \qquad and \qquad S_\infty^{\text{ZF}} = \frac{LM}{2}. \tag{12}$$

On the other hand, under the MF scheme, MIRN is in the interference-limited regime and C^{MF} converges to a fixed constant (which scales like $\log(K)$) as $\frac{E_b}{N_0} \to \infty$; leading to $S_\infty^{\text{MF}} = 0$.

Proof: Each individual single-antenna at the destination terminals decodes its intended signal with no attempt to exploit the knowledge of the codebooks of the interfering streams; and instead the interference is treated as Gaussian noise. Furthermore, the processing at the relay terminals is considered to be linear. Based on these simplifications, the spectral efficiency of L-MIRN expressed as

$$\mathsf{C}^{\text{linear}} = \frac{1}{2}\sum_{l=1}^{L}\sum_{i=1}^{M}\mathbb{E}_{\{\mathbf{H}_{k,l},\mathbf{G}_{k,l}\}}\left[\log_2\left(1+\mathsf{SIR}_{l,i}\right)\right], \tag{13}$$

where $\mathsf{SIR}_{l,i}$ is the signal-to-interference-plus-noise ratio corresponding to stream $s_{l,i}$ at terminal $\mathcal{D}_l$. The rest of the proof involves the analysis of low and high $\frac{E_b}{N_0}$ asymptotic behavior of $\mathsf{SIR}_{l,i}$ in the limit of large K for MF and ZF schemes.

Here we present the detailed power-bandwidth tradeoff analysis for the ZF scheme in the high and low $\frac{E_b}{N_0}$ regimes. The performance of the MF scheme in the low $\frac{E_b}{N_0}$ regime was analyzed in

[33]. It is easy to show that (see [34]) in a ZF-based MIMO relay network, the signal received at the destination terminal $\mathcal{D}_l$ corresponding to the i-th multiplexed stream $s_{l,i}$ is given by

$$y_{l,i}^{\mathrm{ZF}} = \left(\sum_{k=1}^{K} d_{k,l,i} \right) s_{l.i} + \sum_{k=1}^{K} d_{k,l,i} \, \widetilde{n}_{k,l,i} + z_{l,i}, \tag{14}$$

where

$$d_{k,l,i} = \sqrt{\frac{P_{\mathcal{R}} F_{k,l} X_{k,l,i}}{L^2 M^2 \left(\frac{P_{\mathcal{S}}}{M} + \left(\frac{E_{k,l} Y_{k,l,i}}{N_0 B} \right)^{-1} \right)}}, \tag{15}$$

and $\widetilde{n}_{k,l,i}$ denotes the i-th element of the vector $\widetilde{\mathbf{n}}_{k,l} = (E_{k,l})^{-1/2} \mathbf{D}_{k,l} \mathbf{n}_k$ and fading-dependent random variables $X_{k,l,i}$ and $Y_{k,l,i}$ follow the $\Gamma(N - LM + 1)$ probability distribution. As a result ZF-based MIMO relaying scheme effectively decouples the effective channels between source-destination pairs $\{\mathcal{S}_l \to \mathcal{D}_l\}_{l=1}^{L}$ into LM parallel scalar channels. Based on this simplification, we now compute $\mathsf{SIR}_{l,i}$ easily from (14) as

$$\mathsf{SIR}_{l,i}^{\mathrm{ZF}} = \frac{P_{\mathcal{S}} \left(\sum_{k=1}^{K} d_{k,l,i} \right)^2}{M \, N_0 \, B \, \left(1 + \sum_{k=1}^{K} f_{k,l,i}^2 \right)}, \tag{16}$$

in which $d_{k,l,i}$ was defined in (15) and

$$f_{k,l,i} = \sqrt{\frac{P_{\mathcal{R}} F_{k,l} \, X_{k,l,i}}{L^2 M^2 \left(\frac{E_{k,l} P_{\mathcal{S}}}{M} Y_{k,l,i} + N_0 B \right)}}. \tag{17}$$

Substituting (15) and (17) into (16), we have

$$\mathsf{SIR}_{l,i}^{\mathrm{ZF}} = \frac{P_{\mathcal{S}} K^2 \left(\frac{1}{K} \sum_{k=1}^{K} \sqrt{\frac{P_{\mathcal{R}} F_{k,l} X_{k,l,i}}{L^2 M^2 \left(\frac{P_{\mathcal{S}}}{M} + \left(\frac{E_{k,l} Y_{k,l,i}}{N_0 B} \right)^{-1} \right)}} \right)^2}{M N_0 B \left(1 + K \, \frac{1}{K} \sum_{k=1}^{K} \frac{P_{\mathcal{R}} F_{k,l} X_{k,l,i}}{L^2 M^2 \left(\frac{E_{k,l} P_{\mathcal{S}}}{M} Y_{k,l,i} + N_0 B \right)} \right)} \tag{18}$$

If $\mathsf{SNR} \gg 1$ (high $\frac{E_b}{N_0}$ regime), then $\mathsf{SIR}_{l,i}^{\mathrm{ZF}}$ in (18) simplifies to

$$\mathsf{SIR}_{l,i}^{\mathrm{ZF}} = \frac{P_{\mathcal{S}} K^2 \left(\frac{1}{K} \sum_{k=1}^{K} \sqrt{\frac{P_{\mathcal{R}} F_{k,l} X_{k,l,i}}{L^2 M P_{\mathcal{S}}}} \right)^2}{M \, N_0 \, B \, \left(1 + K \, \frac{1}{K} \sum_{k=1}^{K} \frac{P_{\mathcal{R}} F_{k,l} \, X_{k,l,i}}{L^2 M E_{k,l} Y_{k,l,i} P_{\mathcal{S}}} \right)}.$$

Under the assumption that $\{E_{k,l}\}$ and $\{F_{k,l}\}$ are positive and bounded, it is straightforward but tedious to show that the variances of $d_{k,l,i}$ and $f_{k,l,i}^2$ are bounded $\forall k, l, i$. Hence the Kolmogorov condition is satisfied for each of these sequences (as a function of k) and it follows from Theorem 1.8.D in [32] that as $K \to \infty$

$$\sum_{k=1}^{K} \frac{\sqrt{F_{k,l} X_{k,l,i}}}{K} - \sum_{k=1}^{K} \frac{\mathbb{E}\left[\sqrt{F_{k,l} X_{k,l,i}}\right]}{K} \xrightarrow{\text{w.p.1}} 0$$

and

$$\frac{1}{K}\sum_{k=1}^{K} \frac{F_{k,l} X_{k,l,i}}{E_{k,l} Y_{k,l,i}} - \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}\left[\frac{F_{k,l} X_{k,l,i}}{E_{k,l} Y_{k,l,i}}\right] \xrightarrow{\text{w.p.1}} 0.$$

