Project Number:	IST-026905
Project Title:	Multiple-Access Space-Time Coding Testbed
Project Coordinator:	C.F. Mecklenbräuker
Deliverable Number:	D3.2.2

Title of Deliverable:	Signalling/routing strategies and outage properties in Ad-hoc MIMO networks
Workpackage:	WP-3
Nature:	R
Dissemination level:	PU
Editor:	Ari Hottinen
Authors:	see list inside
Contractual Date of Deliverable:	Feb. 28, 2009
Actual Date of Delivery:	Feb. 19, 2009

Abstract:

We consider signalling aspects of distributed beamforming in a relay network and propose new signalling schemes based on relay weight perturbations. Next, we study coalition formation in a cooperative (virtual) MIMO network and propose efficient optimization approach as means to determine cooperative users. Similar optimization approaches are used also in device cooperation in relay networks. Furthermore, we establish the DM tradeoff curve for multi-antenna multi-user relay network under the assumption that relays employ unitary transformations. Finally, we propose and study compact models for the evaluation of scaling laws in large random (ad-hoc) networks.

Contents

Authors

Peter Fertl and Gerald Matz
Institute of Communications and Radio-Frequency Engineering
Vienna University of Technology
Vienna, Austria

Ari Hottinen
Nokia Research Center
Helsinki, Finland

Emanuele Viterbo
DEIS, University of Calabria
Rende (CS), Italy

Cemal Akçaba and Helmut Bölcskei
Eidgenössische Technische Hochschule
8092 Zürich, Switzerland

Maxime Guillaud
FTW, Telecommunications Research Center
Vienna, Austria

Giuseppa Alfano
DELEN, Politecnico of Torino
Torino, Italy

Antonia Tulino
DIET,*Federico II* University
Napoli, Italy

Tiina Heikkinen
Dept. Computer Science, University of Helsinki
Helsinki, Finland

Executive Summary

This deliverable focuses on signalling, routing (cooperation), and outage aspects of ad-hoc MIMO networks.

Chapter 1 contributes to signalling mechanisms for distributed beamforming with multiple half-duplex amplify-and-forward relays. In the proposed method, distributed beamforming applies deterministic perturbations and scalable (1-bit) feedback from the destination to the relays and does not require, in contrast to previous work, CSI knowledge at the relays. Simulation results confirm that the proposed techniques closely approach optimum performance and have satisfactory tracking properties in time-varying environments.

Chapter 2 continues the work the previous Chapter, and considers multiplicative perturbations as opposed to additive perturbations. Multiplicative perturbations are based on Givens rotations and adapt the beamforming weights while guaranteeing a sum power constraint for the relays. This perturbation scheme is shown to be computationally efficient and easy to design, thus allowing for low-complexity relay nodes. An adaptation of the Givens rotation angle allows to approach optimum performance arbitrarily close. Numerical simulations demonstrate noticeable performance gains over additive perturbation schemes that have been exclusively considered up to now.

Chapter 3 studies optimal user pairing in a point-to-multipoint MIMO channel. The optimal user subsets (in terms of sum mutual information) are determined for multiplexing N users in N slots. In each slot at most two users are concurrently transmitting. The brute-force complexity of the subset selection problem is shown to require $N!!$ evaluations, while the complexity of the proposed (yet optimal) method is only $O(N^3)$.

Chapter 4 considers methods for determining user or device coalitions for collaborative signal transmission in ah-hoc relay systems. We restrict the work to a multiuser system, in which one or two devices are allowed transmit simultaneously in an uplink channel, using another device as a amplify-and-forward relay node. The problem is to determine for N total number of users,

the collaboration subsets comprising at most two devices, that are allowed to transmit simultaneously (one as source, the other as relay), under specified constraints on power and channel usage. The optimal pairing strategy is equivalent to a combinatorial optimization problem.

Chapter 5 analyzes fading relay networks, where L users with M-antenna each communicate with an N-antenna destination terminal through a set of half-duplex relays using a half-duplex relaying protocol with linear processing at the relay level. We derive the diversity multiplexing tradeoff curve under the assumption that relays employ unitary transformations. We observe that the benefit (in terms of diversity gain) of having K relay terminals is shared by all the users in the system. We further note that cooperation at the relay level cannot improve performance any further.

Chapter 6 extends work in the analysis of large random wireless ad hoc networks, where the underlying node distribution is almost ubiquitously assumed to be the homogeneous Poisson point process. Despite the nice analytical properties of such model, the spatial randomness has been, however, mainly exploited for connectivity and interference analysis, but has not yet been taken into account explicitly in the scaling laws evaluation. We move here a first step toward the evaluation of an upper bound on the aggregate throughput when the additional randomness due to the spatial node distribution is taken into account, together with the presence of power attenuation and random phase changes. This could be seen as a first attempt to connect some overoptimistic results based on stochastic channel model to more realistic analysis, relying on electromagnetic propagation arguments.

Finally, in Chapter 7, we propose and study a compact model for the evaluation of scaling laws in random wireless networks. The model allows the information-theoretic characterization of both point-to-point as well as distributed communications. It is analyzed under several assumptions about spatial correlation and the utilized channel state information and transmission schemes.

Chapter 1

Perturbation-based Distributed Beamforming for Wireless Relay Networks

P. Fertl, A. Hottinen and G. Matz

This chapter deals with distributed beamforming techniques for wireless networks with half-duplex amplify-and-forward relays. Existing schemes optimize the beamforming weights based on the assumption that channel state information (CSI) is available at the relays. We propose to use adaptive beamforming based on deterministic perturbations and limited feedback (1-bit) from the destination to the relays in order to avoid CSI at the relays. Two scalable perturbation schemes are considered and practical implementation aspects are addressed. Simulation results confirm that the proposed techniques closely approach optimum performance and have satisfactory tracking properties in time-varying environments.

1.1 Introduction

1.1.1 Background

Terminal cooperation in wireless networks has been recognized as a means to form virtual arrays that can realize spatial diversity in a distributed fashion. An important special case is distributed beamforming with half-duplex amplify-and-forward (AF) relays. The coherent AF scheme in [15] requires local channel phase information at the relays to achieve coherent phase combining with equal power at all relays. Beamforming with non-uniform power allocation (PA) under a sum power constraint [37][29] and

under individual relay power constraints [34] offers significant performance gains. However, optimal beamforming with PA places strong requirements regarding channel state information (CSI) at the relays. For centralized arrays with co-located antennas, this requirement has been circumvented by adaptive gradient beamforming techniques that iteratively adjust the beamforming weights using stochastic vector perturbations and limited feedback from the destination [9]. A related approach based on deterministic perturbations is presented in [48]. In a similar spirit, feedback-assisted distributed beamforming with phase perturbation in wireless networks was considered in [41] and extended to the multiuser context in [56]. However, both methods do not assume a relay setup and do not address distributed PA.

1.1.2 Contribution and Organization

We consider perturbation-based beamforming (PB-BF) with 1-bit feedback in a relay network. Under the assumption of a sum power constraint, the relays use the feedback bit to adapt their beamforming weights in order to maximize either the signal-to-noise ratio (SNR) or the received signal power at the destination. This approach does not require any CSI at the relays. Two different perturbation schemes are investigated, both of which are based on deterministic perturbation sets to avoid extensive signaling/feedback overhead. Within this context, we present a scalable protocol, discuss implementation aspects, and provide numerical performance comparisons. Simulation results corroborate that our approach can satisfactorily track time-varying channels in non-static environments. We note that in the context of wireless ad-hoc networks a related idea was touched upon in [38] without explicitly addressing the important practical problem of weight exchange.

The rest of the Chapter is organized as follows. Section 2.2 introduces the system model and Section 1.3 proposes perturbation-based distributed beamforming with 1-bit feedback. A comparison with optimum batch solutions is provided in Section 1.4. Section 1.5 discusses simulation results and conclusions are provided in Section 2.5.

1.2 System Model

We consider a perfectly synchronized wireless network with single antenna nodes where a single source $\mathcal{S}$ communicates with a single destination $\mathcal{D}$ via R half-duplex relays $\mathcal{R}_i$, $i = 1, \ldots, R$ (cf. Fig. 2.1). The half-duplex constraint

necessitates a two-hop protocol. In the first hop, $\mathcal{S}$ transmits the signal $\sqrt{P_s}s$ to the relays which receive

$$x_i = \sqrt{P_s}h_i s + w_i, \quad i = 1, \ldots, R\,. \tag{1.1}$$

Here, s is the transmit symbol normalized as $\mathbb{E}\{|s|^2\} = 1$ ($\mathbb{E}\{\cdot\}$ denotes expectation), P_s denotes the average transmit power of $\mathcal{S}$, h_i is the complex coefficient of the flat fading "backward" channel[1] between $\mathcal{S}$ and $\mathcal{R}_i$, and $w_i \sim \mathcal{CN}(0, N_0)$ denotes i.i.d. complex Gaussian noise. In the AF scenario considered, the second hop amounts to each relay transmitting a complex scaled version of the signal it has received, i.e.,

$$r_i = \alpha_i^* \lambda_i \, x_i\,, \quad \text{with } \lambda_i \triangleq \sqrt{\frac{P}{P_s|h_i|^2 + N_0}}. \tag{1.2}$$

Here, complex conjugation (superscript *) of the beamforming weights α_i will simplify notation later on, and λ_i is a power normalization factor such that the average relay power is $\mathbb{E}\{|r_i|^2|h_i\} = |\alpha_i|^2 P$. The destination receives $y = \sum_{i=1}^{R} g_i r_i + v$, where g_i denotes the complex coefficient of the "forward" channel between $\mathcal{R}_i$ and $\mathcal{D}$, and $v \sim \mathcal{CN}(0, N_0)$ is complex Gaussian noise. Inserting (2.1) and (2.2) yields the compound channel model[2]

$$y = \xi s + \eta, \quad \text{with } \xi \triangleq \boldsymbol{\alpha}^H \bar{\mathbf{h}}, \quad \eta \triangleq \boldsymbol{\alpha}^H \bar{\mathbf{G}} \mathbf{w} + v. \tag{1.3}$$

Here, $\bar{\mathbf{h}} \triangleq [\bar{h}_1 \ldots \bar{h}_R]^T$ with $\bar{h}_i \triangleq h_i g_i \lambda_i \sqrt{P_s}$, $\bar{\mathbf{G}} \triangleq \mathrm{diag}(\bar{g}_1, \ldots, \bar{g}_R)$ with $\bar{g}_i \triangleq g_i \lambda_i$, and $\mathbf{w} \triangleq [w_1 \ldots w_R]^T$. Since the weight vector $\boldsymbol{\alpha} \triangleq [\alpha_1 \ldots \alpha_R]^T$ enters also the noise part in (2.3), it demands careful design to prevent noise amplification.

From (2.3), the average power corresponding to the signal part of y and the SNR at $\mathcal{D}$ are respectively obtained as

$$P_{\mathcal{D}}(\boldsymbol{\alpha}) \triangleq \mathbb{E}\Big\{|\xi s|^2 \Big| \bar{\mathbf{h}}\Big\} = \boldsymbol{\alpha}^H \bar{\mathbf{h}} \bar{\mathbf{h}}^H \boldsymbol{\alpha} = \Big|\boldsymbol{\alpha}^H \bar{\mathbf{h}}\Big|^2, \tag{1.4}$$

$$\rho(\boldsymbol{\alpha}) \triangleq \frac{\mathbb{E}\Big\{|\xi s|^2 \Big| \bar{\mathbf{h}}\Big\}}{\mathbb{E}\Big\{|\eta|^2 \Big| \bar{\mathbf{G}}\Big\}} = \frac{1}{N_0} \frac{\boldsymbol{\alpha}^H \bar{\mathbf{h}} \bar{\mathbf{h}}^H \boldsymbol{\alpha}}{1 + \boldsymbol{\alpha}^H \bar{\mathbf{G}} \bar{\mathbf{G}}^H \boldsymbol{\alpha}}. \tag{1.5}$$

In the following, we will use $\gamma(\boldsymbol{\alpha})$ as generic notation for our objective function, which can either be $P_{\mathcal{D}}(\boldsymbol{\alpha})$ or $\rho(\boldsymbol{\alpha})$. The beamforming vector $\boldsymbol{\alpha}$ can be batch designed to maximize $\gamma(\boldsymbol{\alpha})$ subject to a specific relay power constraint. We resort to two types of power constraints: Constraining the

[1] Note that our discussion does not presume specific channel statistics.

[2] Superscript ^{T} (^{H}) denotes (Hermitian) transposition; $\mathrm{diag}(x_1, \ldots, x_m)$ is the $m \times m$ diagonal matrix with diagonal elements $x_1, \ldots, x_m$.

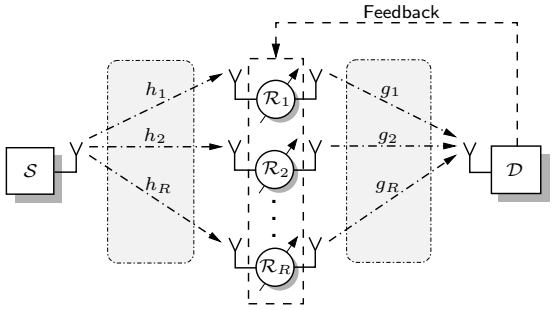


Figure 1.1: Wireless relay network with feedback.

complex beamforming weights to $|\alpha_i|^2 = 1$ (this amounts essentially to phase-matching at the relays [41]) ensures identical per-relay power $\mathbb{E}\{|r_i|^2|h_i\} = P$. In contrast, the total sum power constraint $\mathbb{E}\left\{\sum_{i=1}^{R}|r_i|^2|h_1,\ldots,h_R\right\} = P$ requires that the beamforming vector has unit Euclidean norm, $\|\boldsymbol{\alpha}\|^2 = 1$. However, such batch designs entail stringent requirements regarding the CSI available to the relays (via direct estimation or feedback, cf. Section 1.4).

1.3 Adaptive Perturbation-based Beamforming

1.3.1 Transmission Principle

To avoid CSI at the relays, we study distributed beamforming using feedback-assisted adaptive weight perturbation. The idea underlying this approach is to maximize the objective function $\gamma(\boldsymbol{\alpha})$ by adjusting the beamforming weights at the relays in an iterative manner using limited feedback (see Fig. 2.1). For co-located arrays with centralized processing similar ideas have been proposed in [9][48].

Transmission happens in frames consisting of a training interval $\mathcal{T}_\mathrm{p}$ and a data interval $\mathcal{T}_\mathrm{d}$. The relays use different beamforming weights to forward the training and data parts of each frame received from $\mathcal{S}$ to $\mathcal{D}$ according to (2.2). The idea is to apply the currently best beamforming vector, denoted $\boldsymbol{\alpha}_k$ (k is the frame index), to the data while using a perturbed version $\tilde{\boldsymbol{\alpha}}_k$ of the beamforming vector for the training portion. The destination evaluates the effectiveness of the perturbed weights and checks whether or not the perturbation improved the objective function $\gamma(\boldsymbol{\alpha}_k)$. It then provides the relays with one bit of feedback to indicate which beamforming vector shall be used to forward the data of the next frame.

In the proposed scheme, the weights $\tilde{\boldsymbol{\alpha}}_k$ are obtained by an additive perturbation from the data beamforming vector $\boldsymbol{\alpha}_k$. While in a centralized setup the perturbation vectors can be chosen randomly for each frame, our distributed setup necessitates a deterministic vector set, collected in a $R \times N$ matrix $\mathbf{Q} = [\mathbf{q}_0 \dots \mathbf{q}_{N-1}]$, from which the perturbation vector is picked in a cyclic fashion (cf. Section 1.3.4).

In the following, we present two variants of the proposed PB-BF scheme and describe the individual steps in more detail. For this discussion, we assume that all channel coefficients remain constant during the weight adaptation process.

1.3.2 Take/Reject (T/R) Perturbation

The perturbed weights for the kth frame are computed as

$$\tilde{\boldsymbol{\alpha}}'_k = \boldsymbol{\alpha}_k + \mu\, \mathbf{q}_{k \bmod N}, \tag{1.6}$$

where μ is a step-size parameter determining the adaptation rate, followed by proper normalization, i.e., $\tilde{\boldsymbol{\alpha}}_k = \tilde{\boldsymbol{\alpha}}'_k / \|\tilde{\boldsymbol{\alpha}}'_k\|$ in case of a sum power constraint and $\tilde{\alpha}_{k,i} = \tilde{\alpha}'_{k,i} / |\tilde{\alpha}'_{k,i}|$ for a per-relay power constraint. The weights $\tilde{\boldsymbol{\alpha}}_k$ are applied to the training sequence received at the relays, which is then forwarded to $\mathcal{D}$. At $\mathcal{D}$, the known training sequence and the receive signal are used to evaluate the performance of $\tilde{\boldsymbol{\alpha}}_k$ within $\mathcal{T}_\mathrm{p}$ according to the objective function, i.e., $\tilde{\gamma}_k = \gamma(\tilde{\boldsymbol{\alpha}}_k)$. Recall that $\gamma(\cdot)$ represents either the received signal power $P_\mathcal{D}(\cdot)$ in (1.4) or the SNR $\rho(\cdot)$ in (2.4). The actual estimation of these quantities will be addressed in Section 1.3.5.

The destination then compares $\tilde{\gamma}_k$ to the performance $\gamma_k = \gamma(\boldsymbol{\alpha}_k)$ achieved with the beamforming vectors $\boldsymbol{\alpha}_k$ that up to this point performed best. If $\tilde{\gamma}_k \leq \gamma_k$, $\tilde{\boldsymbol{\alpha}}_k$ does not perform better than $\boldsymbol{\alpha}_k$ and hence the relays should stick with $\boldsymbol{\alpha}_k$ for the data in the next transmission frame ("reject" $\tilde{\boldsymbol{\alpha}}_k$). Otherwise ($\tilde{\gamma}_k > \gamma_k$), the beamforming vector $\tilde{\boldsymbol{\alpha}}_k$ improves on $\boldsymbol{\alpha}_k$ and should thus be used in the next frame to transmit the data ("take" $\tilde{\boldsymbol{\alpha}}_k$). This rationale can be implemented be letting $\mathcal{D}$ provide the relays with a single bit of feedback, given by

$$c_k = \mathrm{u}(\tilde{\gamma}_k - \gamma_k),$$

where u(.) denotes the unit step function. Depending on the feedback bit, the relays update the data beamforming vector for the next frame as

$$\boldsymbol{\alpha}_{k+1} = \begin{cases} \boldsymbol{\alpha}_k, & \text{if } c_k = 0, \\ \tilde{\boldsymbol{\alpha}}_k, & \text{if } c_k = 1. \end{cases}$$

The vector $\boldsymbol{\alpha}_{k+1}$ will be the basis for the next perturbation according to (1.6). The destination performs the corresponding update $\gamma_{k+1}=\max\{\tilde{\gamma}_k,\gamma_k\}$. This process continues in an iterative manner. During the first frame, the perturbation scheme is initialized with $\tilde{\boldsymbol{\alpha}}_0=\boldsymbol{\alpha}_0$ and $\gamma_0=0$.