Now applying Theorem 1.7 in [32], we obtain

$$\mathsf{SIR}_{l,i}^{\text{ZF}} \xrightarrow{\text{w.p.1}} \frac{K^2 \left(\mathbb{E}\left[\sqrt{F_{k,l} X_{k,l,i}}\right]\right)^2}{N_0 B \left(\frac{L^2 M^2}{P_{\mathcal{R}}} + \frac{KM}{P_{\mathcal{S}}} \mathbb{E}\left[\frac{F_{k,l} X_{k,l,i}}{E_{k,l} Y_{k,l,i}}\right]\right)} + o(K),$$

Letting $\beta = P_{\mathcal{R}}/P_{\mathcal{S}}$, we find that SIR-maximizing power allocation (for fixed SNR) is achieved with

$$\beta^* = \sqrt{\frac{L^3 M}{K^2} \left(\mathbb{E}\left[\frac{F_{k,l} X_{k,l,i}}{E_{k,l} Y_{k,l,i}}\right]\right)^{-1}}$$

resulting in ($\mathsf{SNR} \gg 1$)

$$\mathsf{SIR}_{l,i}^{\text{ZF}} \xrightarrow{\text{w.p.1}} \frac{2K\mathsf{SNR}}{LM} \frac{\left(\mathbb{E}\left[\sqrt{F_{k,l} X_{k,l,i}}\right]\right)^2}{\left(\sqrt{LM} + \sqrt{\mathbb{E}\left[\frac{F_{k,l} X_{k,l,i}}{E_{k,l} Y_{k,l,i}}\right]}\right)^2} + o(K). \tag{19}$$

This key result gives a complete picture in terms of how distributed MIMO relaying impacts the SIR statistics for each transmit stream. First, we observe that using ZF-based MIMO relays that exploit the differences in the spatial signatures of the interfering data signals, the effective channel between the source-destination pairs was *orthogonalized* to increase spectral efficiency; creating uncoupled parallel channels for each stream and the multi-stream interference arising from the simultaneous transmission of multiple users and multiple streams per user is completed eliminated. Next, we observe from (19) that $\mathsf{SIR}_{l,i}$ scales *linearly* in the number of relay terminals, K providing higher energy efficiency. This can be interpreted as *distributed array gain*, since it is realized without requiring any cooperation among the relay terminals. In other words, the required transmit power to achieve a given target data rate reduces inversely proportionally with the number of relay terminals. This increased power efficiency could help provide much better coverage in cellular/WLAN networks. Finally, we observe from the convergence result in (19) that distributed signal processing by multiple relay terminals realizes *diversity gain* arising from the deterministic scaling behavior of $\mathsf{SIR}_{l,i}$ in the number of relay terminals K. Hence in the limit of infinite number of relays, our result suggests that a Shannon capacity exists even for slow fading (non-ergodic) channels [35]. This phenomenon of ”relay ergodization” can be interpreted as a form of statistical averaging that ensures the convergence of the SIR statistics to a deterministic scaling behavior even if the fading processes affecting the individual relays are not ergodic. Even more importantly, the deterministic scaling behavior also suggests that the lack of CSI knowledge at the destination terminals does not degrade performance in the limit of infinite number of relay terminals.

Now substituting (19) into (13), we obtain

$$\mathsf{C}^{\mathrm{ZF}} \xrightarrow{\text{w.p.1}} \frac{LM}{2}\log_2\left(\frac{2K\mathsf{SNR}}{LM}\frac{\left(\mathbb{E}\left[\sqrt{F_{k,l}X_{k,l,i}}\right]\right)^2}{\left(\sqrt{LM}+\sqrt{\mathbb{E}\left[\frac{F_{k,l}X_{k,l,i}}{E_{k,l}Y_{k,l,i}}\right]}\right)^2}+o(K)\right).$$

Applying (3)-(4) to C^{ZF}, we obtain the high $\frac{E_b}{N_0}$ power-bandwidth tradeoff relationships in (12) for ZF MIMO relaying.

If $\mathsf{SNR} \ll 1$ (low $\frac{E_b}{N_0}$ regime), then $\mathsf{SIR}^{\mathrm{ZF}}_{l,i}$ in (18) simplifies to

$$\mathsf{SIR}^{\mathrm{ZF}}_{l,i} = \frac{P_{\mathcal{S}}}{N_0\,B}\frac{P_{\mathcal{R}}}{N_0\,B}\frac{K^2}{L^2\,M^3}\left(\frac{1}{K}\sum_{k=1}^{K}\sqrt{E_{k,l}F_{k,l}X_{k,l,i}Y_{k,l,i}}\right)^2.$$

As before, using Theorem 1.8.D of [32], we have

$$\sum_{k=1}^{K}\frac{\sqrt{E_{k,l}F_{k,l}X_{k,l,i}Y_{k,l,i}}}{K}-\sum_{k=1}^{K}\frac{\mathbb{E}\left[\sqrt{E_{k,l}F_{k,l}X_{k,l,i}Y_{k,l,i}}\right]}{K}\xrightarrow{\text{w.p.1}}0$$

as $K\to\infty$, yielding

$$\mathsf{SIR}^{\mathrm{ZF}}_{l,i} \xrightarrow{\text{w.p.1}} \frac{P_{\mathcal{S}}}{N_0\,B}\frac{P_{\mathcal{R}}}{N_0\,B}\frac{K^2}{L^2\,M^3}\left(\mathbb{E}\left[\sqrt{E_{k,l}F_{k,l}X_{k,l,i}Y_{k,l,i}}\right]\right)^2+o(K).$$

Letting $\beta = P_{\mathcal{R}}/P_{\mathcal{S}}$, we find that SIR-maximizing power allocation (for fixed SNR) is achieved with $\beta^* = L/K$ resulting in ($\mathsf{SNR} \ll 1$)

$$\mathsf{C}^{\mathrm{ZF}} = \frac{LM}{2}\log_2\left(1+\mathsf{SNR}^2\left(\frac{K}{L^3M^3}\left(\mathbb{E}\left[\sqrt{E_{k,l}F_{k,l}X_{k,l,i}Y_{k,l,i}}\right]\right)^2+o(K)\right)\right).$$