1.3.3 Plus/Minus (P/M) Perturbation

T/R perturbation has the advantage that performance never deteriorates, i.e., $\gamma(\boldsymbol{\alpha}_{k+1})\geq\gamma(\boldsymbol{\alpha}_k)$. On the other hand, in many cases the perturbation (1.6) will not yield an improvement, which entails $\boldsymbol{\alpha}_{k+1}=\boldsymbol{\alpha}_k$ and hence slow adaptation. We next discuss an alternative perturbations scheme with faster adaptation rate. Here, the training interval $\mathcal{T}_\mathrm{p}$ is split into two halves $\mathcal{T}_\mathrm{p}^+$ and $\mathcal{T}_\mathrm{p}^-$ for which different perturbed beamforming vectors are used, i.e.,[3]

$$\tilde{\boldsymbol{\alpha}}_k^+=\frac{\boldsymbol{\alpha}_k+\mu\,\mathbf{q}_{k\bmod N}}{\|\boldsymbol{\alpha}_k+\mu\,\mathbf{q}_{k\bmod N}\|},\quad \tilde{\boldsymbol{\alpha}}_k^-=\frac{\boldsymbol{\alpha}_k-\mu\,\mathbf{q}_{k\bmod N}}{\|\boldsymbol{\alpha}_k-\mu\,\mathbf{q}_{k\bmod N}\|}. \tag{1.7}$$

The destination $\mathcal{D}$ then measures the performance of $\tilde{\boldsymbol{\alpha}}_k^+$ and $\tilde{\boldsymbol{\alpha}}_k^-$ by evaluating the objective function according to $\tilde{\gamma}_k^+=\gamma(\tilde{\boldsymbol{\alpha}}_k^+)$ and $\tilde{\gamma}_k^-=\gamma(\tilde{\boldsymbol{\alpha}}_k^-)$ within $\mathcal{T}_\mathrm{p}^+$ and $\mathcal{T}_\mathrm{p}^-$, respectively. While in principle we could pick the beamforming weights corresponding to the maximum of $\tilde{\gamma}_k^+$, $\tilde{\gamma}_k^-$, and $\gamma_k=\gamma(\boldsymbol{\alpha}_k)$ (the performance of the current data beamforming vector), 1-bit feedback can only support binary choices. Hence, $\boldsymbol{\alpha}_k$ will be discarded in any case. $\mathcal{D}$ broadcasts the feedback bit $c_k=\mathrm{u}(\tilde{\gamma}_k^-\!-\tilde{\gamma}_k^+)$ to the relays, indicating whether the "plus" perturbation $\tilde{\boldsymbol{\alpha}}_k^+$ or the "minus" perturbation $\tilde{\boldsymbol{\alpha}}_k^-$ performs better. In the next frame, the relays use the beamforming vector

$$\boldsymbol{\alpha}_{k+1}=\begin{cases}\tilde{\boldsymbol{\alpha}}_k^+, & \text{if } c_k=0,\\ \tilde{\boldsymbol{\alpha}}_k^-, & \text{if } c_k=1.\end{cases}$$

Although P/M perturbation shows typically faster adaptation than T/R, sometimes both perturbations in (1.7) deteriorate the performance with respect to γ_k. Thus, P/M performance may fluctuate continually. Furthermore, only half of the training interval can be used to estimate each of $\tilde{\gamma}_k^+$ and $\tilde{\gamma}_k^-$.

1.3.4 Perturbation Set

Vector normalization of the weights in (1.6) and (1.7) ensures that the sum power constraint is satisfied, but requires that each relay knows all elements of

[3] These expressions are valid for the sum power constraint. With a per-relay power constraint, weight normalization has to be performed element-wise.

the beamforming vector. Hence, a *stochastic* gradient algorithm with random perturbation vectors (as in [9]) cannot be applied to relay networks, since it would require to exchange all weights among the relays, thus imposing a tremendous signaling overhead. Rather, we propose to use a matrix $\mathbf{Q}$ of *deterministic* perturbation vectors (cf. [48]) known to each relay. This allows each relay to keep track of *all* beamforming weights and to perform vector normalization locally. Reasonable choices for the deterministic perturbation matrix are $\mathbf{Q} = [\mathbf{F}, j\mathbf{F}]$ (i.e., $N = 2R$ vectors) for P/M perturbation and $\mathbf{Q} = [\mathbf{F}, j\mathbf{F}, -\mathbf{F}, -j\mathbf{F}]$ ($N = 4R$) for T/R perturbation; here, $\mathbf{F}$ is an $R \times R$ unitary matrix, e.g., the discrete Fourier transform (DFT) matrix.

Note that with the per-relay power constraint, element-wise normalization does not require knowledge of all weights at each relay, thus allowing also for stochastic perturbations.

1.3.5 Channel, Power, and SNR Estimation

We next discuss the estimation of the receive signal power $P_{\mathcal{D}}(\boldsymbol{\alpha})$ in (1.4) and the SNR $\rho(\boldsymbol{\alpha})$ in (2.4) which are used as performance measures, as well as the estimation of the compound channel ξ in (2.3) required for coherent detection.

In the following, we omit the frame index k and denote the pilot sequence within a transmission frame as $s_{\mathrm{p}}[n]$, $n \in \mathcal{T}'$. The destination can then compute the maximum likelihood (ML) estimate of the compound channel as

$$\hat{\xi} = \frac{\sum_{n\in\mathcal{T}'} y[n] s_{\mathrm{p}}^*[n]}{\sum_{n\in\mathcal{T}'} |s_{\mathrm{p}}[n]|^2}. \tag{1.8}$$

Using (1.8), the ML estimates of receive signal power and SNR can be obtained as

$$\hat{P}_{\mathcal{D}} = |\hat{\xi}|^2, \qquad \hat{\rho} = \frac{|\hat{\xi}|^2}{\frac{1}{|\mathcal{T}'|}\sum_{n\in\mathcal{T}'} \left| y[n] - \hat{\xi} s_{\mathrm{p}}[n] \right|^2}. \tag{1.9}$$

For T/R perturbation, (1.8) and (1.9) are evaluated using $\mathcal{T}' = \mathcal{T}_{\mathrm{p}}$. After each weight update the destination stores the channel estimate and uses it for detection of the subsequent frames till the next update occurs. With P/M, (1.8) and (1.9) are calculated twice in each frame with $\mathcal{T}' = \mathcal{T}_{\mathrm{p}}^+$ and $\mathcal{T}' = \mathcal{T}_{\mathrm{p}}^-$. The channel estimate corresponding to the better beamforming vector is then kept for data detection in the next frame. Alternatively, an approximate ML estimate for the channel coefficient ξ can be obtained by evaluating (1.8) over the whole training interval ($\mathcal{T}' = \mathcal{T}_{\mathrm{p}}$) within the same frame (cf. [9]), provided that the step-size μ is chosen sufficiently small and $|\mathcal{T}_{\mathrm{p}}^+| = |\mathcal{T}_{\mathrm{p}}^-|$.

1.3.6 Birth and Death of Relays

Our deterministic perturbation approach is scalable in that it can be easily adapted to deal with the situation where relays enter ("birth") or leave ("death") the network, even in the case of a sum power constraint. We assume that the maximum number of relays is $R_{\max}$, of which $R \leq R_{\max}$ are active and can exchange information with $\mathcal{D}$ but not with each other. In essence, the destination and each relay keep track of all the active relays. The relays can then compute the required vector norm locally. Additionally, all relays know their "identity" (index i), which is fixed and enables them to pick their corresponding beamforming weight. If a relay $\mathcal{R}_{i_0}$ drops out, it informs $\mathcal{D}$ which in turn broadcasts the relay index i_0 to the remaining relays using $\log_2(R_{\max})$ bits. These relays then exclude the corresponding beamforming/perturbation weight from the update process. If a new relay enters the system, it contacts $\mathcal{D}$ which in turn broadcasts $R_{\max}$ bits to indicate to all relays (also to the new one) which relays are active. Since the new relay cannot know the current weights of the other relays, the weight adaptation process needs to be re-initialized in this case.

In the case of a per-relay power constraint, element-wise weight normalization allows that the relays only need to track their own weights. This renders a birth-and-death protocol particularly easy, since relays can enter or leave the system completely arbitrarily without informing the other relays.

1.4 Comparison with Optimal Beamforming

We next compare optimal batch beamforming designs with adaptive PB-BF. The former requires each relay having either local CSI (i.e., each relay's own back- and forward channel) or global CSI (i.e., all channels) available, whereas PB-BF exploits limited feedback to avoid CSI at the relays.

1.4.1 Optimal Batch Designs

Equal Gain Combining (EGC). Maximizing $P_{\mathcal{D}}(\boldsymbol{\alpha})$ or $\rho(\boldsymbol{\alpha})$ under a per-relay power constraint yields the beamforming weights $\alpha_i = \bar{h}_i/|\bar{h}_i| = h_i g_i/|h_i g_i|$ that amount to coherent combining [15]. This scheme requires that each relay knows the phase of its backward and forward channel.

Power Maximization under Sum Power Constraint (P-SP). Optimizing $P_{\mathcal{D}}(\boldsymbol{\alpha})$ in (1.4) under a sum power constraint amounts to maximizing $|\boldsymbol{\alpha}^H \bar{\mathbf{h}}|$ subject to $\|\boldsymbol{\alpha}\|^2 = 1$. Via the Cauchy-Schwarz inequality, the solution is obtained as $\alpha_i = \bar{h}_i/\|\bar{\mathbf{h}}\|$ requiring global CSI at $\mathcal{R}_i$. Alternatively, if global

CSI is available at $\mathcal{D}$, each relay needs only local CSI and feedback of $\|\bar{\mathbf{h}}\|$ from $\mathcal{D}$. This shows that the relays optimally allocate their transmit power to match the current local fading coefficients while performing coherent combining.

SNR Maximization under Sum Power Constraint (S-SP). The beamforming vector $\boldsymbol{\alpha}$ can also be chosen to maximize the SNR $\rho(\boldsymbol{\alpha})$ in (2.4) [37][29]. Under the sum power constraint $\|\boldsymbol{\alpha}\|^2 = 1$ this can be shown to lead to a generalized eigenvalue problem whose solution is given by [29]

$$\boldsymbol{\alpha} = \frac{(\mathbf{I}_R + \bar{\mathbf{G}}\bar{\mathbf{G}}^H)^{-1}\bar{\mathbf{h}}}{\|(\mathbf{I}_R + \bar{\mathbf{G}}\bar{\mathbf{G}}^H)^{-1}\bar{\mathbf{h}}\|}. \tag{1.10}$$

Again this essentially requires either global CSI at the relays or local CSI with feedback of $\|(\mathbf{I}_R + \bar{\mathbf{G}}\bar{\mathbf{G}}^H)^{-1}\bar{\mathbf{h}}\|$ from $\mathcal{D}$. In contrast to P-SP, (1.10) also accounts for noise amplification.

1.4.2 Comparison with PB-BF Schemes

It can be shown that $\rho(\boldsymbol{\alpha})$ and $P_{\mathcal{D}}(\boldsymbol{\alpha})$ have only one global maximum (unique up to phase ambiguity) under both power constraints and this maximum is achieved by the corresponding optimal batch design. The proposed PB-BF schemes aim to maximize $\rho(\boldsymbol{\alpha})$ or $P_{\mathcal{D}}(\boldsymbol{\alpha})$, and indeed approach their optimal counterparts (cf. Section 1.5). EGC can be approximated by PB-BF using element-wise normalization and the objective function $\gamma(\boldsymbol{\alpha})$ chosen as received signal power (cf. (1.9)). P-SP and S-SP performance can be approached using $\gamma(\boldsymbol{\alpha}) = P_{\mathcal{D}}(\boldsymbol{\alpha})$ and $\gamma(\boldsymbol{\alpha}) = \rho(\boldsymbol{\alpha})$ as objective function, respectively, and vector normalization of the beamforming weights. The PB-BF schemes can be implemented via T/R or P/M perturbation.

1.5 Simulation Results

We next investigate a network with $R = 3$ relays via numerical simulations; we will refer to the PB-BF schemes by adding the prefix 'PB-' to the corresponding batch design. For a fair comparison, all schemes use the same total relay power $\bar{P}$ (in (2.2) we thus have $P = \bar{P}/R$ under a per-relay power constraint and $P = \bar{P}$ under a sum power constraint). The source $\mathcal{S}$ transmits BPSK symbols with transmit power $P_s = \bar{P}$ and the destination $\mathcal{D}$ employs an ML detector. We further assume error- and delay-free 1-bit feedback, and employ a deterministic perturbation set based on a 3×3 DFT matrix.

1.5.1 Idealized Scenario

In this scenario, all channels are static i.i.d. Rayleigh fading including path loss effects, i.e., $h_i, g_i \sim \mathcal{CN}(0, d_i^{-2})$ with $d_i = 1, 3, 5$. Each relay perfectly knows its backward channel (used in (2.2)), and $\mathcal{D}$ has perfect knowledge of the compound channel ξ and the performance measures $P_{\mathcal{D}}(\boldsymbol{\alpha})$ and $\rho(\boldsymbol{\alpha})$. Unless stated otherwise, $\bar{P}/N_0 = 18\,\text{dB}$.

Convergence Behavior. For the case of PB-S-SP using P/M and T/R perturbation with step size $\mu = 0.1$ and $\mu = 0.3$, Fig. 1.2(a) shows the evolution of the receive SNR $\rho_k = \rho(\boldsymbol{\alpha}_k)$ (normalized by the maximum receive SNR) versus the frame index k for one channel realization. It is seen that with T/R ρ_k is nondecreasing and reaches almost optimal performance; a larger step size results in faster convergence but also in a larger gap to the optimum. Similar observations apply to P/M, which converges significantly faster than T/R, but features continual fluctuations whose amplitude increases with the step size.

For a systematic assessment of the convergence rate of PB-S-SP (with $\mu = 0.1$), Fig. 1.2(b) shows the empirical cumulative distribution function (cdf) of the normalized SNR gap that remains after a certain number of frames (shown as curve labels). The cdfs were obtained with 10^5 fading realizations. P/M converges considerably faster than T/R. To achieve an SNR gap of less than 4.3% in 91% of the cases, P/M and T/R respectively require 40 and 70 iterations. However, after a large number of frames, T/R on average features a considerably smaller SNR gap than P/M. Our simulations also revealed that a larger number of relays entails slower convergence; for space reasons, the corresponding curves can not be shown here.

BER Performance. Fig. 1.2(c) plots bit-error rate (BER) versus nominal SNR $\bar{P}/N_0$ (in dB) for the batch designed beamforming schemes EGC, P-SP, S-SP, and their perturbation-based counterparts. In each simulation run, the first 60 frame were not taken into account for the BER evaluation to ensure that the PB-BF schemes have converged (again $\mu = 0.1$). As a reference, we include an AF scheme that uses uniform PA and no coherent combining (labeled 'no BF').

It can be seen that all PB-BF performance curves are almost indistinguishable from those of their corresponding batch designs and offer significant gains over the no-BF case (e.g., 8 dB SNR improvement at a BER of 10^{-2}). SNR optimization (PB-S-SP) is seen to outperform power optimization (PB-P-SP) at high SNR. In fact, PB-S-SP and S-SP are the only schemes to achieve a diversity larger than 1. Power optimization under a sum power constraint (PB-P-SP) and under a per-relay power constraint (PB-EGC)

perform almost identically; in fact, PB-P-SP appears to suffer from noise amplification at high SNR.

1.5.2 Realistic Scenario

We next use independent, time-varying flat fading channels with Jakes Doppler profile and the same path loss model as above. Furthermore, the destination uses (1.8) and (1.9) to estimate the compound channel, the received signal power, and the instantaneous SNR. To this end, each transmission frame contains $|\mathcal{T}_\mathrm{p}| = 10$ pilot symbols in addition to $|\mathcal{T}_\mathrm{d}| = 40$ data symbols. The normalization in (2.2) is achieved by measuring the receive power at the relays during one frame.

We analyze the tracking capabilities of P/M perturbation (with $\mu = 0.1$ and $\mu = 0.5$) in terms of BER versus normalized Doppler frequency (i.e., Doppler in Hertz times frame length in seconds) for $\bar{P}/N_0 = 22\,\mathrm{dB}$ (see Fig. 1.3). In general, the BER degrades with increasing Doppler. At high Doppler frequencies the relay weights cannot be adapted fast enough to the channel variations (note that in practice, the feedback delay will add on top of this). Moreover, if there are channel variations within a frame, the compound channel and the objective function cannot be estimated accurately. Even at low Doppler, there is an order of magnitude BER penalty for PB-S-SP (cf. Fig. 1.3 with $\mu = 0.1$ and Fig. 1.2(c) at $\bar{P}/N_0 = 22\,\mathrm{dB}$).

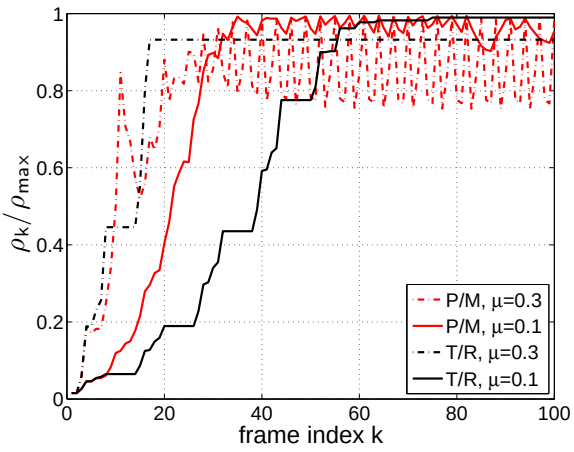
We observe that at at low Doppler frequencies a small step size ($\mu = 0.1$) performs better whereas at higher Doppler frequencies a larger step size ($\mu = 0.5$) is advantageous since it allows quicker adjustment of the relay weights. Note that with $\mu = 0.1$, PB-S-SP looses its entire performance advantage over PB-P-SP at high Doppler frequencies.

In time-varying scenarios, T/R suffers from the fact that the beamforming weights are not updated when the channel quality gets worse. This can be circumvented by building a forgetting factor into the performance measure γ_k.

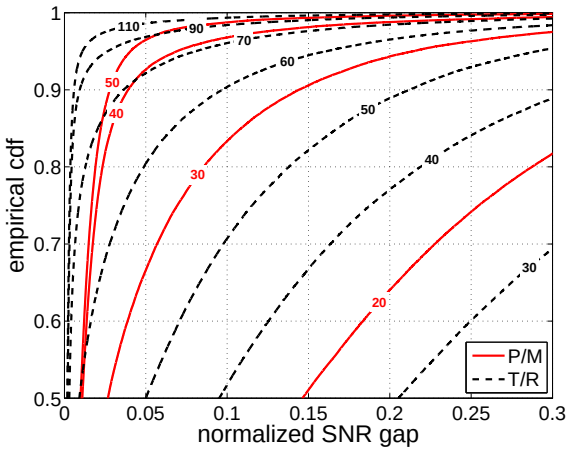
1.6 Conclusion

We have investigated scalable perturbation-based distributed beamforming protocols in wireless relay networks that exploit 1-bit feedback to approach the optimal beamforming weights in an adaptive manner while avoiding CSI at the relay nodes. We used a deterministic perturbation set to optimize either received signal power or SNR at the destination under per-relay or sum power constraints. The best performance was observed with SNR as

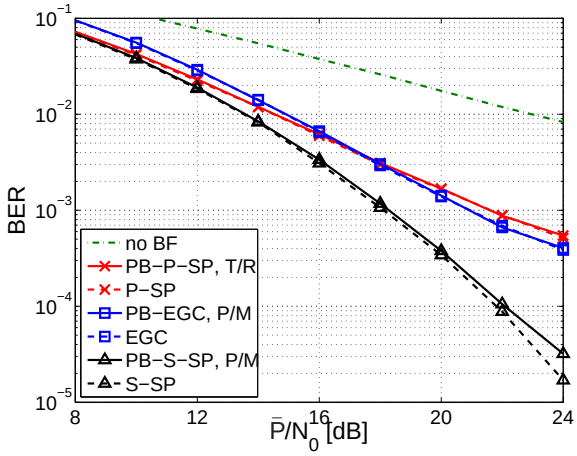
objective function under a sum power constraint. At high SNR, equal gain combining appears to be preferable over power optimization under a sum-power constraint. In time-varying environments, the proposed perturbation scheme require a careful choice of the step-size parameter and the transmission frame length.



(a)



(b)



(c)

Figure 1.2: PB-BF performance for idealized scenario with static channels in a 3-relay network: (a) example for SNR evolution at $\mathcal{D}$ ($\bar{P}/N_0 = 18\,\mathrm{dB}$), (b) cdf of the SNR gap after a fixed number of frames ($\bar{P}/N_0 = 18\,\mathrm{dB}$, $\mu = 0.1$), and (c) BER versus nominal SNR ($\mu = 0.1$).

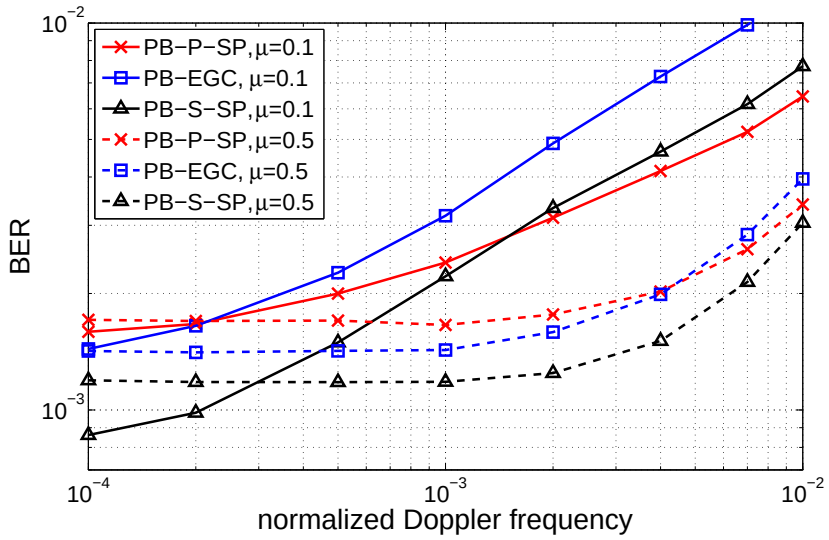


Figure 1.3: BER versus normalized Doppler frequency for various PB-BF schemes using P/M perturbation under a realistic scenario ($\bar{P}/N_0 = 22\,\mathrm{dB}$).

Chapter 2

A Multiplicative Weight Perturbation Scheme for Distributed Beamforming in Wireless Relay Networks with 1-bit Feedback

P. Fertl, A. Hottinen and G. Matz

This Chapter uses a perturbation-based distributed beamforming with 1-bit feedback for wireless amplify-and-forward relay networks. We propose to use multiplicative perturbations based on Givens rotations to adapt the beamforming weights while guaranteeing a sum power constraint for the relays. This perturbation scheme is shown to be computationally efficient and easy to design, thus allowing for low-complexity relay nodes. An adaptation of the Givens rotation angle allows to approach optimum performance arbitrarily close. Numerical simulations demonstrate noticeable performance gains over additive perturbation schemes that have been exclusively considered up to now.

2.1 Introduction

Motivation. Distributed beamforming with half-duplex amplify-and-forward (AF) relays has recently attracted much attention due to its ability to exploit spatial diversity in a distributed fashion [37][29][34]. However, this approach

imposes stringent requirements on the availability of channel state information (CSI) at the relay nodes; either global CSI (i.e., all channels) or local CSI (i.e., each relay's own backward and forward channel) is required. The requirement for CSI at the relays can be avoided by using feedback from the destination to the relays in order to adaptively adjust the beamforming weights. Perturbation-based beamforming (PB-BF) is a well known example for such an approach, applicable to centralized arrays with co-located antennas [9][48]. In [18], we proposed deterministic, additive vector perturbations and 1-bit feedback to extend PB-BF to relay networks. This scheme has the potential to approach optimum performance without any CSI at the relays. However, to satisfy a sum power constraint each update of the beamforming weights involves a vector normalization at each relay.

Contributions and Organization. In this Chapter, we propose multiplicative perturbations in terms of elementary Givens rotations [23] for PB-BF in wireless relay networks. The relays receive 1-bit feedback from the destination to adapt their beamforming weights with the goal of maximizing the signal-to-noise ratio (SNR) at the destination. Our multiplicative perturbation scheme has the advantage that it inherently maintains constant sum power and has a computational complexity at each relay which is independent of the number of relays in the network. This is in striking contrast to additive vector perturbation techniques (cf. [18]). We further show how the adaptation behavior of our proposed scheme can be controlled in an intuitive manner and provide numerical results that illustrate the performance of our method in comparison to additive PB-BF.

The rest of the paper is organized as follows. Section 2.2 provides the background for distributed beamforming in relay networks. The proposed multiplicative perturbation scheme is introduced and discussed in detail in Section 2.3. Simulation results are shown in Section 2.4. Finally, conclusions are given in Section 2.5.

2.2 Network Beamforming

System Model. We consider a half-duplex wireless relay network with single antenna nodes where a single source $\mathcal{S}$ communicates with a single destination $\mathcal{D}$ via R amplify-and-forward relays $\mathcal{R}_i$, $i = 1, \ldots, R$ (cf. Fig. 2.1). There is no direct link between $\mathcal{S}$ and $\mathcal{D}$, and we assume perfect synchronization among the nodes. The half-duplex constraint necessitates transmission in two hops. In the first hop, $\mathcal{S}$ transmits the signal $\sqrt{P_s}\, s$ to the relays which receive

$$x_i = \sqrt{P_s}\, h_i s + w_i. \tag{2.1}$$

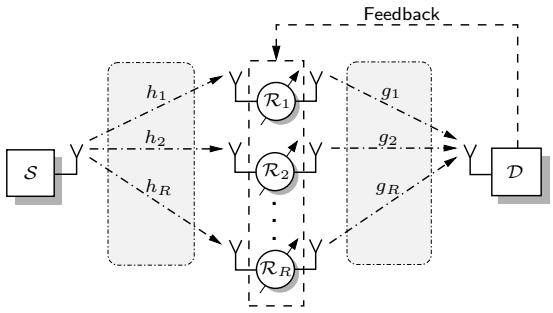


Figure 2.1: Wireless relay network with feedback.

Here, s is the transmit symbol normalized as $\mathbb{E}\{|s|^2\} = 1$ ($\mathbb{E}\{\cdot\}$ denotes expectation), P_s denotes the average transmit power of the source $\mathcal{S}$, h_i is the complex fading coefficient of the "backward" channel[1], and $w_i \sim \mathcal{CN}(0, N_0)$ denotes i.i.d. circularly complex Gaussian noise. In the second hop, each relay applies a complex beamforming weight α_i to the signal it has received and forwards

$$r_i = \lambda_i \, \alpha_i^* \, x_i \,, \tag{2.2}$$

to the destination $\mathcal{D}$; here, $P_{x_i} = \mathbb{E}\{|x_i|^2|h_i\}$ is the average receive power at relay $\mathcal{R}_i$ and complex conjugation (superscript *) will simplify notation later on. Throughout this paper, we assume a sum power constraint $\sum_{i=1}^{R} \mathbb{E}\Big\{|r_i|^2|h_i\Big\} = P$. With the power normalization in (2.2), this requires the beamforming vector $\boldsymbol{\alpha} = (\alpha_1 \ldots \alpha_R)^T$ to have unit Euclidean norm, $\|\boldsymbol{\alpha}\| = 1$. The destination receives $y = \sum_{i=1}^{R} g_i r_i + v$, where g_i denotes the complex fading coefficient of the "forward" channel between $\mathcal{R}_i$ and $\mathcal{D}$ and $v \sim \mathcal{CN}(0, N_0)$ is circularly complex Gaussian noise. Combining this with (2.1) and (2.2) yields the compound channel model[2]

$$y = \xi s + \eta, \quad \text{with } \xi \triangleq \boldsymbol{\alpha}^H \bar{\mathbf{h}}, \quad \eta \triangleq \boldsymbol{\alpha}^H \bar{\mathbf{G}} \mathbf{w} + v. \tag{2.3}$$

In this expression, $\bar{\mathbf{h}} \triangleq (\bar{h}_1 \ldots \bar{h}_R)^T$ with $\bar{h}_i \triangleq h_i g_i \sqrt{P_s P / P_{x_i}}$, $\bar{\mathbf{G}} \triangleq \mathrm{diag}(\bar{g}_1, \ldots, \bar{g}_R)$ with $\bar{g}_i \triangleq g_i \sqrt{P / P_{x_i}}$, and $\mathbf{w} \triangleq (w_1 \ldots w_R)^T$. Note that the beamforming vector $\boldsymbol{\alpha}$ not only affects the effective channel gain ξ in (2.3) but also the noise η.

The receive SNR at the destination $\mathcal{D}$ can be obtained from (2.3) as ($\mathbf{I}$ is the identity matrix)

$$\rho(\boldsymbol{\alpha}) \triangleq \frac{\mathbb{E}\Big\{|\xi s|^2 \Big| \bar{\mathbf{h}}\Big\}}{\mathbb{E}\Big\{|\eta|^2 \Big| \bar{\mathbf{G}}\Big\}} = \frac{1}{N_0} \frac{|\boldsymbol{\alpha}^H \bar{\mathbf{h}}|^2}{\|(\mathbf{I} + \bar{\mathbf{G}} \bar{\mathbf{G}}^H)^{1/2} \boldsymbol{\alpha}\|^2}. \tag{2.4}$$

[1] Note that we do not assume a specific channel statistics.

[2] Superscript ^{T} (^{H}) denotes (Hermitian) transposition; $\mathrm{diag}(x_1, \ldots, x_m)$ is the $m \times m$ diagonal matrix with diagonal elements $x_1, \ldots, x_m$.

Note that the SNR is invariant to a common phase factor, i.e., $\rho(\boldsymbol{\alpha}) = \rho(e^{j\psi}\boldsymbol{\alpha})$. Up to this phase ambiguity, $\rho(\boldsymbol{\alpha})$ can be shown to have a unique global maximum $\rho_{\max}$ (and no local maxima). The beamforming weights $\boldsymbol{\alpha}$ can be designed to maximize the received SNR $\rho(\boldsymbol{\alpha})$ in (2.4) subject to the sum power constraint $\|\boldsymbol{\alpha}\|^2 = 1$. This leads to the optimum beamforming vector [37][29]

$$\boldsymbol{\alpha}_{\text{opt}} = \frac{(\mathbf{I} + \bar{\mathbf{G}}\bar{\mathbf{G}}^H)^{-1}\bar{\mathbf{h}}}{\|(\mathbf{I} + \bar{\mathbf{G}}\bar{\mathbf{G}}^H)^{-1}\bar{\mathbf{h}}\|}\,, \tag{2.5}$$

which is unique up to a phase factor. However, calculating the optimum beamforming weights can be shown to require either global CSI at all relays or local CSI for each relay and global CSI at $\mathcal{D}$ with feedback of a scalar normalization factor to the relays.

Perturbation-based Beamforming. Motivated by beamforming techniques for co-located arrays [9][48], we proposed distributed BF for relay networks based on *additive* weight perturbations in [18]. These techniques circumvent the need for CSI at the relays and allow to approach the maximum of the objective function[3] $\gamma(\boldsymbol{\alpha})$.

We next briefly review the transmission principle, termed *take/reject (T/R) perturbation* (see [18] for more details). The idea is to approximate the optimum beamforming vector $\boldsymbol{\alpha}_{\text{opt}}$ by iteratively updating the beamforming weights at the relays according to 1-bit feedback provided by the destination. To this end, the source transmits frames that consist of two training blocks $\tilde{\mathcal{B}}_k^{(\text{p})}$ and $\mathcal{B}_k^{(\text{p})}$, and a data block $\mathcal{T}_{\text{d}}$ (k is the frame index). The relays forward these frames according to (2.2), using the currently best beamforming vector (denoted $\boldsymbol{\alpha}_k$) for $\mathcal{B}_k^{(\text{p})}$ and $\mathcal{T}_{\text{d}}$, while using a perturbed version $\tilde{\boldsymbol{\alpha}}_k$ for the training portion $\tilde{\mathcal{B}}_k^{(\text{p})}$. Up to now, only the following additive perturbations have been considered (cf. [18]):

$$\tilde{\boldsymbol{\alpha}}_k = \frac{\boldsymbol{\alpha}_k + \mu\,\mathbf{q}_{k \bmod \tilde{N}}}{\|\boldsymbol{\alpha}_k + \mu\,\mathbf{q}_{k \bmod \tilde{N}}\|}. \tag{2.6}$$

Here, μ is a step-size parameter and $\mathbf{q}_n$ denotes the additive perturbation vector taken cyclically from a deterministic $R \times \tilde{N}$ matrix $(\mathbf{q}_0 \ldots \mathbf{q}_{\tilde{N}-1})$ with $\tilde{N} \geq 4R$ [18]. Note that (2.6) involves a normalization which is necessary to satisfy the sum power constraint and requires that all beamforming weights are tracked at each relay. This implies that the computation of (2.6) in general needs $10R$ real flops and one square root operation per relay.

[3] Note that the receive power $P_{\mathcal{D}}(\boldsymbol{\alpha}) \triangleq \mathbb{E}\{|\xi s|^2 | \bar{\mathbf{h}}\} = |\boldsymbol{\alpha}^H \bar{\mathbf{h}}|^2$ can as well be used as objective function (cf. [18]).

The destination evaluates the effectiveness of the beamforming weights $\boldsymbol{\alpha}_k$ and $\tilde{\boldsymbol{\alpha}}_k$ with regard to the objective function $\gamma(\boldsymbol{\alpha})$ within $\mathcal{B}_k^{(\mathrm{p})}$ and $\tilde{\mathcal{B}}_k^{(\mathrm{p})}$, respectively. It then provides a single bit c_k of feedback to the relays, indicating which weights perform better, i.e., $c_k = 0$ if $\gamma(\boldsymbol{\alpha}_k) \geq \gamma(\tilde{\boldsymbol{\alpha}}_k)$ and $c_k = 1$ if $\gamma(\boldsymbol{\alpha}_k) < \gamma(\tilde{\boldsymbol{\alpha}}_k)$. The relays update the beamforming vector $\boldsymbol{\alpha}_{k+1}$ to be used in the next frame according to $\boldsymbol{\alpha}_{k+1} = \boldsymbol{\alpha}_k$ ("reject" $\tilde{\boldsymbol{\alpha}}_k$) if $c_k = 0$ and $\boldsymbol{\alpha}_{k+1} = \tilde{\boldsymbol{\alpha}}_k$ ("take" $\tilde{\boldsymbol{\alpha}}_k$) if $c_k = 1$. The new vector $\boldsymbol{\alpha}_{k+1}$ will then be the basis for the next perturbation and the whole process continues in an iterative manner. The first frame is initialized by setting $\gamma_0 = 0$ and $\tilde{\boldsymbol{\alpha}}_0 = \boldsymbol{\alpha}_0$, where $\boldsymbol{\alpha}_0$ can be an arbitrary vector with $\|\boldsymbol{\alpha}_0\|^2 = 1$.

2.3 Proposed Perturbation Scheme

Beamforming Manifold. This section proposes to replace (2.6) with a multiplicative weight perturbation scheme. This is motivated by the fact that the sum power constraint $\|\boldsymbol{\alpha}\|^2 = 1$ implies that the real-valued representation $(\mathrm{Re}\{\boldsymbol{\alpha}^T\}\ \mathrm{Im}\{\boldsymbol{\alpha}^T\})^T$ of admissible beamforming vectors lies on a $(2R-1)$-dimensional hypersphere in the $2R$-dimensional (real) Euclidean space. In addition, the phase invariance of our objective function means that there are disjoint one-dimensional equivalence classes of beamforming vectors within which the SNR $\gamma(\boldsymbol{\alpha})$ remains constant. It is sufficient to consider only one representative of each equivalence class, which reduces the number of degrees of freedom by one. Without loss of generality, we choose this representative to be $\boldsymbol{\alpha}' = e^{-j\arg(\alpha_R)}\boldsymbol{\alpha}$ such that $\mathrm{Re}\{\alpha'_R\} = |\alpha_R| \geq 0$ and $\mathrm{Im}\{\alpha'_R\} = 0$. In the following, we thus restrict to the beamforming vectors

$$\mathbf{a} = \Big(\mathrm{Re}\{\boldsymbol{\alpha}^T\}\ \mathrm{Im}\{\alpha_1\}\ \ldots\ \mathrm{Im}\{\alpha_{R-1}\} \Big)^T$$

which have length $\bar{R} = 2R-1$ but, due to the constraint $\|\mathbf{a}\|^2 = 1$, lie on a $(2R-2)$-dimensional hyper-hemisphere $\mathcal{H}$. Note that

$$\boldsymbol{\alpha} = \Big(a_1 \ldots\ a_R\Big)^T + j\Big(a_{R+1} \ldots\ a_{2R-1}\ 0\Big)^T.$$

Rewriting the cost function $\gamma(\boldsymbol{\alpha})$ in terms of $\mathbf{a}$ reveals that it has a unique global maximum on $\mathcal{H}$, i.e., without phase ambiguity.

As compared to the Euclidean perspective underlying (2.6), we have reduced the problem dimension by two. Furthermore, (2.6) is intended to approximate Euclidean-space steepest ascent (gradient) techniques but does not account for the manifold structure of the hypersphere (this necessitates the renormalization). Specifically, the natural notion of a translation on the

hypersphere is *rotation*, amounting to *multiplication* by a matrix $\mathbf{F}$ belonging to the *special orthogonal group* $\mathrm{SO}(\bar{R})$ [25], defined by $\mathbf{F}^T\mathbf{F}=\mathbf{I}$ and $\det(\mathbf{F})=1$.