Substituting $\mathsf{SNR} = \mathsf{C}\frac{E_b}{N_0}$ and solving for $\frac{E_b}{N_0}$, we obtain the result in (9). The rest of the proof follows from the strict convexity of

$$\frac{2^{2\mathsf{C}/(LM)}-1}{\mathsf{C}^2}$$

in C for all $\mathsf{C}\geq 0$. □

Simulation Example 1: SIR Statistics. In this section, we consider an MIRN with $L = 2$, $M = 1$ and $N = 2$ and analyze (based on Monte Carlo simulations) the SIR statistics of distributed MIMO relaying techniques based on the ZF algorithm and compare the performance with that under direct transmissions. For direct transmission, we assume that two users share the total fixed power P equally and communicate over a fading interference channel with single-user decoders at the destination terminals. For fair comparison with distributed relaying, no transmit CSI is considered at the sources while the receivers possess perfect CSI. Denoting the overall channel gain (including path loss, shadowing and fading) between source $i \in \{1, 2\}$ and destination $j \in \{1, 2\}$ by $\xi_{i,j}$, the SIR for each stream under direct transmission is

$$\mathsf{SIR}_{l,i}^{\text{direct}} = \frac{|\xi_{i,i}|^2 \frac{P}{2}}{N_0 B + |\xi_{i,j}|^2 \frac{P}{2}}, \quad i \in \{1, 2\},\ j \neq i.$$

We set $\mathsf{SNR} = 20\ \text{dB}$ and plot the cumulative distribution function (CDF) of $\mathsf{SIR}_{l,i}$ for direct transmission and for distributed MIMO relaying based on the ZF scheme, and varying $K = 1, 2, 4, 8, 16$ in Fig. 5.

As predicted by (19), we observe that the mean of SIR increases by $3\ \text{dB}$ for every doubling of K due to the array gain improvement proportional in the number of relay terminals. Furthermore, the tightening of the SIR CDFs with increasing K demonstrate the distributed diversity gain due to processing by multiple relay terminals. Finally, we note the huge improvement in SIR with respect to direct transmissions due to increased interference cancelation capability of the relay-assisted wireless network.

Simulation Example 2: Power and Bandwidth Efficiency. We consider an MIRN with $K = 10$, $L = 2$, $M = 1$ and $N = 2$ and plot (based on Monte Carlo simulations) power-bandwidth tradeoff curves for the upper-limit based on the cutset bound, L-MIRN schemes using MF, ZF and L-MMSE techniques and direct transmission in Fig. 6.

Our analytical results in (10)-(12) supported with the numerical results in Fig. 6 show that even low complexity linear-beamformer relay schemes could potentially yield *significant power and bandwidth savings* over direct transmissions. We observe that a significant portion of the set of energy and spectral efficiency pairs within the cutset outer bound (that is infeasible with direct transmission) is covered by our low-complexity L-MIRN schemes. From our asymptotic analysis, we find that $\frac{E_b}{N_0}$ reduces like $K^{-1/2}$ in the low $\frac{E_b}{N_0}$ regime for the MF, ZF and L-MMSE schemes and like K^{-1} in the high $\frac{E_b}{N_0}$ regime for the ZF and L-MMSE schemes.[8] Note that ZF and L-MMSE schemes achieve *optimal energy scaling* (in K) for high $\frac{E_b}{N_0}$ verified through the best-case power-bandwidth tradeoff in Theorem 1. Furthermore, unlike MF, the spectral efficiency of ZF and L-MMSE based MIRN grows without bound with $\frac{E_b}{N_0}$ due to their interference cancelation capability and achieves the *optimal* high-SNR slope of $S_\infty = \frac{LM}{2}$.

All linear-beamforming schemes achieve the highest energy efficiency at a finite spectral efficiency. In other words, the most efficient power utilization under linear-beamformer relays is achieved at a *finite bandwidth* and there is *no power-bandwidth tradeoff* above a certain bandwidth. Additional bandwidth requires more power. A similar observation was made in [36] and [37] in the context of Gaussian parallel relay networks. This phenomenon is due to noise amplification during relay linear-beamforming operations; which significantly degrades performance at high bandwidths when the MIRN becomes noise-limited. We find that the ZF scheme performs worse than the MF and L-MMSE schemes in the low $\frac{E_b}{N_0}$ regime because of its inherent inability of noise suppression.

One solution to the problem of noise amplification in the low SNR regime is *bursty* transmission [38]. For the duty cycle parameter $\alpha \in \{0, 1\}$, this means that the sources and relays transmit only α fraction of time over which they consume total power P/α and remain silent otherwise; and hence satisfying the average power constraints. The result of bursty transmission is that the network is forced to operate in the high SNR regime at the expense of lower spectral efficiency. This is achieved, for instance under the ZF scheme, through the adjustment of signal burstiness by choosing the duty cycle parameter α such that even though $\mathsf{SNR} \ll 1$, the SIR for each stream in (18) simplifies to (note the additional α term in the denominator)

$$\mathsf{SIR}_{l,i}^{\mathrm{ZF}} = \frac{P_{\mathcal{S}} K^2 \left(\frac{1}{K} \sum_{k=1}^{K} \sqrt{\frac{P_{\mathcal{R}} F_{k,l} X_{k,l,i}}{L^2 M P_{\mathcal{S}}}} \right)^2}{\alpha\, M\, N_0\, B\, \left(1 + K \frac{1}{K} \sum_{k=1}^{K} \frac{P_{\mathcal{R}} F_{k,l}\, X_{k,l,i}}{L^2 M E_{k,l} Y_{k,l,i} P_{\mathcal{S}}} \right)}.$$

as in the high $\frac{E_b}{N_0}$ regime and the network spectral efficiency is computed as

$$\mathsf{C}^{\mathrm{ZF}} = \frac{\alpha}{2} \sum_{l=1}^{L} \sum_{i=1}^{M} \mathbb{E}_{\{\mathbf{H}_{k,l}, \mathbf{G}_{k,l}\}} \left[\log_2 \left(1 + \mathsf{SIR}_{l,i}^{\mathrm{ZF}} \right) \right].$$

[8] L-MMSE converges to ZF as $\mathsf{SNR} \to \infty$ and to MF as $\mathsf{SNR} \to 0$.