Multiplicative Perturbation. In the light of the foregoing discussion we propose to replace the additive perturbation (2.6) with

$$\tilde{\boldsymbol{\alpha}}'_k = \mathbf{F}_{k \bmod N}\, \mathbf{a}_k\,, \tag{2.7}$$

where $\mathbf{F}_n \in \mathrm{SO}(\bar{R})$ denotes orthogonal matrices cyclically taken from an appropriately chosen set $\mathcal{I} \triangleq \{\mathbf{F}_0, \ldots, \mathbf{F}_{N-1}\}$ of size $|\mathcal{I}| = N$. The set $\mathcal{I}$ is known to all relays so that each relay can keep track of all beamforming weights. Note that by construction $\|\tilde{\boldsymbol{\alpha}}'_k\| = \|\mathbf{a}_k\| = 1$. Particularly simple and useful examples for orthogonal matrices are *Givens rotations* [23] by an angle $\phi \in (-\pi, \pi]$ within the (l, m)-plane (with $l \neq m$):

$$\boldsymbol{\Gamma}_{lm}(\phi) = (\mathbf{e}_l\, \mathbf{e}_m) \begin{pmatrix} c & s \\ -s & c \end{pmatrix} (\mathbf{e}_l\, \mathbf{e}_m)^T + \Big[\mathbf{I} - (\mathbf{e}_l\, \mathbf{e}_m)(\mathbf{e}_l\, \mathbf{e}_m)^T\Big], \tag{2.8}$$

where $c = \cos(\phi)$, $s = \sin(\phi)$, and $\mathbf{e}_l$ denotes the lth canonical unit vector. Applying $\boldsymbol{\Gamma}_{lm}(\phi)$ to a vector performs a clockwise rotation of the lth and mth element by the angle ϕ (first term in (2.8)), while all other elements remain unaffected (last term in (2.8)).

For any given initial vector $\mathbf{a}_0$, there is an orthogonal matrix $\mathbf{F}'$ that rotates $\mathbf{a}_0$ into the optimum beamforming vector $\mathbf{a}_{\mathrm{opt}}$ in one step. This matrix can be factored into $\bar{R}-1$ Givens rotations [23] as

$$\mathbf{F}' = \prod_{l=1}^{\bar{R}-1} \boldsymbol{\Gamma}_{l,l+1}(\phi_l), \tag{2.9}$$

with properly chosen angles ϕ_l. In our distributed beamforming setup, the angles ϕ_l and hence $\mathbf{F}'$ is not available. Nonetheless, it appears promising to perform the multiplicative perturbation (2.7) using a set $\mathcal{I}$ consisting of appropriately chosen Givens rotations.

Givens Perturbations. We next show that perturbations based on Givens rotations have the advantage of being intuitive, computationally efficient, and simple to design.

Intuition. As argued previously, rotations are the natural translation on the hyper-(hemi)sphere $\mathcal{H}$ and thus more intuitive than additive perturbations. Specifically, additive perturbations can have arbitrary orientation relative to $\mathcal{H}$, thereby hindering an interpretation of the parameter μ in (2.6) as step size. In the extreme case where $\mathbf{q}_{k \bmod \tilde{N}} = \boldsymbol{\alpha}_k$, the perturbation is orthogonal to $\mathcal{H}$ at $\boldsymbol{\alpha}_k$, resulting in $\tilde{\boldsymbol{\alpha}}_k = \boldsymbol{\alpha}_k$, i.e., no perturbation at all. In contrast, the angle of the Givens rotation gives a clear indication of

the amount of perturbation on $\mathcal{H}$. Fig. 2.2 illustrates this behavior in two dimensions. Starting from the initial vector $\mathbf{a}_0$, the Givens perturbations continually rotate the beamforming vector (marked with bullets) closer to the optimum weights $\mathbf{a}_{\text{opt}}$ while retaining the sum power constraint. In contrast, additive perturbations (marked with crosses) suffer from strongly varying step sizes (e.g., in the first perturbation) and require normalization.

Design. The action of Givens rotations is geometrically intuitive and simplifies the design of the set $\mathcal{I}$. In particular, any Givens rotation $\mathbf{\Gamma}_{lm}(\phi)$ is completely specified in terms of the index pair (l, m) and the angle ϕ. Thus, instead of specifying the set $\mathcal{I}$ in terms of N orthogonal matrices of dimension $\bar{R}\times\bar{R}$, it is sufficient to specify the corresponding N index pairs and angles. For the moment, we consider a fixed choice of the rotation angle. Then, there are $\bar{R}(\bar{R}-1)/2$ different index pairs and corresponding rotation planes in total. However, following (2.9), the minimum number of rotation planes is given by $\bar{R}-1$ (in this case, each index has to occur at least once in the list). For our T/R scheme, we have to allow for clockwise and counter-clockwise rotations within each rotation plane, yielding a set $\mathcal{I}$ of maximum size $N=\bar{R}(\bar{R}-1)$ and minimum size $N=2(\bar{R}-1)$. Note that counter-clockwise rotations can be achieved by swapping indices, i.e., $\mathbf{\Gamma}_{ml}(\phi)=\mathbf{\Gamma}_{lm}(-\phi)$. Choosing a large perturbation set increases the chance of picking a rotation plane that allows a perturbation within the direction of the steepest gradient; however, it potentially requires more trials until the "right" rotation plane is getting used. With small N, each rotation plane is tested more frequently but certain rotations not available within the perturbation set can only be approximated over several iterations.

Complexity. Any Givens rotation involves only two elements of $\mathbf{a}_k$, i.e., $\tilde{a}_{k,l}=c\,a_{k,l}-s\,a_{k,m}$, $\tilde{a}_{k,m}=s\,a_{k,l}+c\,a_{k,m}$, and $\tilde{a}_{k,i}=a_{k,i}$ for $i\neq l,m$. This means that the beamforming weights of at most two relays are updated within each iteration, with each update requiring only 6 flops per relay. In contrast to additive perturbation, the complexity per relay of our multiplicative perturbation scheme thus is *independent of the number of relays*. Furthermore, similarities between our perturbation scheme and CORDIC algorithms [62] can be exploited to reduce complexity even further via appropriate choice of the rotation angle.

Angle Adaptation. In the following, we present a modification of our Givens rotations based perturbation scheme which is motivated by the fact that the method will get stuck as soon as the angular distance between the current beamforming vector $\mathbf{a}_k$ and the optimum vector $\mathbf{a}_{\text{opt}}$ is less than the fixed angle ϕ. In this case none of the available rotations further improves the objective function. An obvious way to evade this deadlock is a reduction of

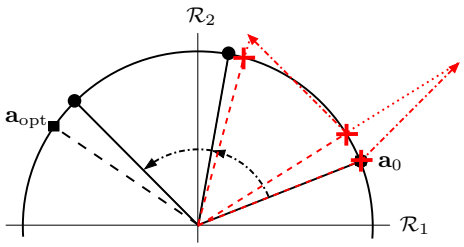


Figure 2.2: Example for perturbations in two dimensions under a unit-length constraint (bullets: multiplicative, crosses: additive).

the rotation angle. Specifically, we propose that all relays count the number of successive Givens perturbations that have been rejected by the destination because they did not improve $\gamma(\boldsymbol{\alpha})$; this essentially amounts to accumulating the number of successive feedback bits equal to zero. Whenever this number is larger than a certain integer $M \leq N$, all relays switch to a smaller angle (e.g., according to $\phi \leftarrow \gamma\,\phi$, $\gamma < 1$). It may be advantageous to shrink the rotation angle even before all available rotations have been rejected. Note that the perturbation index set $\mathcal{I}$ remains unchanged, however. It follows from the properties of the objective function that our Givens perturbation scheme with angle adaptation asymptotically achieves optimum performance. Yet, the convergence speed depends on the specific choice of initial angle, angle reduction, and M.

2.4 Simulation Results

We next study the performance of the multiplicative perturbation scheme with angle adaptation via numerical simulations and provide a comparison with the additive PB-BF scheme proposed in [18]. In our simulations, all channels were chosen static i.i.d. Rayleigh fading. The source $\mathcal{S}$ transmitted BPSK symbols with power $P_s = P$. The destination $\mathcal{D}$ had perfect knowledge of the compound channel ξ to perform ML detection. We further ensured exact evaluation of the objective funcion (in practice, the SNR is estimated at the destination using the training blocks, cf. [18]) and error-free 1-bit feedback. All results shown were obtained using 10^5 fading realizations. Unless stated otherwise, we chose an initial rotation angle of $\phi = 45°$, an angle reduction factor of $\gamma = 0.25$, and M equal to the size of the perturbation set for the multiplicative scheme. For the additive scheme, we used a constant step-size $\mu = 0.5$ and a perturbation set of size $\tilde{N} = 4R$ based on a discrete Fourier transform matrix (cf. [18]).

Convergence Behavior. We first study the convergence rate for a network

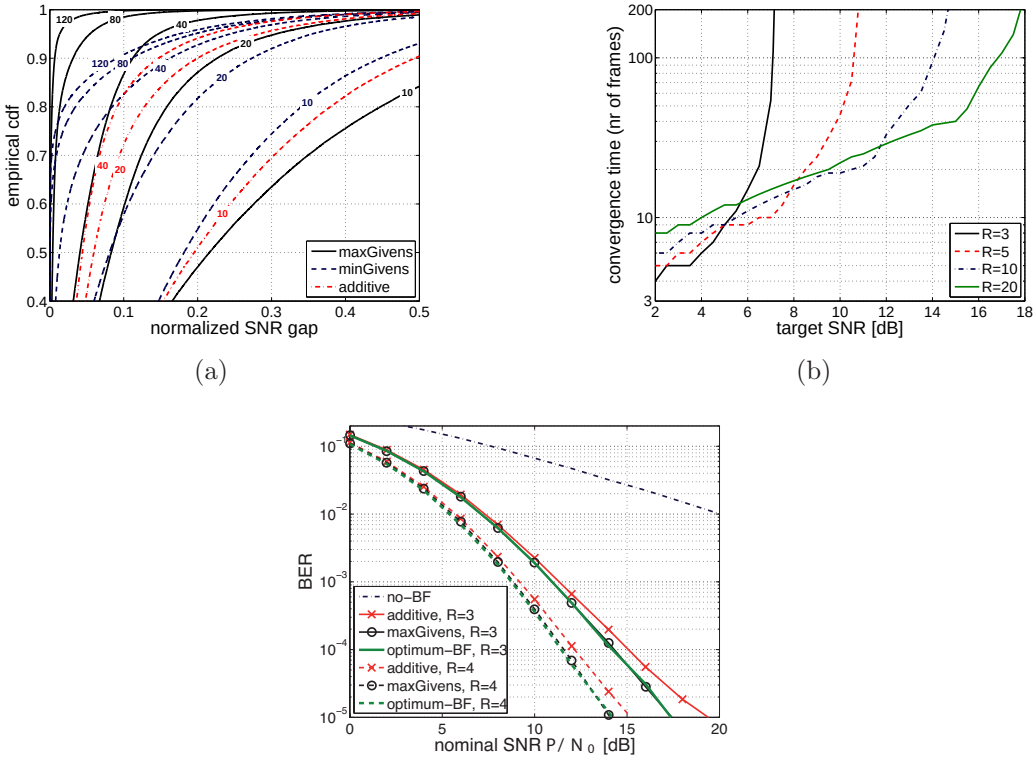


Figure 2.3: Illustration of PB-BF performance: (a) cdf of the normalized SNR gap after a fixed number of frames ($R=3$, $P/N_0=14\,\mathrm{dB}$), (b) convergence time (in frames) versus target SNR ($P/N_0=14\,\mathrm{dB}$) for different network sizes, and (c) BER versus nominal SNR ($R=3,4$).

of $R=3$ relays at nominal SNR $P/N_0=14\,\mathrm{dB}$. Fig. 2.3(a) shows the empirical cumulative distribution function (cdf) of the normalized SNR gap $\frac{\rho_{\max}-\gamma(\boldsymbol{\alpha}_k)}{\rho_{\max}}$ that remains after a fixed number of transmission frames (shown as curve labels). Results are shown for additive perturbation and for Givens perturbations with perturbation sets[4] of minimum size $N=2(\bar{R}-1)=8$ (labeled 'minGivens') and of maximum size $N=\bar{R}(\bar{R}-1)=20$ ('maxGivens'). We note that no noticeable improvements are observed for additive perturbations beyond 40 transmission frames (iterations). It is seen that initially (after 10 iterations) minGivens performs best; here, rapid improvement is achieved since only a few rotations have to be checked. However, if a small SNR gap has to be ensured with high probability, maxGivens is preferable. For

[4]Here, we chose the sets such that all the different rotations planes first undergo clockwise and then counter-clockwise rotations.

example, to achieve a normalized SNR gap less than 11% in 90% of the cases, minGivens and maxGivens require 40 and 80 iterations, respectively. After 120 frames, an SNR gap less than 5% is achieved in 99% of the cases with maxGivens but only in 85% of the cases with minGivens. With additive perturbations, initial convergence is poorer than with minGivens and the SNR gap after many iterations is larger than with maxGivens.

Network Size. Next, we analyze the impact of the network size (i.e., number of relays R) on the convergence of PB-BF with minGivens perturbations ($N = 2(\bar{R}-1)$) from the perspective of rapidly achieving a certain target SNR at the destination given that $P/N_0 = 14\,\text{dB}$ (note that the sum relay power is independent of the network size). Fig. 2.3(b) shows the convergence time versus target SNR. Here, we define convergence time as the minimum number of frames required to achieve the target SNR in 98% of the cases. Obviously, the achievable SNR is higher for larger R due to increasing array gain (i.e., about 7.5 dB for $R = 3$ and about 15 dB for $R = 10$). However, very large convergence times are implied if the target SNR approaches the SNR limit. It is further seen that the curves for different R intersect, and hence, for a given target SNR there is an optimum network size minimizing the convergence time. For example, 7 dB target SNR can most rapidly be obtained using $R=5$ relays whereas the optimum number of relays to achieve 10 dB is $R=10$.

BER Performance. Bit error rate (BER) versus nominal SNR P/N_0 for the case of $R = 3$ and $R = 4$ relays is shown in Fig. 2.3(c). Since close-to-optimum performance is desired, we here use maxGivens (20 and 42 rotation planes, respectively) and compare the results with additive perturbation. As ultimate benchmarks, we also include the results for optimum beamforming using the weights in (2.5) (labeled 'optimum-BF') and a scheme without beamforming ('no-BF'), i.e., uniform power allocation $\alpha_i = \sqrt{P/R}$ and no coherent combining. For each fading realization, we excluded the initial convergence phase (first 60 frames) from the BER evaluation. It can be seen that the angle adaptation allows maxGivens to closely approach optimum performance and to outperform additive PB-BF in the high-SNR regime (e.g., 0.8 dB SNR gain at a BER of 10^{-4} for $R=3$), even though the complexity of maxGivens is smaller than that of additive perturbations. Also, the curves show that our scheme is able to fully exploit the spatial diversity offered by cooperative relaying and offers significant gains over the no-BF case (e.g., 13 dB SNR improvement at a BER of 10^{-2} for $R=3$).

2.5 Conclusion

We have proposed to use multiplicative perturbations based on Givens rotations for distributed beamforming in wireless relay networks with 1-bit feedback. These perturbations are much better matched to the non-Euclidean manifold underlying the problem setup than additive perturbations considered previously. It further allows direct step-size control in terms of rotation angles and is computationally very efficient. In fact, the per-relay complexity is independent of the network size. Numerical simulations showed that our scheme approaches optimum performance arbitrarily close at a satisfactory convergence speed.

Chapter 3

Optimal user pairing for multiuser MIMO

E. Viterbo and A. Hottinen

In this chapter we show how the capacity of the uplink of a multiuser system can be increased by a scheduling strategy, which pairs the transmission of users in different time/frequency/code slots according to the channel quality. The optimal scheduling strategy is equivalent to a combinatorial optimization problem. We show how this problem can be solved efficiently by using the Hungarian method. We then show that, by using the proposed scheduling scheme, the performance of Minimum Mean Square Error detection approaches the one of Maximum Likelihood detection, as the number of users increases.

3.1 Introduction

A multiuser multiple-input-multiple output (MU-MIMO) system consists of K user with n_t antennas each communicating to a base station with n_r receive antennas. Since each user faces a different channel condition, in different time/frequency/code (TFC) slots it is possible to improve the overall system capacity by *multiuser scheduling*. This technique attempts to increase the system capacity by smartly allocating the channel to different subgroups of users. A general introduction to this topic can be found in [5]. Among the most popular multiuser scheduling schemes we have opportunistic scheduling and best subset selection. All scheduling schemes are confronted with the *fairness* issue that forces to sacrifice the overall network optimality, in order to guarantee to all users a minimum service requirement.

In this contribution we will focus on a scheduling scheme based on user pairing and assume as objective function the total instantaneous mutual information between users and the base station when both ML and MMSE receivers are considered. We first show that the combinatorial optimization problem, which yields an optimal scheduling, can be solved efficiently by using the Hungarian method [35][36][44]. We then show that, by using the proposed scheduling scheme, the performance of Minimum Mean Square Error (MMSE) detection approaches the one of Maximum Likelihood (ML) detection, as the number of users increases.

3.2 System Model

In this section, we describe the multiuser system model and we state the scheduling problem based on user pairing as a combinatorial optimization problem.

Considering the uplink channel, we assume that the users are multiplexed in the code domain, i.e., all user's signals overlap both in time and in frequency within a channel use. For K users we have

$$\mathbf{y} = \sum_{k=1}^{K} \mathbf{H}^{(k)}\mathbf{x}^{(k)} + \mathbf{z} \tag{3.1}$$

where $\mathbf{x}^{(k)} \in \mathbb{C}^{n_t}$ is the transmitted column vector from user k, $\mathbf{H}^{(k)} \in \mathbb{C}^{n_r \times n_t}$ the channel coefficient matrix, $\mathbf{z} \in \mathbb{C}^{n_r}$ the white Gaussian noise vector distributed as $\mathcal{N}_c(0, \mathbf{I}_{n_r})$. Let P be the total transmitted power by each user (i.e., $P = \mathbb{E}[\|\mathbf{x}^{(k)}\|^2]$), then we define $\mathsf{SNR} = P$.

We assume the transmitter does not know the channel (open loop) and the receiver has knowledge of each user channel matrix. Furthermore, we assume that a power control scheme is used to compensate the path-loss, so that the average received power from each user is balanced and equal to P.

Let us rewrite (3.1) in equivalent matrix form

$$\mathbf{y} = \left[\mathbf{H}^{(1)}|\cdots|\mathbf{H}^{(K)}\right] \begin{bmatrix} \mathbf{x}^{(1)} \\ \vdots \\ \mathbf{x}^{(K)} \end{bmatrix} + \mathbf{z} = \mathcal{H}\mathbf{X} + \mathbf{z} \tag{3.2}$$

where we assume that the joint channel $n_r \times Kn_t$ matrix $\mathcal{H}$ is constant during the channel use and $\mathbf{X}$ is the joint input vector of length Kn_t.

Assuming the receiver performs ML detection the *mutual information per user* (conditioned by the channel realization) for channel (3.2) is given by

$$I_{ML}(\mathbf{X}; \mathbf{y}|\mathcal{H}) = \frac{1}{K}\log_2\left(\det\left(\mathbf{I}_{n_r} + \frac{P}{Kn_t}\mathcal{H}\mathcal{H}^{\dagger}\right)\right) \tag{3.3}$$

Due to the high complexity of ML detection, the simpler MMSE receiver is generally adopted, and in this case we have

$$I_{MMSE}(\mathbf{X};\mathbf{y}|\mathcal{H}) = \frac{1}{K}\sum_{j=1}^{Kn_t}\log_2\left(1+\mathbf{h}_j^{\dagger}\mathbf{A}_j^{-1}\mathbf{h}_j\right) \tag{3.4}$$

where $\mathbf{h}_j$ are the column vectors of $\mathcal{H}$ and

$$\mathbf{A}_j = \frac{Kn_t}{P}\mathbf{I}_{n_r} + \sum_{i=1,i\neq j}^{Kn_t}\mathbf{h}_i^{\dagger}\mathbf{h}_i \ .$$

The above expressions represent a measure of the per-user throughput, given that the system is occupying a total bandwidth B.

The above scheme requires a K user multiuser detection which can be still rather complex for large numbers of users and transmit antennas. For this reason it is common to consider joint TDMA/FDMA/CDMA/SDMA schemes to reduce the number of simultaneous users by allocating them in different TFC slots within a frame. Since the channel matrices for the users are different and determine how the users signals interfere at the receiver, scheduling the users that simultaneously transmit in the same TFC slot, can improve the total system throughput.

3.2.1 Pairing users

Let us first consider the case where K is even and users are paired to transmit simultaneously in the same TFC slot. The total number of TFC slots (or channel orthogonal resources) is then $N = K/2$ and we assume the total occupied bandwidth is still B. Fairness is provided by the fact that all users access the channel exactly once, within a *frame* of N TFC slots. We let $\mathbf{H}^{(k)}$ denote the channel for user k and assume it is constant over the entire frame.

In this case we have that the received signal in the n-th TFC slot is

$$\mathbf{y}(n) = \mathbf{H}^{(k_1)}\mathbf{x}^{(k_1)} + \mathbf{H}^{(k_2)}\mathbf{x}^{(k_2)} + \mathbf{z}(n) \qquad n = 1,\ldots,N \tag{3.5}$$

where $k_1 \neq k_2$. Note that the multiuser detection now handles only two overlapping users per TFC slot and thus even multiuser ML detection could become viable.

We denote by π a particular pairing configuration, within the set of all configurations Π. The number of ways to choose N disjoint pairs of items from $2N$ items is ([1])

$$|\Pi| = (2N-1)!! = (2N-1)(2N-3)\cdots 3\ 1 \ .$$

For example, with $K = 4$ users we have three configurations

$$\Pi = \{\{(12),(34)\},\ \{(13)(24)\},\ \{(14)(23)\}\}$$

Given (3.5), the per-user mutual information between $\mathbf{X}$ and $\mathbf{Y} = (\mathbf{y}(1)^T, \ldots, \mathbf{y}(N)^T)^T$, given a pairing configuration π, is

$$I_{ML}(\mathbf{X};\mathbf{Y}|\mathcal{H},\pi) = \frac{1}{N}\sum_{(k_1,k_2)\in\pi}^{(N)} \log_2\left(\det(\mathbf{I}_{n_r} + \frac{P}{2n_t}\mathbf{H}^{(k_1,k_2)}\mathbf{H}^{(k1,k2)^\dagger})\right) \tag{3.6}$$

where $\mathbf{H}^{(k_1,k_2)} = [\mathbf{H}^{(k_1)}|\mathbf{H}^{(k_2)}]$. Similarly

$$I_{MMSE}(\mathbf{X};\mathbf{Y}|\mathcal{H},\pi) = \frac{1}{N}\sum_{(k_1,k_2)\in\pi}^{(N)}\sum_{j=1}^{2n_t} \log_2\left(1 + \mathbf{h}_j^{(k1,k2)^\dagger}\mathbf{A}_j^{(k1,k2)^{-1}}\mathbf{h}_j^{(k1,k2)}\right) \tag{3.7}$$

where $\mathbf{h}_j^{(k1,k2)}$ are the $2n_t$ columns of $\mathbf{H}^{(k_1,k_2)}$ and

$$\mathbf{A}_j^{(k1,k2)} = \frac{2n_t}{P}\mathbf{I}_{n_r} + \sum_{i=1,i\neq j}^{2n_t} \mathbf{h}_i^{(k1,k2)^\dagger}\mathbf{h}_i^{(k1,k2)}\ .$$

Both (3.6) and (3.7) can be written as additive objective functions to be maximized over the choice of $\pi \in \Pi$

$$\max_{\pi\in\Pi} \sum_{(k_1,k_2)\in\pi}^{(N)} f_{k_1,k_2}(\pi) \tag{3.8}$$

Selecting the pairing configuration that maximizes the above mutual information can become a formidable task even for a small number of users due to the exponential complexity of an exhaustive search. For example, for $K = 2, 4, 6, 8, 10, 16$ we have a number of configurations $|\Pi| = 1, 3, 15, 105, 945, 2027025$. We will show in Sec. 3.2.4 how this problem can be solved in polynomial time using a technique known as Hungarian method.

3.2.2 Both single users and paired users

We now consider the case where we allow some users to transmit alone and some others to be paired in the TFC slots. The total number of users is $K = 2N_{pair} + N_{sing}$, where N_{pair} is the number of pairs of users that transmit simultaneously in a TFC slot and N_{sing} is the number of users that transmit

alone. In this case the total number of TFC slots used in one transmission frame would be $N = N_{pair} + N_{sing}$ and the total number of configurations Π is much larger than before, namely:

$$|\Pi| = \sum_{k=0}^{\lfloor K/2 \rfloor} \frac{K!}{(K-2k)!\, 2^k\, k!}$$

This number corresponds to the number of partitions of a set of K distinguishable elements into sets of size 1 and 2 or equivalently to the number of $K \times K$ symmetric permutation matrices [2]. For example, for $K = 2, 4, 6, 8, 16$ we have $|\Pi| = 2, 10, 76, 764,\ 46206736$ and with $K = 4$ users we have the following 10 configurations

$$\begin{aligned} \Pi \;=\; & \{\{(12)(34)\},\ \{(13)(24)\},\ \{(14)(23)\}, \\ & \{(1)(2)(34)\},\ \{(1)(3)(24)\},\ \{(1)(4)(23)\}, \\ & \{(12)(3)(4)\},\ \{(13)(2)(4)\},\ \{(14)(2)(3)\}, \\ & \{(1)(2)(3)(4)\}\} \end{aligned}$$

In this case the optimization problem becomes

$$\max_{\pi \in \Pi} \left\{ \frac{1}{N_{pair}(\pi)} \sum_{(k_1,k_2)\in\pi}^{(N_{pair}(\pi))} f^{(pair)}_{k_1,k_2}(\pi) \;+\; \frac{1}{N_{sing}(\pi)} \sum_{(k_3)\in\pi}^{(N_{sing}(\pi))} f^{(sing)}_{k_3}(\pi) \right\} \tag{3.9}$$

We can think of the single users (k_3) as paired with themselves, i.e., (k_3, k_3). Unfortunately, this problem cannot be solved by the Hungarian method, since the objective function is not a sum of terms only depending on one pair due to the factors $\frac{1}{N_{pair}(\pi)}$ and $\frac{1}{N_{sing}(\pi)}$ (see Section 3.2.4 for details). Due to the exponential complexity required to solve 3.9 we are motivated to consider the new scheduling scheme of the following section.

3.2.3 New scheduling scheme

In order to have the same total bandwidth for all configurations with different N_{pair} and N_{sing}, we assume that the N_{pair} paired users access two TFC slots, essentially doubling their rate. As a compensation, the N_{sing} unpaired users, that only use one TFC slot, are allowed to double their transmit power. This will produce comparable out-of-cell interfering power during all TFC slots. Now the total number of TFC slots used in one transmission frame would

be $N = 2N_{pair} + N_{sing} = K$. By transmitting with double power, unpaired users can employ a higher order modulation in order to double their spectral efficiency and compensate for their use of only one TFC slot.

Let us now show how a pairing configuration $\pi = \{\pi_{pair}, \pi_{sing}\}$ can be mapped to a permutation σ of K elements of the form

$$\sigma : \begin{pmatrix} 1 & 2 & \cdots & K \\ \sigma(1) & \sigma(2) & \cdots & \sigma(K) \end{pmatrix} . \tag{3.10}$$

Let the pairs $(k_1, k_2) \in \pi_{pair}$ correspond to the two columns of (4.7) $(k_1, k_2 = \sigma(k_1))^T$ and $(k_2, k_1 = \sigma(k_2))^T$, while the unpaired users $(k_3) \in \pi_{sing}$ correspond to the fixed elements of the permutation, i.e., columns of (4.7) of the type $(k_3, k_3)^T$. For example

$$\pi = \{(1,5)(2,4)(3)\} \quad \Rightarrow \quad \sigma : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 4 & 3 & 2 & 1 \end{pmatrix}$$

Clearly, under the assumptions of the Sections 3.2.1 and 3.2.2 this will limit the permutations σ to have at most cycles of length 2 of the type (k_1, k_2), [16].

In this new scenario we can further expand the possible pairing configurations to include any user permutation σ, i.e., we will consider K pairs of users $(k, \sigma(k))$. For example we can have

$$\pi = \{(1,5)(2,4)(3,3)(4,5)(5,2)\}$$
$$\Rightarrow \quad \sigma : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 4 & 3 & 1 & 2 \end{pmatrix}$$

which is a permutation with a cycle $(1, 5, 2, 4)$ of length 4.

The optimization problem can now be written as

$$\max_{\sigma \in S_K} \sum_{(k,\sigma(k))}^{(K)} f_{(k,\sigma(k))} \tag{3.11}$$

where S_K denotes the group of all permutations (symmetric group).

3.2.4 Solving the combinatorial optimization problem

Here, we show how the above combinatorial optimization problems (3.8) and (3.11) can be solved in polynomial time $O(n^3)$ using a technique known as the

Hungarian method commonly used to solve the so called *assignment problem* [35][36][44].

Assignment problem: Given a weighted complete bipartite graph $G = (X \cup Y; X \times Y)$, where edge xy has weight $w(xy)$, find a matching M from X to Y with maximum weight.

In a common application, X could be a set of workers, Y could be a set of jobs, and $w(xy)$ could be the profit made by assigning worker x to job y. By adding virtual jobs or workers with 0 profitability, we may assume that X and Y have the same size, n, and can be written as $X = \{x_1; x_2; \ldots, x_n\}$ and $Y = \{y_1, y_2, \ldots, y_n\}$.

Mathematically, the problem can be stated as follows: given an $n \times n$ matrix $W = [w_{k,\ell}] = [w(x_k y_\ell)]$, find a permutation $\sigma \in S_n$ of n elements for which

$$\sum_{k=1}^{n} w(x_k y_{\sigma(k)})$$

is a maximum. This form coincides with (3.11) when $w(x_k y_{\sigma(k)}) = f_{(k,\sigma(k))}$.

In order to solve the problem (3.8) in the case of even K, where no users are allowed to be unpaired it is enough to initialize the matrix W with zero entries on the diagonal and symmetric entries $w_{k_1,k_2} = w_{k_2,k_1} = f_{(k_1,k_2)}$. The final solution is found by taking only the pairs $(k, \sigma(k))$, for $k = 1, \ldots, K/2$.

3.3 Simulation results

In this section we show some examples of the gains provided by the proposed scheduling schemes in Sec. 3.2.1 and Sec. 3.2.3. In our results we evaluate system capacity as the average over the channel realizations of the mutual information.

We assume $n_t = n_r = 2$. Figure 3.1 shows capacity expressed in bits per channel use (bpcu) as a function of the number of users, when all the users are paired as for (3.8). Both ML and MMSE receivers are compared when the scheduling is optimized and when it is based on a fixed random assignment. The single user (SU) case is plotted for reference and coincides for MMSE and ML detection. Similar results are shown in Fig. 3.2 for the case (3.11).

The following observations are in order in both cases.

- There is a substantial capacity gain over the single user case, thanks to the spatial multiplexing of the users.
- The gain of optimal scheduling increases for increasing number of users (for $K = 2$ there is obviously no difference).

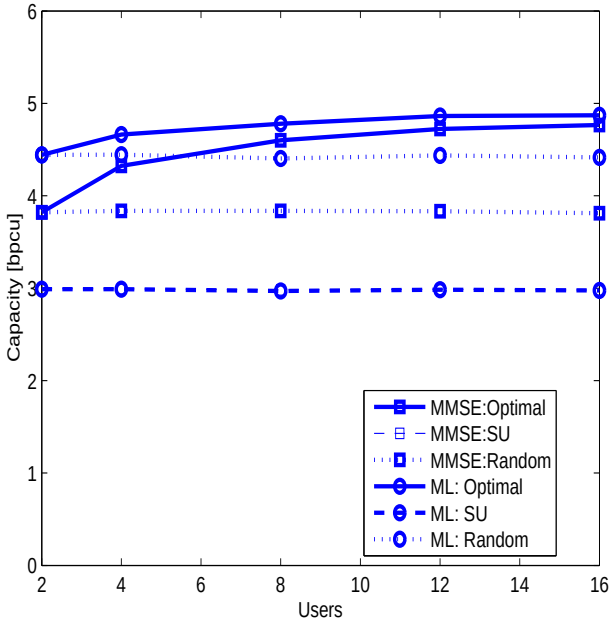


Figure 3.1: Paired users only at $\mathsf{SNR} = 16dB$. Note that ML and MMSE single user curves fully overlap.

- The gain of optimal scheduling is larger for MMSE receiver since ML can handle better ill-conditioned situations.
- For large K the MMSE seems to approach the ML capacity.

3.4 Conclusion

In this Chapter we have proposed a new signalling scheme based on the optimal scheduling of pairs of users. This particular scheme provides capacity gains using a polynomial time algorithm to compute optimum scheduling. In particular, it enables to improve the performance of the suboptimal MMSE receiver and to approach, for large number of users, the performance of the ML receiver which has a higher complexity.

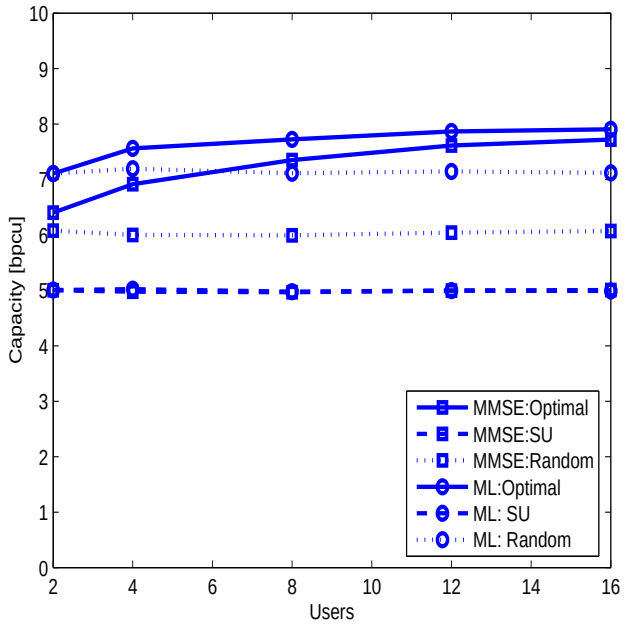


Figure 3.2: Both pairs and single users at $\mathsf{SNR} = 13dB$. Note that ML and MMSE single user curves fully overlap.

Chapter 4

Device collaboration in ad-hoc MIMO networks

A. Hottinen, T. Heikkinen, and E. Viterbo

In this Chapter we consider methods for determining device coalitions for collaborative signal transmission, where different devices act as relay nodes to peers. The problem is to determine for R total number of users and R transmission slots the subsets of at most two devices that are allowed to transmit simultaneously. The subset selection problem is shown to be equivalent to an assignment problem. We consider both optimal assignment and greedy assignment and demonstrate the performance benefit due to device cooperation with simulations in a network model that models path loss between devices.

4.1 Introduction

In future networks different devices could potentially help each other in signal transmission, using each others hardware in an opportunistic way. Amplify-forward (AF) relaying is a potential candidate for such systems, since with AF, the relaying node need not know all transport parameters of the source node (as it does not decode the signal). On the other hand, AF relays are known to enhance also noise. Therefore, a randomly selected AF device can amplify noise to the extent that it has detrimental effect on network capacity.

In a practical network there are typically multiple AF-relaying devices and a limited number of orthogonal subchannels (time-frequency slots). The device population needs to be divided into subsets of active devices for each

transmission subchannel. In addition, the roles (if a device acts as source or as a relay) for each device in each subset and channel use need to be determined.

Related subset selection and scheduling problems have appeared in uplink MU-MIMO [8], relay scheduling [32][30], and in sensor networks [50]. Here, the subset selection problem considered from a MIMO relay network viewpoint, where a source and a co-channel relay jointly form a MIMO channel to a common destination node.

In the current application, we allow at most two devices to collaborate in a given channel use. We use sum-throughput (mutual information) of a MIMO relay channel as a performance measure when determining cooperative user coalitions. Unpaired devices are also allowed, if deemed beneficial. Unpaired devices transmit directly to the destination node (no relaying). A paired device transmits a part of its signal to a peer device during one channel use. In the next channel use, the paired devices transmit simultaneously to the destination node.

4.2 System Model

4.2.1 Relay model

We have a population of R devices each with one transmit antenna. Signal transmission is divided into two hops. In the first hop a source is allowed to communicate with the selected $K < R$ peers. In the second transmission hop the source and the selected peers transmit simultaneously to the destination node, which is assumed to have $N_{\mathrm{r}} \geq K$ receive antennas. The second hop channel is a Multiple Input Multiple Output (MIMO) channel. Formally, the signal model follows that of a MIMO relay network.

During the first hop, the source device transmits signal vector $\mathbf{x}$ with power P_1 through a $K \times K$ first hop channel $\mathbf{F}$, where K designates the number of active devices in the second hop channel. The off-diagonal terms of $\mathbf{F}$ (i.e. $|f_{k,n}|^2, n \neq k$) designate interference power due to source n at relay k input. Obviously, interference power vanishes for all relays if matrix $\mathbf{F}$ is diagonal. In this case, each device receives and retransmits a fraction $1/K$ of signal vector $\mathbf{x}$ during the second hop.

The $N_{\mathrm{r}} \times K$ second hop MIMO channel from the (selected) K devices to the destination is given by $\mathbf{H}$. During the second hop, each of the K devices multiply the signal with a relay-specific weighting coefficient w_k to satisfy a

transmit power constraint at relay. We let

$$w_k = \sqrt{\frac{P_2/K}{\sum_{n=1}^{K} |f_{k,n}|^2 + \sigma_k^2}} \tag{4.1}$$

where σ_k^2 designates noise power at kth relay and P_2 is the desired sum transmit power over all K relay nodes. Note that if interference terms and noise power vanish, the relay only modifies the transmit power of the original signal. In the second hop channel, for notational simplicity, the original source device is modelled a special AF relay with zero noise and interference power at relay input.