Hence, the results of Theorem 2 in (11) can immediately be applied, with slight modifications, resulting in the power-bandwidth tradeoff relation

$$\frac{E_b}{N_0}^{\mathrm{ZF}}(\mathsf{C}) = \frac{2^{2\mathsf{C}/(\alpha LM)}}{2\mathsf{C}/(\alpha LM)} \frac{\left(\sqrt{\epsilon_2} + \sqrt{L\,M}\right)^2}{K\,\epsilon_3^2} + o(\frac{1}{K}), \quad K \to \infty,$$

and consequently the energy and spectral efficiency performance can be quantified by

$$\frac{E_b}{N_0}^{\mathrm{ZF}}_{\mathrm{min}} = \frac{(\ln 2)\,L\,M}{2\,K\,\epsilon_3^2}\left(\sqrt{\epsilon_2} + \sqrt{L\,M}\right)^2 + o(\frac{1}{K}) \qquad \text{and} \qquad S_0^{\mathrm{ZF}} = \alpha LM,$$

in the low $\frac{E_b}{N_0}$ regime and

$$\frac{E_b}{N_0}^{\mathrm{ZF}}_{\mathrm{imp}} = \frac{L\,M}{2\,K\,\epsilon_3^2}\left(\sqrt{\epsilon_2} + \sqrt{L\,M}\right)^2 + o(\frac{1}{K}) \qquad \text{and} \qquad S_\infty^{\mathrm{ZF}} = \frac{\alpha LM}{2},$$

in the high $\frac{E_b}{N_0}$ regime. As a result, we have shown that with sufficient amount of burstiness, the optimal energy scaling of K^{-1} can be achieved with the ZF (and L-MMSE) linear relaying protocols; while the spectral efficiency slopes for the low and high $\frac{E_b}{N_0}$ regimes scale down by the duty cycle factor α. This final remark establishes the asymptotic optimality of linear multi-antenna relaying in the sense that with proper signaling it can achieve the best possible (i.e., as in the cutset bound) energy and spectral efficiency scaling for any SNR. We also emphasize that our results prove that the energy scaling of K^{-1} is achievable with linear relaying (provided the necessary relay interference cancelation mechanisms) while previous work showed that it could only achieve the scaling of $K^{-1/2}$ [26]. In Fig. 7, we plot the power-bandwidth tradeoff under the ZF scheme (setting $K = 10$, $L = 2$, $M = 1$, $N = 2$) for various values of α. Clearly, we find that in the lower spectral efficiency (and hence lower SNR) regime, it is desirable to increase the level of burstiness by reducing the α parameter in order to achieve higher energy efficiency.

Simulation Example 3: Enhancements from MIMO. We consider the L-MMSE MIRN scheme with $K = 10$ and $L = 2$, and plot (based on Monte Carlo simulations) power-bandwidth tradeoff curves for different values of M and N, to understand the impact of MIMO techniques at the source-destination pair (larger M) and relay (larger N) terminals on energy and spectral efficiency. From Fig. 8, it is clear that multiple antennas at the relay terminals improve energy efficiency (through the downward shift of the power-bandwidth tradeoff curve) while multiple antennas at the source-destination pairs improve spectral efficiency (through the improvement in the wideband slope and high SNR slope of the power-bandwidth tradeoff curve).

V. CONCLUSIONS

As an additional leverage for supporting high data rate next generation wireless networks, we proposed low-complexity distributed MIMO relaying algorithms that exploit locally available channel state information (CSI) at each relay terminal to beamform multiple users' signals at the intended destinations. From a Shannon-theoretic perspective, we analyzed the power-bandwidth tradeoff for various MIMO relaying schemes and demonstrated significant gains in terms of power efficiency, bandwidth efficiency and link reliability over direct transmissions. We quantified these gains by investigating the energy and spectral efficiency scaling in the low and high SNR regimes. We established that in the limit of large number of relay terminals ($K \to \infty$), linear-beamformer relays achieve asymptotically optimal power-bandwidth tradeoff at any SNR with the energy efficiency scaling like K. Finally, we verified our results through the numerical investigation of SIR statistics and power-bandwidth tradeoffs.

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

[6] 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.

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

[8] 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.

[9] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*, Cambridge University Press, Cambridge, UK, 1st edition, 2003.

[10] T. Guess and M. K. Varanasi, "An information-theoretic derivation of the MMSE decision-feedback equalizer," in *Proc. 36th Allerton Conf.*, Monticello, IL, Sep. 1998.

[11] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, Autumn 1996.

[12] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity, part 1: System description," *IEEE Trans. Comm.*, vol. 51, no. 11, pp. 1939–1948, Nov. 2003.

[13] J. N. Laneman and G. Wornell, "Distributed space-time coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2415–2425, Oct. 2003.

[14] R. U. Nabar, H. Bölcskei, and F. W. Kneubühler, "Fading relay channels: Performance limits and space-time signal design," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1099–1109, Aug. 2004.

[15] R. Pabst, B. Walke, D. Schultz, P. Herhold, H. Yanikomeroglu, S. Mukherjee, H. Viswanathan, M. Lott, W. Zirwas, M. Dohler, H. Aghvami, D. Falconer, and G. Fettweis, "Relay-based deployment concepts for wireless and mobile broadband radio," *IEEE Communications Magazine*, pp. 80–89, Sept. 2004.

[16] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, Jun. 2002.

[17] A. Lozano, A. Tulino, and S. Verdú, "Multiple-antenna capacity in the low-power regime," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2527–2543, Oct. 2003.

[18] S. Shamai (Shitz) and S. Verdú, "The impact of flat-fading on the spectral efficiency of CDMA," *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 1302–1327, May 2001.

[19] G. Caire, D. Tuninetti, and S. Verdú, "Suboptimality of TDMA in the low-power regime," *IEEE Trans. Inf. Theory*, vol. 50, no. 4, pp. 608–620, Apr. 2004.

[20] A. Lapidoth, I. E. Telatar, and R. Urbanke, "On wide-band broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3250–3258, Dec. 2003.

[21] T. Muharemović and B. Aazhang, "Robust slope region for wideband CDMA with multiple antennas," in *Proc. 2003 IEEE Information Theory Workshop*, Paris, France, March 2003, pp. 26–29.

[22] A. El Gamal and S. Zahedi, "Minimum energy communication over a relay channel," in *Proc. 2003 IEEE International Symposium on Information Theory (ISIT'03)*, Yokohama, Japan, June-July 2003, p. 344.

[23] X. Cai, Y. Yao, and G. Giannakis, "Achievable rates in low-power relay links over fading channels," *IEEE Trans. Communications*, vol. 53, no. 1, pp. 184–194, Jan. 2005.

[24] Ö. Oyman and A. J. Paulraj, "Spectral efficiency of relay networks in the power-limited regime," in *Proc. Allerton Conf. on Communication, Control and Computing*, Monticello,IL, Sep. 2004.

[25] A. Lozano, A. Tulino, and S. Verdú, "High-SNR power offset in multi-antenna communication," in *Proc. IEEE International Symposium on Information Theory (ISIT'04)*, Chicago, IL, June 2004.

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

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

[28] 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.

[29] H. Bölcskei, R. U. Nabar, Ö. Oyman, and A. J. Paulraj, "Capacity scaling laws in MIMO relay networks," *IEEE Trans. Wireless Communications*, to appear in 2006.

[30] 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.

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

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

[33] Ö. Oyman and A. J. Paulraj, "Energy efficiency in MIMO relay networks under processing cost," in *Proc. Conference on Information Sciences and Systems (CISS'05)*, Baltimore, MD, March 2005.

[34] R. U. Nabar, Ö. Oyman, H. Bölcskei, and A. J. Paulraj, "Capacity scaling laws in MIMO wireless networks," in *Proc. Allerton Conf. on Communication, Control and Computing*, Monticello,IL, Sep. 2003.

[35] L. H. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, no. 2, pp. 359–378, May 1994.

[36] I. Maric and R. Yates, "Forwarding strategies for Gaussian parallel-relay networks," in *Proc. Conference on Information Sciences and Systems (CISS'04)*, Princeton, NJ, March 2004.

[37] B. Schein, *Distributed coordination in network information theory*, Ph.D. thesis, Massachusetts Institute of Technology, Sep. 2001.

[38] A. El Gamal, M. Mohseni, and S. Zahedi, "On reliable communication over additive white Gaussian noise relay channels," *IEEE Trans. Inf. Theory*, Sep. 2004, submitted.

Table 1. L-MIRN Beamformer Schemes.

Channel Description	MF	ZF	L-MMSE
$\{\mathcal{S}_l\}_{l=1}^{L} \rightarrow \mathcal{R}_k$ Links $: \mathbf{H}_k = \begin{bmatrix} \sqrt{E_{k,1}}\,\mathbf{H}_{k,1}^T \\ \vdots \\ \sqrt{E_{k,L}}\,\mathbf{H}_{k,L}^T \end{bmatrix}^T$	$\mathbf{U}_k = \mathbf{H}_k^H$	$\mathbf{U}_k = (\mathbf{H}_k^H\mathbf{H}_k)^{-1}\mathbf{H}_k^H$	$\mathbf{U}_k = (\frac{MN_0B}{P_\mathcal{S}}\mathbf{I} + \mathbf{H}_k^H\mathbf{H}_k)^{-1}\mathbf{H}_k^H$
$\mathcal{R}_k \rightarrow \{\mathcal{D}_l\}_{l=1}^{L}$ Links $: \mathbf{G}_k = \begin{bmatrix} \sqrt{F_{k,1}}\,\mathbf{G}_{k,1} \\ \vdots \\ \sqrt{F_{k,L}}\,\mathbf{G}_{k,L} \end{bmatrix}$	$\mathbf{V}_k = \mathbf{G}_k^H$	$\mathbf{V}_k = \mathbf{G}_k^H(\mathbf{G}_k\mathbf{G}_k^H)^{-1}$	$\mathbf{V}_k = \mathbf{G}_k^H(\frac{NN_0B}{P_\mathcal{R}}\mathbf{I} + \mathbf{G}_k\mathbf{G}_k^H)^{-1}$

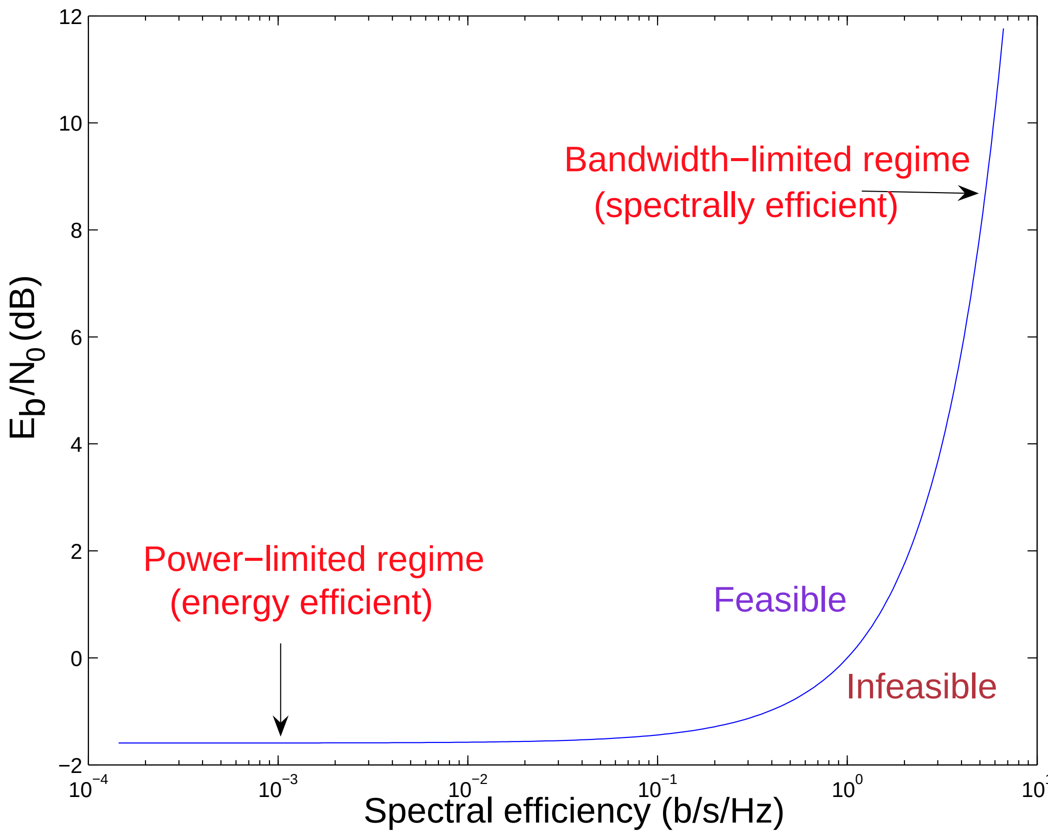

Fig. 1. Power-bandwidth tradeoff in the AWGN channel.