We collect the relay weights into a diagonal matrix

$$\Lambda = \mathrm{diag}(w_1, ..., w_K).$$

The destination receives

$$\mathbf{y} = \mathbf{H}\Lambda\mathbf{F}\mathbf{x} + \mathbf{H}\Lambda\mathbf{n}_r + \mathbf{n}_d$$

where the elements of complex Gaussian vector

$$\mathbf{n}_r = (n_1, ..., n_K)^T$$

designate noise with variance σ_k^2 at k'th relay node, and elements of

$$\mathbf{n}_d = (n_1, ..., n_{N_r})^T$$

designate complex Gaussian noise in each destination antenna. We assume that noise power is identical in each receiver antenna, i.e each has variance σ_d^2. The mutual information with i.i.d. Gaussian sources (in terms of bits-per-channel-use (bpcu)) for the considered signal model is [63]

$$\alpha = \frac{1}{2} \log_2 \det(\mathbf{I} + \mathbf{H}\Lambda\mathbf{F}\mathbf{F}^\dagger\Lambda^\dagger\mathbf{H}^\dagger\mathbf{C}_{nn}^{-1}), \tag{4.2}$$

where the noise correlation matrix is

$$\mathbf{C}_{nn} = (\sigma_d^2\mathbf{I} + \mathbf{H}\Lambda\mathrm{diag}(\sigma_1^2, ..., \sigma_K^2)\Lambda^\dagger\mathbf{H}^\dagger).$$

Factor 1/2 in model (4.2) is due to two-hop relaying.

4.2.2 Subset selection

We consider a special case of the subset selection problem to reduce computational burden of the optimization algorithm. Instead of allowing arbitrary-sized subsets, we determine identities of only $K \leq 2$ second-hop devices for each channel use. We assume that each of the R devices is a source in exactly one of R channel uses. Assuming that $N_{\mathrm{r}} = 2$, each second hop MIMO channel supports $K \leq 2$ simultaneously transmitting devices. Moreover, each of the devices acts as a relay exactly once in the R channel uses, to incorporate a notion of fairness to relay selection. That is, we determine for R sources and R transmission slots the distinct ordered subsets of at most two devices. We first describe the optimal (in system throughput sense) algorithm used for subset selection and then summarize the reference cases, greedy subset selection and random subset selection.

Since, $K \leq 2$, we need to compute the mutual information α_{r_1,r_2} when device $r_1 \in \{1, ..., R\}$ is the source device and device $r_2 \in \{1, ..., R\}$ is the relay device. In general, $\alpha_{r_1,r_2} \neq \alpha_{r_2,r_1}$ since $\mathbf{F}$,$\mathbf{H}$ and Λ matrices also depend on these indices (omitted to simplify notation). When $r_1 = r_2$ ($K = 1$), the source transmits directly to destination with double power.

Optimal selection: Consider the selection of devices over R channel uses (via the following linear programming problem ([35]):

$$\arg \max_{(z_{r_1,r_2})} \sum_{r_2}^{R} \sum_{r_1}^{R} \alpha_{r_1,r_2} z_{r_1,r_2} \tag{4.3}$$

subject to

$$\sum_{r_1=1}^{R} z_{r_1,r_2} = 1, \forall r_2 \tag{4.4}$$

$$\sum_{r_2} z_{r_1,r_2} = 1, \forall r_1, \tag{4.5}$$

$$z_{r_1,r_2} \geq 0, \forall r_1, r_2, \tag{4.6}$$

The variables z_{r_1,r_2}, solved from above problem, dictate which devices become active source and relay nodes in each of the R slots. The model implicitly assumes all assignments involve either direct transmission or device pairing. When considering matrix (z_{r_1,r_2}), the solution to problem (4.3)-(4.6) dictates that there is exactly one non-zero element in each row and column, thus ensuring that all nodes act as sources equal number of times. When two nodes are active, either node may take the role of a source, while the other functions as a relay node. Whenever the $z_{r_1,r_2} = 1$, and $r_1 < r_2$, r_1 acts as source and r_2 relays. This convention results from the way the indices in

eq. 4.3 are enumerated. When $z_{r_1,r_2} = 1$, with $r_1 = r_2$, only the direct link is activated, and relaying is disabled. Recall that problem (4.3)-(4.6) and the resulting permutation matrix can be solved efficiently (with polynomial complexity) applying transportation algorithm [35].

Naturally, considerably simpler subset selection algorithms exist:

Random selection: In random subset selection, the matrix (z_{r_1,r_2}) is defined as a random permutation matrix.

Greedy selection: In the first iteration of a Greedy subset selection the column and row indices of the largest element of (α_{r_1,r_2}) determine an element of the solution matrix. Then, the elements of these rows and columns are set to zero and maximum indices are sought in the following iteration from the modified matrix. This guarantees that the indices are unique for each iteration and that after R iterations a permutation matrix emerges.

4.3 Numerical results

We study the arising collaboration patterns in a simple two-dimensional network. The R devices are placed randomly (uniformly) on a 20×20 rectangular area (meter units) with lower-left corner at coordinate $(0, 25)$. The destination position is $(30, 50)$. We assume $K = 2, N_{\mathrm{r}} = 2$, so that only device pairing or direct transmission is allowed. The 2×2 network matrices $\mathbf{F}$ and $\mathbf{H}$ are computed using a simple path-loss model as follows: the distance between nodes r_1 and r_2 is d_{r_1,r_2} meters and the first-hop link matrix is set to

$$\mathbf{F} = \mathrm{diag}(1, \sqrt{P}_1/d_{r_1,r_2}^{2.3/2})$$

when devices r_1 and r_2, $r_1 \neq r_2$ are paired. The transmit power $P_1 = 27$ dB. For direct transmission ($r_1 = r_2$) the path-loss model is obviously neither applicable or relevant due to the weighting method given in eq. 1. Thus, to model direct transmission in the relay framework, we set $\mathbf{F} = \mathrm{diag}(1, 1)$.

Due to applied weighting, the total second-hop transmit power is identical for direct and paired transmission. The second-hop matrix is of form

$$\mathbf{H} = \mathrm{diag}(\sqrt{P}_2/d_{r_1,d}^{2.3/2}, \sqrt{P}_2/d_{r_2,d}^{2.3/2})\tilde{\mathbf{H}},$$

where $d_{r_1,d}$ and $d_{r_2,d}$ is the distance device r_1 and r_2 and the destination node, respectively, and P_2 is the transmit power on second hop. We set $P_2 = 31.7$ dB. Matrix $\tilde{\mathbf{H}}$ is an i.i.d. complex Gaussian-distributed MIMO matrix, where each element has unit power.

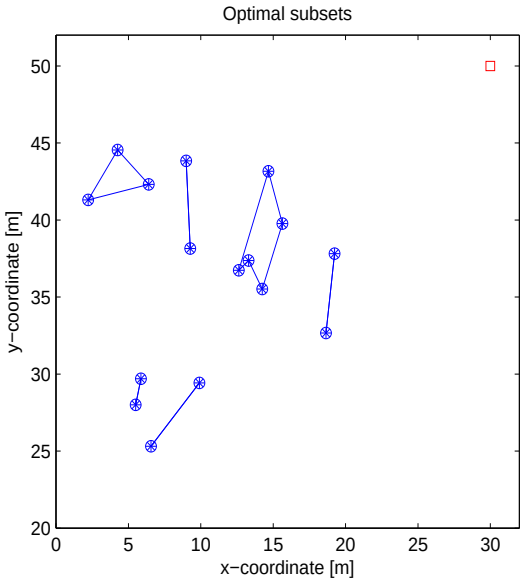


Figure 4.1: Example of device collaboration patterns for *optimal subset selection* with $R = 16$. Destination is located on top-right corner, marked with character '□'. Collaboration patterns include 4 cycles of length 2, 1 cycle of length 3, and one cycle of length 5.

4.3.1 Collaboration patterns

The optimization schemes in previous section each determine a permutation of matrix of dimension R. The non-zero value on the rth row of the permutation matrix is mapped to element $\sigma(r)$, i.e. $z_{r,\sigma(r)} = 1$ in terms of notation in section 2.2. We say that devices $(r, \sigma(r))$ form a collaboration pair. The permutation matrices arising from optimal, greedy or random subset selection can each be mapped to a permutation σ of R elements of the form

$$\sigma : \begin{pmatrix} 1 & 2 & \cdots & R \\ \sigma(1) & \sigma(2) & \cdots & \sigma(R) \end{pmatrix} . \tag{4.7}$$

If $r = \sigma(r)$, device r is unpaired. The unpaired devices correspond to the fixed elements of the permutation. In our relay model, this corresponds to the case, where a device transmits directly to the destination node.

If two devices, say r_1 and r_2, use each other as their respective relays, these devices form a pair (r_1, r_2). If in addition, r_2 uses r_1 as a relay, the corresponding permutation includes columns (4.7) $(r_1, r_2 = \sigma(r_1))^T$ and $(r_2, r_1 = \sigma(r_2))^T$.

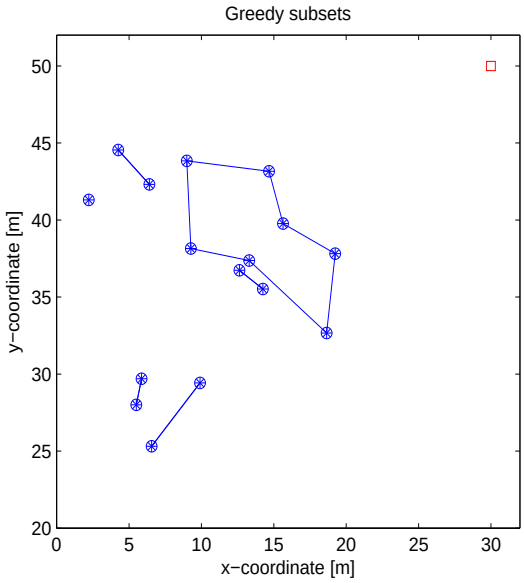


Figure 4.2: Example of device collaboration patterns for *greedy subset selection* with $R = 16$. Destination is located on top-right corner, marked with character '□'. Collaboration patterns include 1 cycle of length 1 (no relaying), 4 cycles of length 2, and 1 cycle of length 7.

In terms of [16][8], unpaired devices correspond to cycles of length 1, while paired users that use each other as relays correspond to cycles of length 2. Naturally, an arbitrary permutation σ, e.g.

$$\sigma : \quad \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 2 & 3 & 1 & 4 \end{pmatrix}$$

can have longer cycles. Above we have a cycle $(1, 5, 4)$ of length 3. In the relay model, device 1 uses devices 5 as relay in the first channel use, device 2 is unpaired in the second channel use, and so on. The collaboration pattern is thus $\{(1, 5)(2, 2)(3, 3)(4, 1)(5, 4)\}$ in 5 channel uses, and it includes two unpaired users.

We first illustrate the emerging cooperation patterns using one realization of device locations and channels. In Fig. 4.1 the optimal device collaboration patterns for each transmission slot are computed by solving problem (4.3)-(4.6). In Fig. 4.2 the same is done for greedy heuristics. In both figures, the destination receiver is located on top-right corner with character '□'. In the two figures a line is drawn between two devices cooperative devices. For

cycles of length 2, the two devices act as source and relay nodes for each other in alternate channel uses. For cycles with length 3 or higher, a device acts as source and relay for two different devices in separate channel uses. For example, in Fig. 4.1 a cycle of length 3 appears in top-left corner. It takes three channel uses to serve all three devices.

4.3.2 Performance

Fig. 4.3 depicts the ergodic performance (average mutual information) for four different subset selection schemes (optimal, greedy, random, direct/no pairing) with $R \in \{2, 4, 8, 12, 16\}$ single-antenna devices and one dual-antenna destination node. The results are averaged for each R over 1000 device locations each with independently generated MIMO channel. For optimal subset selection, the device collaboration patterns for each are computed from problem (4.3)-(4.6) and related mutual information is recorded. The mutual information arising from optimal subsets are shown in figures with legend 'Optimal'. For comparison, we also depict the performance with random subsets - these results are associated with legend 'Random'. The following observations are in order:

- Channel-aware subset selection provides a substantial capacity gain over both direct transmission and random device pairing, thanks to its ability to select network-optimal MIMO relays for the second-hop channel.
- The gain due to optimal subset increases with increasing number of devices. This is in part due to the fact that network is denser and cooperation occurs with devices that are closer.

4.4 Conclusions

We have considered device cooperation as means to form relay-based MIMO uplink. In the considered scheme optimal device collaboration patters (device subsets) are computed (up to pairs) using optimal and greedy matching algorithms. The subset selection algorithms determine which of the R devices should be paired and which should transmit directly to the destination in R channel uses. We demonstrated the performance gain (in terms sum mutual information) with simulations. It is observed that the device subsets have cyclic structure. If the cycle length is three, three devices need to form a coalition when forming source-relay pairs. A topic for future work is to consider subset selection from the point of view of cooperative game theory.

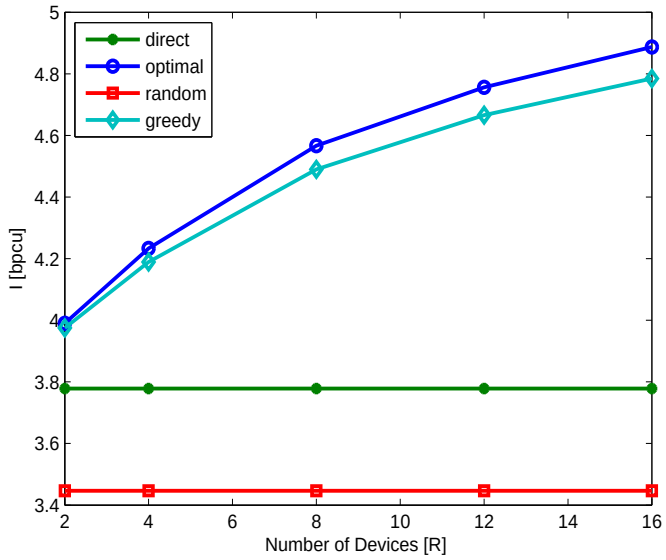


Figure 4.3: Average mutual information (I) at destination node with different number of devices (R) and different pairing schemes. Direct transmission is depicted as reference.

4.5 Conclusions

We have considered device cooperation as means to form relay-based MIMO uplink. In the considered scheme optimal device collaboration patters are computed (up to pairs, so far) using matching (assignment) algorithms. We demonstrated the performance gain (in terms sum capacity) with simulations, and remarked that optimal device subsets tend to have cyclic structure.

Chapter 5

Diversity-Multiplexing Tradeoff in Multi-User Relay Channels

Cemal Akçaba and Helmut Bölcskei

We analyze fading relay networks, where L users with M-antenna each communicate with an N-antenna destination terminal through a set of half-duplex relays using a half-duplex relaying protocol with linear processing at the relay level. We derive the diversity multiplexing tradeoff curve under the assumption that relays employ unitary transformations. We observe that the benefit (in terms of diversity gain) of having K relay terminals is shared by all the users in the system. We further note that cooperation at the relay level cannot improve performance any further.

5.1 Introduction

Previous work and contributions

We analyze fading relay networks, where L users with M-antenna each communicate with an N-antenna destination terminal through a set of K half-duplex single-antenna relays using a half-duplex relaying protocol with linear processing at the relay level (see Fig. 5.1). The contributions in this Chapter can be summarized as follows:

- Our work leads to a characterization of the DMT of the multi-access relay channel for half-duplex relaying protocols[1].

[1]Half-duplex relaying protocols refer to protocols where cooperation takes place over equal-duration receive and transmit phases. No cooperating terminal is allowed to transmit during the receive phase, and to receive during the transmit phase.

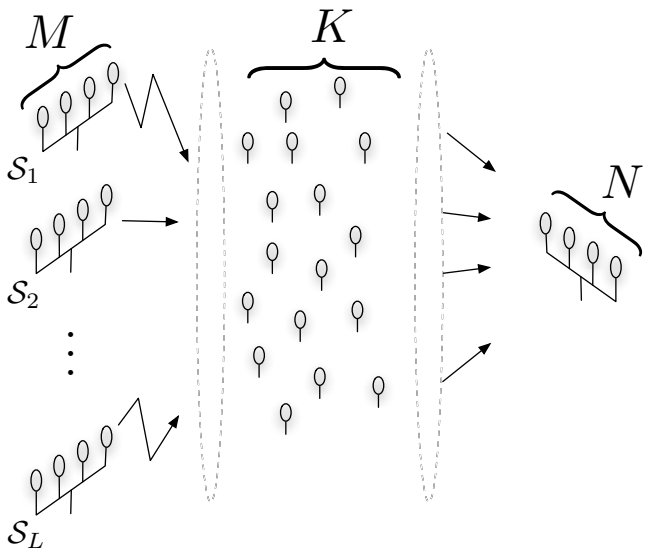


Figure 5.1: System depiction.

- We demonstrate that using relays as active scatterers have the benefit of increasing the diversity gain even when the number of destination antennas is kept fixed. The relays do not need channel state information to reap diversity benefits, and using an appropriate unitary matrix at each relay suffices to have the diversity effect.
- Each user benefits (equally) from having relays in the network.
- The relays do not need to be aware of the number of users in the network. Likewise, the users need not to be aware of the number of relays. The destination terminal however needs to track all the channels in the network.

Notation

The superscripts ^{T,H} and * stand for transpose, conjugate transpose, and conjugation, respectively. x_i represents the ith element of the column vector $\mathbf{x}$, and $X_{i,j}$ stands for the element in the ith row and jth column of the matrix $\mathbf{X}$. $\mathbf{X} \circ \mathbf{Y}$ denot es the Hadamard product of the matrices $\mathbf{X}$ and $\mathbf{Y}$. $\operatorname{rank} \mathbf{X}$ stands for the rank of $\mathbf{X}$. $\operatorname{Tr}(\mathbf{X})$, $\|\mathbf{X}\|_F$, and $\lambda_i(\mathbf{X})$ $(i = 0, 1, \ldots, N-1)$ denote the trace, the Frobenius norm, and the ith eigenvalue (sorted in descending order) of $\mathbf{X}$, respectively. For $N \times N$ positive semi-definite matrices $\mathbf{A}, \mathbf{B}$, $\mathbf{A} \succ \mathbf{B}$ means $\lambda_i(\mathbf{A}) > \lambda_i(\mathbf{B})$ $(i = 0, 1, \ldots, N-1)$, $\mathbf{A} \succeq \mathbf{B}$ is similarly

defined. $\mathbf{I}_N$ is the $N \times N$ identity matrix. $\mathbf{0}$ denotes the all zeros matrix of appropriate size. We say that the square matrices $\mathbf{X}$ and $\mathbf{Y}$ are orthogonal if $\langle \mathbf{X}, \mathbf{Y} \rangle = \mathrm{Tr}\left(\mathbf{X}\mathbf{Y}^H\right) = 0$. All logarithms are to the base 2 and $(a)^+ = \max(a, 0)$. $\mathrm{diag}(a_1, \ldots, a_N)$ denotes the $N \times N$ diagonal matrix with a_i on diagonal entry i. The $N \times N$ discrete Fourier transform (DFT) matrix, $\mathbf{F}$, has entries $F_{ln} = \frac{1}{\sqrt{N}} e^{-j\frac{2\pi}{N}(l-1)(n-1)}$. $X \sim \mathcal{CN}(0, \sigma^2)$ stands for a circularly symmetric complex Gaussian random variable (RV) with variance σ^2. $f(\rho) \doteq g(\rho)$ denotes exponential equality in ρ of the functions $f(\cdot)$ and $g(\cdot)$, i.e.,

$$\lim_{\rho \to \infty} \frac{\log f(\rho)}{\log \rho} = \lim_{\rho \to \infty} \frac{\log g(\rho)}{\log \rho}.$$

The symbols $\dot{\geq}$, $\dot{\leq}$, $\dot{>}$ and $\dot{<}$ are defined analogously.

5.2 System Model

Assumptions and Signal Model

We assume that communication takes place over T slots. Each source terminal i transmits a $T \times M$ matrix, $\mathbf{S}_{(i)}$, via its transmit antennas over T time slots. Each relay k receives a faded copy of each source transmission

$$\mathbf{r}_{(k)} = \sqrt{\rho} \sum_{i=1}^{L} \mathbf{S}_{(i)} \mathbf{f}_{(i,k)} + \mathbf{n}_{(k)} \tag{5.1}$$

where $\mathbf{f}_{(i,k)}$ denotes the $M \times 1$ vector channel from source i to relay k and $\mathbf{n}_{(k)}$ is complex circularly symmetric white gaussian noise at relay k. The vector $\mathbf{f}_{(i,k)}$ has i.i.d. $\mathcal{CN}(0, 1)$ entries with $f_{(i,k)_j}$ representing the fading coefficient from the j^{th} antenna of source i to relay k and $\mathbf{f}^{(i,k)} = [f_{(ik)_1}\ f_{(ik)_2}\ \cdots\ f_{(ik)_M}]^T$. We assume that each $T \times M$ transmit codeword obeys $\|\mathbf{S}_i\|_F^2 = TM$. Each relay multiplies its received signal by a $T \times T$ unitary matrix $\mathbf{G}_k$ $(k = 1, 2, \ldots, K)$ and transmits this signal sequentially after adjusting to its average power constraint[2]. Each relay has a transmit power constraint of $T\rho/K$

$$\mathbb{E}\{\mathbf{r}_{(k)}^H \mathbf{r}_{(k)}\} \tag{5.2}$$

$$= \rho\, \mathbb{E}\{\sum_{i=1}^{L} \sum_{j=1}^{L} \mathbf{f}_{(j,k)}^H \mathbf{S}_{(j)}^H \mathbf{S}_{(i)} \mathbf{f}_{(i,k)}\} + \mathbb{E}\{\mathbf{n}_{(k)}^H \mathbf{n}_{(k)}\} \tag{5.3}$$

[2]The average power constraint at the relays can be replaced with instantaneous power constraint without changing the import of our results

$$= \rho \sum_{i=1}^{L} \mathbb{E}\{\tilde{\mathbf{f}}_{(i,k)}^{H} \Lambda_{\mathbf{S}_{(i)}^{H}\mathbf{S}_{(i)}} \tilde{\mathbf{f}}_{(i,k)}\} + T \tag{5.4}$$

$$= \rho \sum_{i=1}^{L} \mathbb{E}\{\sum_{n=1}^{M} \lambda_n(\Lambda_{\mathbf{S}_{(i)}^{H}\mathbf{S}_{(i)}})\} + T \tag{5.5}$$

$$= \rho \sum_{i=1}^{L} \mathbb{E}\{\mathrm{Tr}\Big(\mathbf{S}_{(i)}^{H}\mathbf{S}_{(i)}\Big)\} + T \tag{5.6}$$

$$= (\rho LM + 1)T \tag{5.7}$$

where step (5.5) follows by the singular value decomposition (SVD) of $\mathbf{S}_{(i)}\mathbf{S}_{(i)}^{H}$ and unitary invariance in distribution of the Gaussian vector $\mathbf{f}_{(i,k)}$. Therefore each relay scale their received signal by $\beta = \sqrt{\frac{\rho}{K(\rho LM+1)}}$ prior to retransmission. The received signal at destination is then given by:

$$\mathbf{Y}^T = \sum_{k=1}^{K} \beta \mathbf{G}_{(k)} \mathbf{r}_{(k)} {\mathbf{g}_{(k)}}^T + \mathbf{Z}^T \tag{5.8}$$

$$= \sum_{k=1}^{K} \beta \mathbf{G}_{(k)} \left(\sqrt{\rho} \sum_{i=1}^{L} \mathbf{S}_{(i)} \mathbf{f}_{(i,k)} + \mathbf{n}_{(k)} \right) {\mathbf{g}_{(k)}}^T + \mathbf{Z}^T$$

$$= \sqrt{\rho} \sum_{k=1}^{K} \sum_{i=1}^{L} \beta \mathbf{G}_{(k)} \mathbf{S}_{(i)} \mathbf{f}_{(i,k)} {\mathbf{g}_k}^T + \tilde{\mathbf{Z}}^T \tag{5.9}$$

$$= \beta\sqrt{\rho} \sum_{i=1}^{L} \sum_{k=1}^{K} \mathbf{G}_{(k)} \mathbf{S}_{(i)} \mathbf{f}_{(i,k)} {\mathbf{g}_k}^T + \tilde{\mathbf{Z}}^T$$

$$= \sqrt{\rho}\beta \sum_{i=1}^{L} \mathbf{X}_{(i)}^T \mathbf{H}_{(i)}^T + \tilde{\mathbf{Z}}^T \tag{5.10}$$

where $KM \times N$ effective channel matrices are defined as

$$\mathbf{H}_{(i)}^T = \begin{bmatrix} \mathbf{f}_{(i,1)} {\mathbf{g}_1}^T \\ \mathbf{f}_{(i,2)} {\mathbf{g}_2}^T \\ \vdots \\ \mathbf{f}_{(i,K)} {\mathbf{g}_K}^T \end{bmatrix} \tag{5.11}$$

for $(i = 1, 2, \ldots, L)$ and $T \times KM$ effective codewords are given by

$$\mathbf{X}_i^T = \begin{bmatrix} \mathbf{G}_{(1)}\mathbf{S}_{(i)} & \mathbf{G}_{(2)}\mathbf{S}_{(i)} & \cdots & \mathbf{G}_{(K)}\mathbf{S}_{(i)} \end{bmatrix} \tag{5.12}$$

and the $T \times N$ effective noise term is as follows

$$\tilde{\mathbf{Z}}^T = \beta \sum_{k=1}^{K} \mathbf{G}_{(k)} \mathbf{n}_{(k)} \mathbf{g}_{(k)}^T + \mathbf{Z}^T \tag{5.13}$$

and β is the power normalization at relays given above. The effective channel matrices $\mathbf{H}_{(i)}$ are correlated for different values of i, and furthermore they are also correlated with the effective noise term $\tilde{\mathbf{Z}}$. We assume that the destination operates with full channel state information of all the links in the network. The receiver left multiplies its received vector at time instance n, in order to make the effective noise term spatially and temporally white

$$\mathbf{Y} = \sqrt{\rho}\beta \sum_{i=1}^{L} \tilde{\mathbf{H}}_{(i)} \mathbf{X}_{(i)} + \mathbf{N} \tag{5.14}$$

where the i^{th} column of $\tilde{\mathbf{H}}$ is $\tilde{\mathbf{H}}_i = \mathbf{W}(\mathbf{H}^T)_i$, $\mathbf{N}$ is the circularly symmetric unit variance complex Gaussian noise matrix and $\mathbf{W}$ is given by

$$\mathbf{W} = \mathbf{U}\Lambda^{-\frac{1}{2}} \tag{5.15}$$

$$\mathbf{R}_{\tilde{\mathbf{Z}}} = \mathbf{U}\Lambda\mathbf{U}^H = (\mathbf{I}_N + \mathbf{N}_K\mathbf{N}_K^H) \tag{5.16}$$

where $\mathbf{N}_K = [\mathbf{g}_1\ \mathbf{g}_2\ \cdots \mathbf{g}_K]$. Note that this normalization does not alter the signal-to-noise ratio, but renders the effective noise term in (5.14) spatially and temporarily white.

5.3 Overview of Single User Results

With only one user present in the system, the system model is given by

$$\mathbf{Y} = \sqrt{\rho}\beta\tilde{\mathbf{H}}\mathbf{X} + \mathbf{N}. \tag{5.17}$$

The mutual information between $\mathbf{X}$ and $\mathbf{Y}$ given the channel is maximized by choosing Gaussian codebooks that are i.i.d. over space and time. That is,

$$I(\mathbf{Y};\mathbf{X}|\tilde{\mathbf{H}}) = \frac{1}{2T}\sum_{i=1}^{T} \log\det\left(\mathbf{I}_N + \tilde{\rho}\tilde{\mathbf{H}}\mathbf{R}_{\mathbf{xx}}^{(i)}\tilde{\mathbf{H}}^H\right) \tag{5.18}$$

where $\tilde{\rho} = \rho\beta^2$ and $\mathbf{R}_{\mathbf{xx}}^{(i)} = \mathbb{E}\{\mathbf{X}_i\mathbf{X}_i^H\}$, is maximized by choosing $\mathbf{R}_{\mathbf{xx}}^{(i)} = \mathbf{I}_N$. Note that Gaussian codebooks are in general not permitted by the power constraint on $\mathbf{S}$, however as we are seeking an upper bound on mutual information, we are allowed to choose an input distribution from a larger set. Further we point out that certain choices of linear processing matrices will not permit such a structure on the codebook covariance. For the sake of obtaining an upper bound, we assume that $\{\mathbf{G}_i\}$ are chosen such that

$\mathbf{R}_{\mathbf{xx}}^{(i)} = \mathbf{I}_N$ for all $i = (1, 2, \ldots, T)$. Under these assumptions, (5.18) simplifies to

$$I(\mathbf{Y}; \mathbf{X}|\tilde{\mathbf{H}}) = \frac{1}{2} \log \det \left(\mathbf{I}_N + \tilde{\rho} \tilde{\mathbf{H}} \tilde{\mathbf{H}}^H \right) \tag{5.19}$$

Following the framework in [64], we define the probability of outage at multiplexing gain r and SNR ρ as the probability of mutual information between the transmit signal $\mathbf{x}$ and the received signal $\mathbf{y}$ given the channel matrix falling below the target rate $r \log \rho$, i.e.

$$P_{\mathcal{O}}(\rho, r) = \mathbb{P}\Big[I(\mathbf{Y}; \mathbf{X}|\tilde{\mathbf{H}}) < r \log \rho \Big]. \tag{5.20}$$

We need the high SNR behavior of this outage probability in the remainder of the text.

Theorem 1 (Rao and Hassibi [49])**.** *The outage probability given in (5.20) satisfies*

$$d_{M,K,N}(r) = - \lim_{\rho \to \infty} \frac{\log P_{\mathcal{O}}(\rho, r)}{\log \rho} \tag{5.21}$$

$$= (\min(M, N) - 2r)^+ (K - 2r)^+ \tag{5.22}$$

for $r = 0, 1, \ldots, \min(M, N, K)$.

Proof. See [49]. The statement of theorem in [49] contains a typo; k in the statement of the theorem should run from 0. Also justification of the effective noise cancellation in [49] is missing several important steps which should be provided; for example as done in [7][Eqns. (53-55) and (79-82)]. The rest of the proof is by establishing an equivalence to the eigenvalue distribution of a Wishart matrix. □

The DM-tradeoff realized by a family (one at each SNR ρ) of codebooks $\mathcal{C}_r$ with rate $R = r \log \rho$ is given by the function

$$d(r) = - \lim_{\rho \to \infty} \frac{\log P_e(\rho, r)}{\log \rho}$$

where $P_e(\rho, r)$ is the error probability obtained through maximum likelihood (ML) decoding. We say that $\mathcal{C}_r$ operates at multiplexing gain r. For a given SNR ρ, the codebook $\mathcal{C}_r(\rho) \in \mathcal{C}_r$ contains ρ^{2MTr} codewords $\hat{\mathbf{S}}$. For any two codewords $\hat{\mathbf{S}}, \tilde{\mathbf{S}} \in \mathcal{C}_r(\rho)$, we define the codeword difference matrix as

$$\mathbf{\Phi} = \begin{bmatrix} \Delta \mathbf{S}^T \mathbf{G}_{(1)}^T \\ \Delta \mathbf{S}^T \mathbf{G}_{(2)}^T \\ \vdots \\ \Delta \mathbf{S}^T \mathbf{G}_{(K)}^T \end{bmatrix} \tag{5.23}$$

where $\Delta\mathbf{S} = \hat{\mathbf{S}} - \tilde{\mathbf{S}}$. Let $\eta = \min(N, MK)$ and λ_i denote the i^{th} eigenvalue of $\mathbf{\Phi}$. For the remainder of this section, fix $T > MK$. First part of the following theorem is also presented in [49].

Theorem 2 (DMT)**.** *For the system described described by (5.20), the optimal DM-tradeoff curve is given by piecewise linear curve joining the points*

$$d^*(r) = (K - 2r)(\min(M, N) - 2r), 2r \in [0, \min(K, M, N)]. \tag{5.24}$$

Let $[\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_K]$ *be a set of matrices and* $\mathcal{C}_r$ *be a family of codebooks such that for any codebook* $\mathcal{C}_r(\rho) \in \mathcal{C}_r$ *and any two codewords* $\tilde{\mathbf{S}}, \hat{\mathbf{S}} \in \mathcal{C}_r(\rho)$ *the condition*

$$|\lambda_1|^2|\lambda_2|^2 \cdots |\lambda_\eta|^2 \doteq \rho^{-2r} \tag{5.25}$$

holds. Then the ML decoding error probability satisfies

$$P_e(\rho, r) = \rho^{-d^*(r)} \tag{5.26}$$

Proof. The proof is essentially due to [54]. We provide it here for completeness and some insights that are not directly stated in [54]. We start by calculating the worst case pairwise error probability given the channel

$$\mathbb{P}\left[\hat{\mathbf{X}} \to \tilde{\mathbf{X}}|\tilde{\mathbf{H}}\right] = Q\left(\sqrt{\frac{\tilde{\rho}}{2}\|\tilde{\mathbf{H}}\mathbf{\Phi}\|^2}\right). \tag{5.27}$$

The singular value decompositions of the effective channel matrix $\tilde{\mathbf{H}}$ and the codeword difference matrix $\mathbf{\Phi}$ are given by

$$\tilde{\mathbf{H}} = \mathbf{U}_1\mathbf{\Psi}\mathbf{V}_1^H \tag{5.28}$$

$$\mathbf{\Phi} = \mathbf{U}_2\mathbf{\Lambda}\mathbf{V}_2^H. \tag{5.29}$$

Rewriting the pairwise error probability given the channel (5.27) as

$$\mathbb{P}\left[\hat{\mathbf{X}} \to \tilde{\mathbf{X}}|\tilde{\mathbf{H}}\right] = Q\left(\sqrt{\frac{\tilde{\rho}}{2}\|\mathbf{\Psi}\mathbf{V}_1^H\mathbf{U}_2\mathbf{\Lambda}\|^2}\right) \tag{5.30}$$

The worst-case channel aligns the weaker singular values of the channel matrix with the stronger singular values of the codeword difference matrix. As given in [54], the worst-case rotation is given by

$$\mathbf{V}_1 = \mathbf{U}_1 \tag{5.31}$$

when the singular-values are placed in ascending order in $\mathbf{\Lambda}$ and in descending order in $\mathbf{\Psi}$. That is

$$\text{diag}(\mathbf{\Lambda}) = [\lambda_1 \ \lambda_2 \ \cdots \ \lambda_{MK}] \tag{5.32}$$

$$\text{diag}(\mathbf{\Psi}) = [\psi_\eta \ \psi_{\eta-1} \ \cdots \ \psi_1] \tag{5.33}$$

where $\eta = \min(MK, N)$. Hence, the worst-case pairwise error probability is given by

$$\mathbb{P}\left[\hat{\mathbf{X}} \to \tilde{\mathbf{X}}|\tilde{\mathbf{H}}\right] = Q\left(\sqrt{\frac{\tilde{\rho}}{2}\sum_{i=1}^{\eta}|\psi_{\eta+1-i}|^2|\lambda_i|^2}\right). \tag{5.34}$$

Next natural question to ask is the following: What is the worst-case pairwise error probability for channels that are not in outage? The condition for the effective channel not to be in outage is given by

$$I(\mathbf{Y};\mathbf{X}|\tilde{\mathbf{H}}) = \sum_{i=1}^{\eta}\log(1+\tilde{\rho}|\psi_i|^2) \geq 2R(1+\epsilon). \tag{5.35}$$

Hence, we have to minimize the argument of the $Q(\cdot)$ function in (5.34) subject to the constraint on the singular values of the channels that are not in outage (5.35)

$$\min_{\psi_1,\psi_2,\ldots,\psi_\eta} \frac{\tilde{\rho}}{2}\sum_{i=1}^{\eta}|\psi_{\eta+1-i}|^2|\lambda_i|^2 \tag{5.36}$$

As done in [54], this problem can be solved using Lagrange multiplier method and the solution is given by standard water-filling. However, in the proof provided in [54], it is implicitly assumed that all $\psi_i > 0$ (meaning that *the worst effective channel realization* has *full rank*), leading to the solution

$$\mathbb{P}\left[\hat{\mathbf{X}} \to \tilde{\mathbf{X}}|\tilde{\mathbf{H}}\right] = Q\left(\sqrt{\frac{1}{2}\sum_{i=1}^{\eta}\left(\frac{1}{\mu}-|\lambda_i|^2\right)^+}\right) \tag{5.37}$$

where μ satisfies

$$\sum_{i=1}^{\eta}\left[\log\left(\frac{1}{\mu|\lambda_i|^2}\right)\right]^+ = 2R(1+\epsilon). \tag{5.38}$$

The goal of a good code is to minimize the worst-case error probability given in (5.37) or in other words to maximize the argument of $Q(\cdot)$ function in (5.37)

$$\sum_{i=1}^{\eta}\left(\frac{1}{\mu}-|\lambda_i|^2\right)^+. \tag{5.39}$$

Next we investigate this quantity. In the following, we say that a singular value λ_i is *a*ctive if

$$\frac{1}{\mu} \geq |\lambda_i|^2. \tag{5.40}$$

For the sake of discussion, assume that only the smallest singular value is active, i.e.

$$\frac{1}{\mu} \geq |\lambda_1|^2. \tag{5.41}$$

(5.41) implies two important properties. The first is the value of $\frac{1}{\mu}$, which is given by

$$\frac{1}{\mu} = 2^{2R(1+\epsilon)}|\lambda_1|^2. \tag{5.42}$$

The second implication is the relation between $\frac{1}{\mu}$ and the second smallest singular value λ_2, which must obey

$$|\lambda_2|^2 > \frac{1}{\mu}. \tag{5.43}$$

Next, putting (5.41), (5.41) and (5.42) together, we get

$$|\lambda_2|^2 \geq 2^{2R(1+\epsilon)}|\lambda_1|^2 \geq |\lambda_1|^2. \tag{5.44}$$