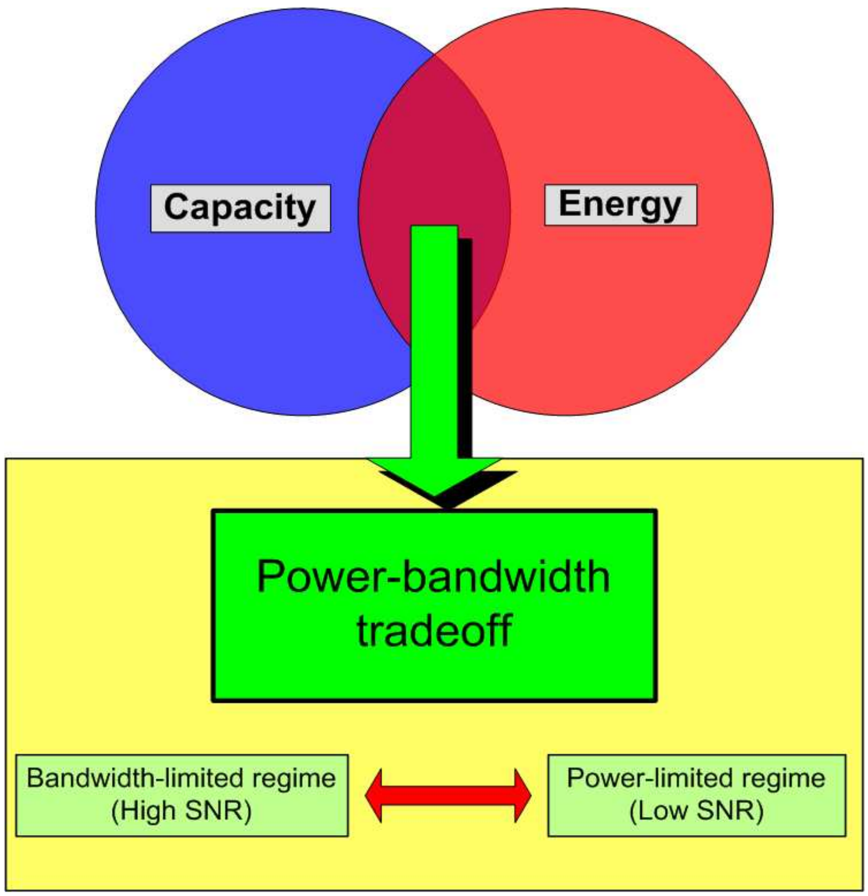

Fig. 2. Power-bandwidth tradeoff formulation to jointly capture capacity and energy gains of distributed MIMO relaying with special focus on low and high SNR regimes.

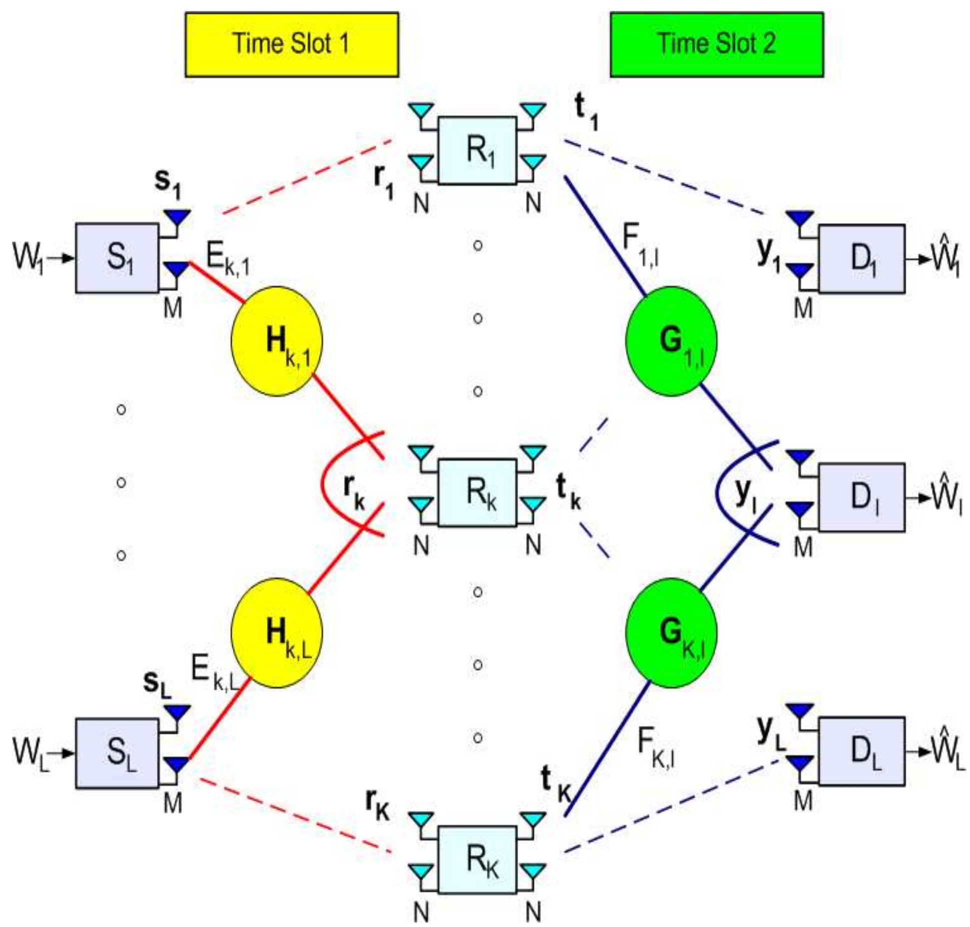

Fig. 3. MIRN source-relay and relay-destination channel models.

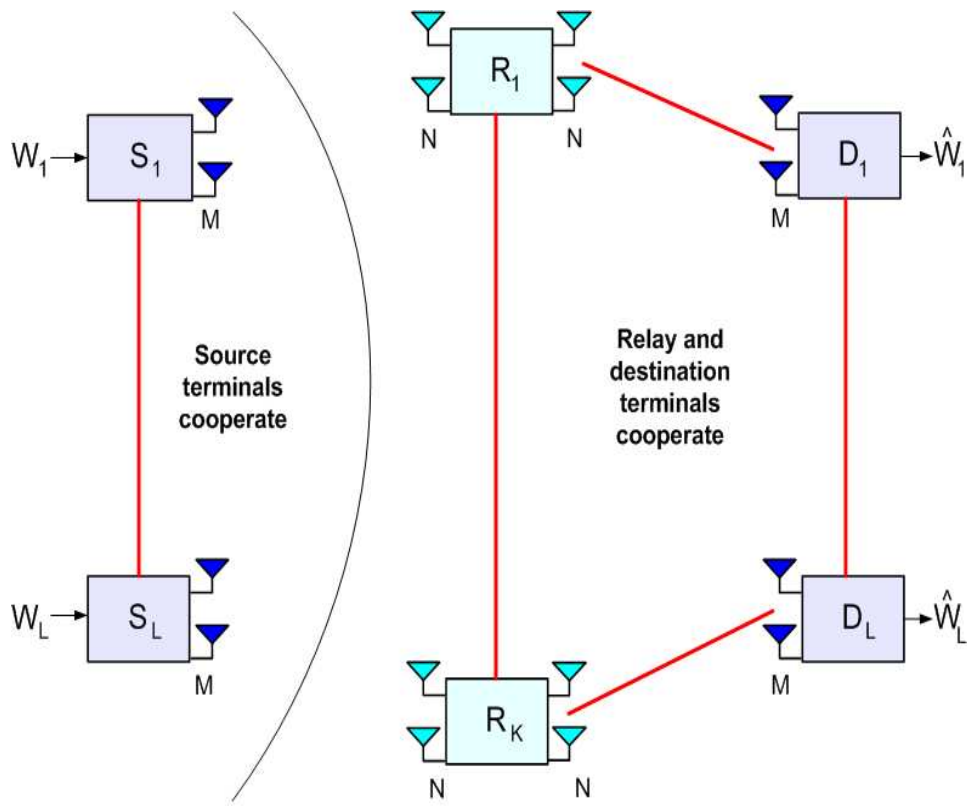

Fig. 4. Illustration of the broadcast cut over the MIRN.

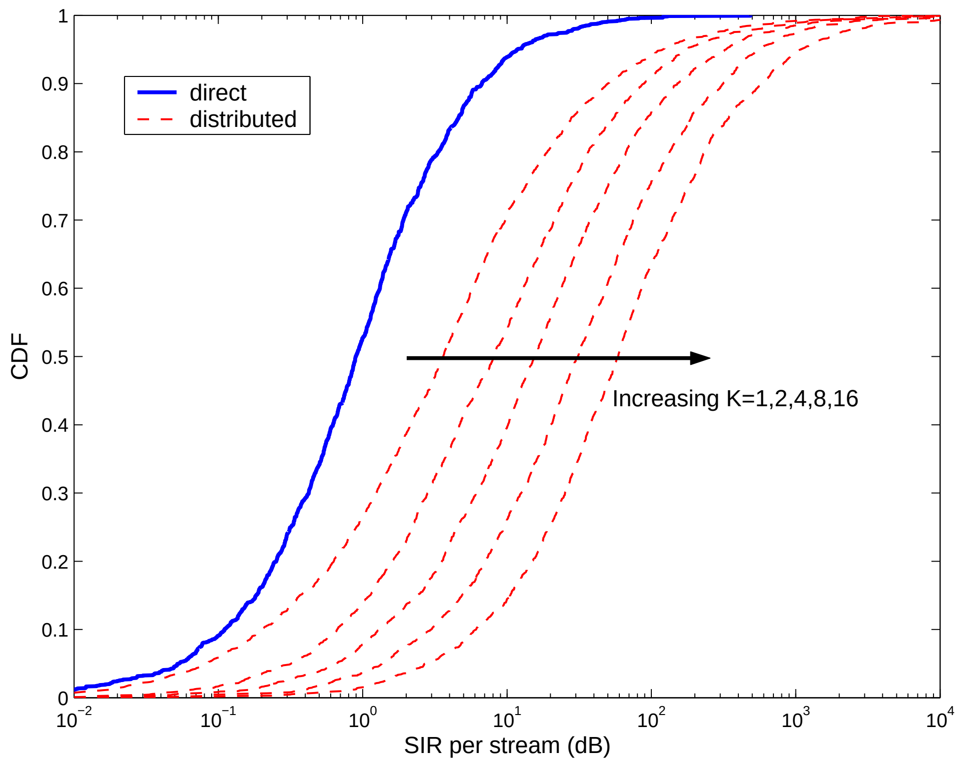

Fig. 5. CDF of SIR for direct and ZF based distributed MIMO relaying for various values of K at $\mathsf{SNR} = 20$ dB.

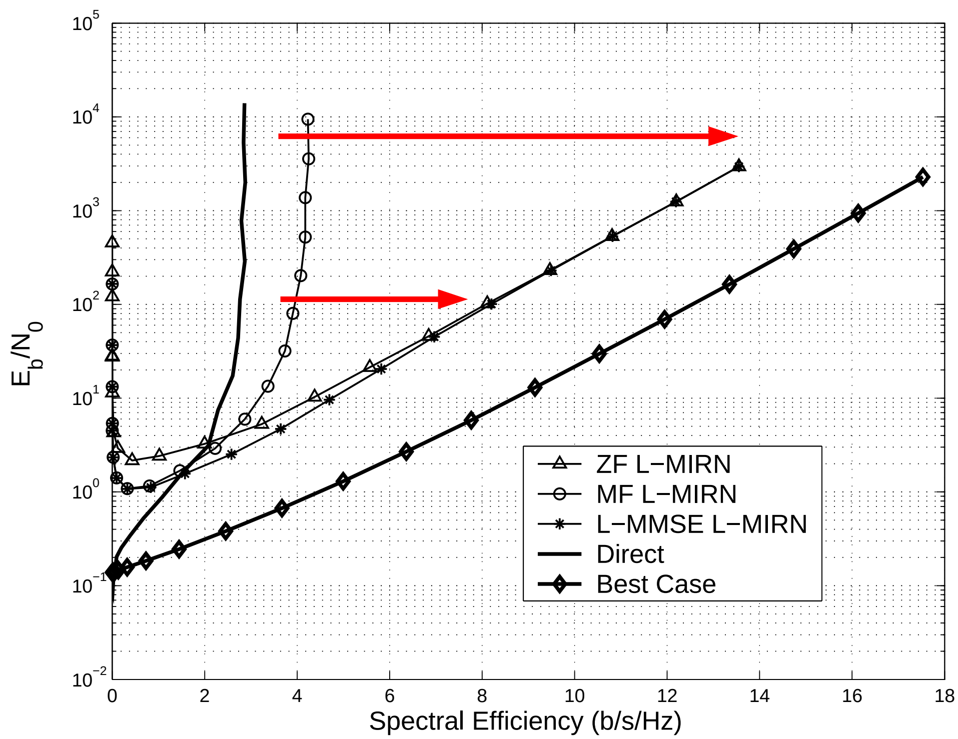

Fig. 6. MIRN power-bandwidth tradeoff comparison: Upper-limit, linear-beamformer relaying schemes and direct transmission.

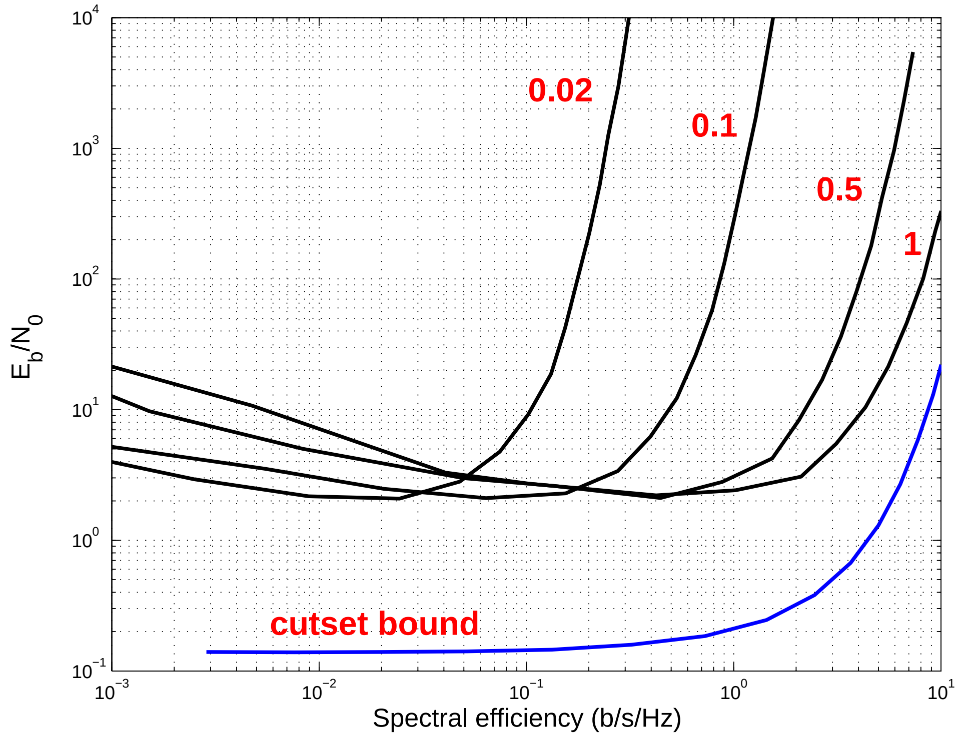

Fig. 7. Power-bandwidth tradeoff for the ZF MIRN scheme under bursty transmission for duty cycle parameters $\alpha = 0.02, 0.1, 0.5, 1$.

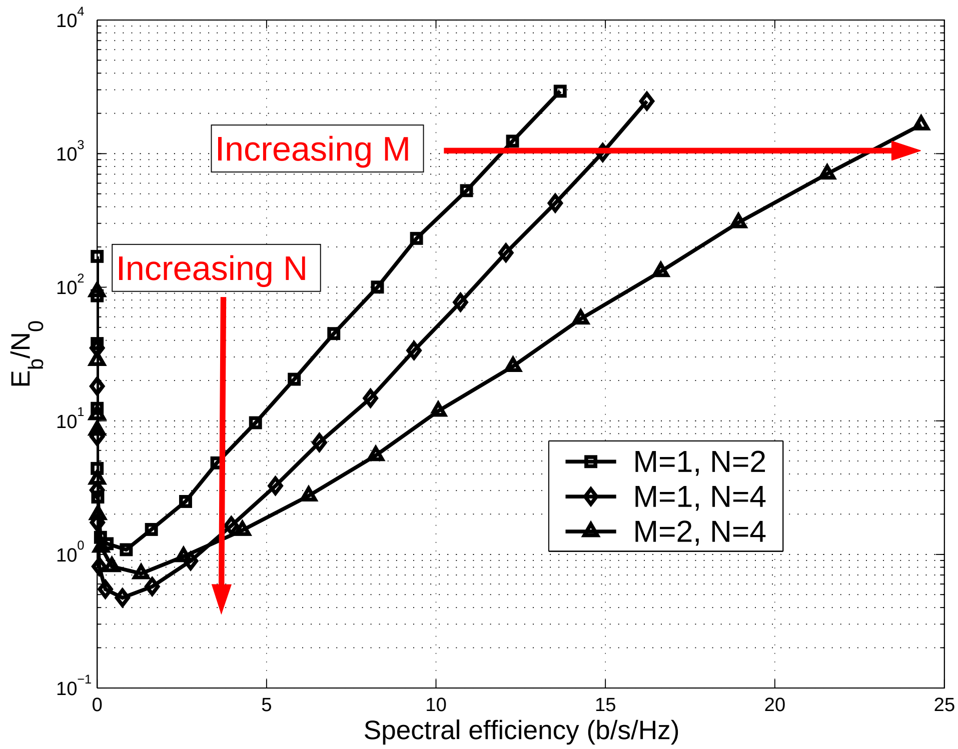

Fig. 8. MIRN power-bandwidth tradeoff for the L-MMSE relaying scheme with varying number of antennas at the source-destination pair (M) and relay terminals (N).