The relation (5.44) is for the case when only 1 singular value is active. If there are $k \leq \eta$ singular values that are active, then we have from (5.38) that

$$\sum_{i=1}^{k} \log\left(\frac{1}{\mu|\lambda_i|^2}\right) = 2R(1+\epsilon). \tag{5.45}$$

which can be explicitly solved to give

$$\log \prod_{i=1}^{k}\left(\frac{1}{\mu|\lambda_i|^2}\right) = 2R(1+\epsilon) \tag{5.46}$$

$$\left(\frac{1}{\mu^k \prod_{i=1}^{k}|\lambda_i|^2}\right) = 2^{2R(1+\epsilon)} \tag{5.47}$$

$$\frac{1}{\mu} = \left(2^{2R(1+\epsilon)} \prod_{i=1}^{k}|\lambda_i|^2\right)^{\frac{1}{k}}. \tag{5.48}$$

Hence for any $k < \eta$ we have

$$|\lambda_{k+1}|^2 \geq \left(2^{2R(1+\epsilon)} \prod_{i=1}^{k} |\lambda_i|^2\right)^{\frac{1}{k}} \geq |\lambda_k|^2 \tag{5.49}$$

and for $k = \eta$, we have

$$\left(2^{2R(1+\epsilon)} \prod_{i=1}^{\eta} |\lambda_i|^2\right)^{\frac{1}{\eta}} \geq |\lambda_\eta|^2. \tag{5.50}$$

Now let us consider the error probability (5.37) that we would like to minimize. We would like to choose the argument of the $Q(\cdot)$ function as large as possible. In other words, we need to find out how many singular values need to be active in order to maximize the quantity

$$\sum_{i=1}^{k} \left(\frac{1}{\mu} - |\lambda_i|^2\right). \tag{5.51}$$

We start by manipulating this expression

$$\begin{aligned} \sum_{i=1}^{k} \left(\frac{1}{\mu} - |\lambda_i|^2\right) &= \frac{k}{\mu} - \sum_{i=1}^{k} |\lambda_i|^2 \\ k \left(2^{2R(1+\epsilon)} \prod_{i=1}^{k} |\lambda_i|^2\right)^{\frac{1}{k}} &- \sum_{i=1}^{k} |\lambda_i|^2 \end{aligned} \tag{5.52}$$

where (5.52) follows from (5.48). We define

$$\zeta(k) = k \left(2^{2R(1+\epsilon)} \prod_{i=1}^{k} |\lambda_i|^2\right)^{\frac{1}{k}} - \sum_{i=1}^{k} |\lambda_i|^2. \tag{5.53}$$

Note further that for two integers $k_1, k_2 \in [0, \eta]$, if $k_2 \geq k_1$, then

$$\zeta(k_2) \geq \zeta(k_1). \tag{5.54}$$

To demonstrate how (5.54) follows, we consider the difference $\zeta(k_2) - \zeta(k_1)$ and show that this difference is always non-negative;

$$\zeta(k_2) - \zeta(k_1) \tag{5.55}$$

$$= k_2 \left(2^{2R(1+\epsilon)} \prod_{i=1}^{k_2} |\lambda_i|^2\right)^{\frac{1}{k_2}} - \sum_{i=1}^{k_2} |\lambda_i|^2 - k_1 \left(2^{2R(1+\epsilon)} \prod_{i=1}^{k_1} |\lambda_i|^2\right)^{\frac{1}{k_1}} + \sum_{i=1}^{k_1} |\lambda_i|^2 \tag{5.56}$$

$$\geq k_2 \left(2^{2R(1+\epsilon)} \prod_{i=1}^{k_2} |\lambda_i|^2 \right)^{\frac{1}{k_2}} - \sum_{i=1}^{k_2} |\lambda_i|^2 - k_1 |\lambda_{k_1+1}|^2 + \sum_{i=1}^{k_1} |\lambda_i|^2 \tag{5.57}$$

$$= k_2 \left(2^{2R(1+\epsilon)} \prod_{i=1}^{k_2} |\lambda_i|^2 \right)^{\frac{1}{k_2}} - \sum_{i=k_1+1}^{k_2} |\lambda_i|^2 - k_1 |\lambda_{k_1+1}|^2 \tag{5.58}$$

$$\geq k_2 |\lambda_{k_2}|^2 - (k_2 - k_1)|\lambda_{k_2}|^2 - k_1 |\lambda_{k_1+1}|^2 \tag{5.59}$$

$$\geq k_2 |\lambda_{k_2}|^2 - (k_2 - k_1)|\lambda_{k_2}|^2 - k_1 |\lambda_{k_2}|^2 = 0. \tag{5.60}$$

Step (5.57) is due to (5.49), (5.57) is obtained by subtracting out common terms in two sums and (5.59) is due to the RHS of (5.49) (for values of $k_2 < \eta$) and (5.50) (for $k_2 = \eta$).

Therefore the value of k that maximizes the argument of the error function in (5.37) is $k = \eta$. In other words, if we would like to design a good code for the worst case full rank channel which is not in outage, we need to have a code difference matrix with η non-zero singular values. Hence, the worst-case pairwise error probability for a good code given the set of effective channels that are not in outage is upper bounded by

$$\mathbb{P}\left[\hat{\mathbf{X}} \to \tilde{\mathbf{X}}, \mathcal{O}^c\right] = \mathbb{E}_{\tilde{\mathbf{H}} \in \mathcal{O}^c} \left\{ Q\left(\frac{1}{\sqrt{2}} \sqrt{\eta \left(2^{2R(1+\epsilon)} \prod_{i=1}^{\eta} |\lambda_i|^2 \right)^{\frac{1}{\eta}} - \sum_{i=1}^{\eta} |\lambda_i|^2} \right) \right\} \mathbb{P}[\mathcal{O}^c] \tag{5.61}$$

$$\leq \exp - \left[\frac{1}{4} \left(\eta \left(2^{2R(1+\epsilon)} \prod_{i=1}^{\eta} |\lambda_i|^2 \right)^{\frac{1}{\eta}} - \sum_{i=1}^{\eta} |\lambda_i|^2 \right) \right] \mathbb{P}[\mathcal{O}^c] \tag{5.62}$$

$$\leq \exp - \left[\frac{1}{4} \left(\eta \left(2^{2R(1+\epsilon)} \prod_{i=1}^{\eta} |\lambda_i|^2 \right)^{\frac{1}{\eta}} - \sum_{i=1}^{\eta} |\lambda_i|^2 \right) \right]. \tag{5.63}$$

(5.61) follows by Bayes' rule, whereas (5.62) follows by dropping the expectation operator since we are considering the worst channel that is not in outage and by the inequality $Q\left(\sqrt{\frac{x}{2}}\right) \leq \exp[-\frac{x}{4}]$. (5.63) is justified since $\mathbb{P}[\mathcal{O}^c] \leq 1$. By the conditions of the theorem we have that $|\lambda_1|^2 |\lambda_2|^2 \cdots |\lambda_\eta|^2 \doteq \rho^{-2r}$, hence

$$\mathbb{P}\left[\hat{\mathbf{X}} \to \tilde{\mathbf{X}}, \mathcal{O}^c\right] \leq \exp - \left[\frac{1}{4} \left(\eta \rho^{\frac{\epsilon}{\eta}} - \sum_{i=1}^{\eta} |\lambda_i|^2 \right) \right] \tag{5.64}$$

$$\leq \exp - \left[\frac{1}{4} \left(\eta \rho^{\frac{\epsilon}{\eta}} - 4TM \right) \right] \tag{5.65}$$

$$\dot{\leq} \exp \left[-\frac{1}{4} \eta \rho^{\frac{\epsilon}{\eta}} \right] \tag{5.66}$$

where (5.65) follows from the power constraint on the transmit codewords, and (5.66) is due to the exponential growth of $\eta\rho^{\frac{\epsilon}{\eta}}$ compared to the constant $4TM$ in the limit of high SNR. Since the ML-decoder error probability can be upper bounded by

$$P_{\text{error}}(\rho, r) \leq P_{\mathcal{O}}(\rho, r) + \rho^{2MTr}\mathbb{P}\left[\hat{\mathbf{X}} \to \tilde{\mathbf{X}}, \mathcal{O}^c\right] \tag{5.67}$$

$$\dot{\leq} P_{\mathcal{O}}(\rho, r) + \rho^{2MTr} \exp\left[-\frac{1}{4}\eta\rho^{\frac{\epsilon}{\eta}}\right] \doteq P_{\mathcal{O}}(\rho, r). \tag{5.68}$$

In $|\lambda_1|^2|\lambda_2|^2\cdots|\lambda_\eta|^2 \dot{>} \rho^{-2r}$, we have provided a sufficient condition for the worst case PEP to decay exponentially in SNR. However as pointed out above, our derivation was valid for a channel matrix that has η eigenvalues. In order to show that the condition is still sufficient when the efficient channel matrix is rank deficient, we consider an illustrative example[3]. Assume that channel matrix has only one non-zero singular value, i.e.

$$\psi_\eta > 0 \text{ and } \psi_i = 0 \ \forall i \ \in [1, \eta - 1]. \tag{5.69}$$

Then the worst case PEP given the channel is

$$\mathbb{P}\left[\hat{\mathbf{X}} \to \tilde{\mathbf{X}}|\tilde{\mathbf{H}}\right] = Q\left(\sqrt{\frac{\tilde{\rho}}{2}|\psi_\eta|^2|\lambda_1|^2}\right). \tag{5.70}$$

which can be solved explicitly subject to the no outage condition (5.35) to give

$$\mathbb{P}\left[\hat{\mathbf{X}} \to \tilde{\mathbf{X}}|\tilde{\mathbf{H}}\right] = Q\left(\sqrt{\frac{(2^{2R(1+\epsilon)} - 1)|\lambda_1|^2}{2}}\right) \tag{5.71}$$

$$\leq \exp\left[-\frac{(2^{2R(1+\epsilon)} - 1)|\lambda_1|^2}{4}\right]. \tag{5.72}$$

From which we see that choosing $|\lambda_1|^2 \dot{>} 2^{-2R}$, suffices for this probability to decay exponentially in SNR. Put differently, $|\lambda_1|^2 \dot{>} \rho^{-2r}$ is the condition on the smallest singular value of the code difference matrix. Now in general, the condition $|\lambda_1|^2|\lambda_2|^2\cdots|\lambda_\eta|^2 \dot{>} \rho^{-2r}$ does not imply $|\lambda_1|^2 \dot{>} \rho^{-2r}$. This implication, i.e.,

$$|\lambda_1|^2|\lambda_2|^2\cdots|\lambda_\eta|^2 \dot{>} \rho^{-r} \Rightarrow |\lambda_1|^2 \dot{>} \rho^{-2r} \tag{5.73}$$

[3]Personal communication with P.E. Coronel.

holds only when we have the additional constraint that

$$\sum_{i=}^{\eta} |\lambda_i|^2 \doteq \rho^0. \tag{5.74}$$

Note that (5.74) guarantees that each eigenvalue of the codeword difference matrix will grow at most at the same rate with SNR, i.e.

$$|\lambda_i|^2 \dot{\leq} \rho^0 \quad \forall i \in [1, \eta]. \tag{5.75}$$

However, (5.75) and $|\lambda_1|^2|\lambda_2|^2 \cdots |\lambda_\eta|^2 \dot{\geq} \rho^{-2r}$ together imply $|\lambda_1|^2 \dot{\geq} \rho^{-2r}$. We have now proved conclusively that if $|\lambda_1|^2|\lambda_2|^2 \cdots |\lambda_\eta|^2 \dot{\geq} \rho^{-2r}$ for every pair of codewords in every codebook in the family of codebooks, we are guaranteed to have the ML-decoder error probability decay as fast as the outage probability. We have also showed that having a power constrained on the SNR-normalized codebook is also required for these results to hold. This dependency of the proof on the power constrained is not mentioned in [54].

We have not argued about the optimality of this tradeoff. However, an application of a cut-set bound [10][6] yields that the outage probability is lower bounded by

$$P_{\text{MIMO}M\times K}(\rho, 2r) + P_{\text{MIMO}K\times N}(\rho, 2r) \;\leq\; P_{\mathcal{O}}(\rho, r) \tag{5.76}$$

$$P_{\text{MIMO}\,\min(M,N)\times K}(\rho, 2r) \dot{\leq} P_{\mathcal{O}}(\rho, r) \tag{5.77}$$

$$P_{\text{MIMO}\,\min(M,N)\times K}(\rho, 2r) \;\doteq\; P_{\mathcal{O}}(\rho, r) \tag{5.78}$$

where $P_{\text{MIMO}m\times n}(\rho, 2r)$ is the outage probability of the Rayleigh-fading $m \times n$ MIMO channel [64] after the half-duplex nature of the relaying scheme is taken into account. (5.76) is due to the cut-set upper bound on the mutual information, (5.77) is by retaining the slowest decaying (in SNR) probability in the LHS of (5.76) and finally (5.78) since the SNR exponents of $P_{\text{MIMO}\,\min(M,N)\times K}(\rho, 2r)$ and $P_{\mathcal{O}}(\rho, r)$ are equal to the first order. We have hence showed that

$$P_{\text{MIMO}\,\min(M,N)\times K}(\rho, 2r) \doteq P_{\mathcal{O}}(\rho, r) \leq P_e(\rho, r) \dot{\leq} P_{\mathcal{O}}(\rho, r) \tag{5.79}$$

$$P_e(\rho, r) \doteq \rho^{-(K-2r)^+(\min(M,N)-2r)^+}. \tag{5.80}$$

□

5.4 Multi-User Relay System

Our contribution in this section is the characterization of the achievable DMT of a multi-user system described by (5.14). The characterization also

sheds light into the structure of space-time codes that would be achieving the corresponding DMT. For the remainder of this section, we assume without loss of generality that $T \geq LMK$.

Theorem 3 (DMT for Half-Duplex Protocols)**.** *For a multiple access system operating under a half-duplex protocol with L users of M antenna each, K single-antenna relays and an N antenna destination, assume that user i is operating at rate $R_i = r_i \log \rho$. The ML-decoder error probability behaves like*

$$P_e(\rho, \mathbf{r}) = \rho^{-d^*(\mathbf{r})} \tag{5.81}$$

where $\mathbf{r} = (r_1, r_2, \ldots, r_L)$ and

$$d^*(\mathbf{r}) = \min_{\mathcal{S}} d_{|\mathcal{S}|M,K,N}\left(\sum_{i\in\mathcal{S}} r_i\right) \tag{5.82}$$

with $\mathcal{S} \subseteq \{1, 2, \ldots, L\}$ if the normalized codeword difference matrices $\mathbf{\Phi}_{\mathcal{S}}$ for all pairs of codewords in the family of codebooks satisfy

$$\prod_{j=1}^{\min(N,|\mathcal{S}|MK)} |\lambda_j(\mathbf{\Phi}_{\mathcal{S}})|^2 \dot{>} \rho^{-2\sum_{i\in\mathcal{S}} r_i} \tag{5.83}$$

for every subset $\mathcal{S}$.

Proof. As in [57], we define the outage event as follows for a multiple access channel. For a multiple access channel with L users, each equipped with M transmit antennas, and a receiver with N receive antennas the outage event is defined as

$$\mathcal{O} \triangleq \bigcup_{\mathcal{S}} \mathcal{O}_{\mathcal{S}} \tag{5.84}$$

where the union in (5.84) is taken over all subsets $\mathcal{S} \subseteq \{1, 2, \ldots, L\}$ and $\mathcal{O}_{\mathcal{S}}$ is defined as

$$\mathcal{O}_{\mathcal{S}} \triangleq \left\{\tilde{\mathbf{H}} \in \mathbb{C}^{LMK\times N} : I(\mathbf{Y}; \mathbf{C}_{\mathcal{S}}|\mathbf{C}_{\mathcal{S}^c}, \mathbf{H} = \tilde{\mathbf{H}}) < \sum_{i\in\mathcal{S}} R_i\right\}$$

where $\mathbf{C}_{\mathcal{S}}$ contains the input signals from the users in the set $\mathcal{S}$ and $\mathbf{C}_{\mathcal{S}^c}$ contains the signals from the users in the complementary set $\mathcal{S}^c$. By subtracting the signals from the users in $\mathcal{S}^c$ and allowing cooperation between the users in $\mathcal{S}$, we can write

$$\mathbf{Y} = \sqrt{\rho}\beta\tilde{\mathbf{H}}_{\mathcal{S}}\mathbf{C}_{\mathcal{S}} + \mathbf{N} \tag{5.85}$$

where $\tilde{\mathbf{H}}_{\mathcal{S}}$ contains the fading coefficients for the system where $|\mathcal{S}|$ users are allowed to cooperate. By allowing the users in $\mathcal{S}$ to cooperate, the problem takes a simpler form with $|\mathcal{S}|M$ transmit antennas, K relays and N receive antennas communicating over a half-duplex communication protocol. To get a lower bound on the outage probability, we can assume $\mathbf{C}_{\mathcal{S}}$ to have an i.i.d. Gaussian distribution (as done in the proofs of theorems 1 and 2). Now let the target data rate for user i be $R_i = r_i \log \rho$ for $i \in (1, 2, ..., L)$. By theorem 1, the outage probability in the limit of high SNR for the system where the users in the set $\mathcal{S}$ are allowed to cooperate satisfies

$$\mathbb{P}[\mathcal{O}_{\mathcal{S}}] \doteq \rho^{-d^*_{|\mathcal{S}|M,K,N}(\sum_{i\in\mathcal{S}} r_i)}. \tag{5.86}$$

From the definition of outage event in (5.84), we know that

$$\mathbb{P}[\mathcal{O}] = \mathbb{P}\left[\bigcup_{\mathcal{S}} \mathcal{O}_{\mathcal{S}}\right] \leq \sum_{\mathcal{S}} \mathbb{P}[\mathcal{O}_{\mathcal{S}}] \doteq \mathbb{P}[\mathcal{O}_{\mathcal{S}^*}] \tag{5.87}$$

where $\mathcal{S}^*$ is the subset of $\{1, 2, \ldots, L\}$ with the slowest decay rate of $\mathbb{P}[\mathcal{O}_{\mathcal{S}}]$, i.e.,

$$\mathcal{S}^* = \arg\min_{\mathcal{S}} d^*_{|S|M,K,N}\left(\sum_{i\in\mathcal{S}} r_i\right). \tag{5.88}$$

Combining with the fact that $\mathbb{P}[\mathcal{O}] \geq \mathbb{P}[\mathcal{O}_{\mathcal{S}^*}]$, we must have

$$\mathbb{P}[\mathcal{O}_{\mathcal{S}^*}] \leq \mathbb{P}[\mathcal{O}] = \mathbb{P}\left[\bigcup_{\mathcal{S}} \mathcal{O}_{\mathcal{S}}\right] \dot{\leq} \mathbb{P}[\mathcal{O}_{\mathcal{S}^*}]. \tag{5.89}$$

From (5.89), we conclude

$$\mathbb{P}[\mathcal{O}] \doteq \mathbb{P}[\mathcal{O}_{\mathcal{S}^*}]. \tag{5.90}$$

To obtain an upper bound on the error probability, we use the smart union bound approach [54][13] and analyze the worst case pairwise error probability (PEP). Consider the error probability of the joint ML decoder. We define the type $\mathcal{S}$-error event as

$$\mathcal{E}^{\mathcal{S}} \triangleq \{\hat{m}_i = m_i, \forall i \in \mathcal{S}^c \text{and } \hat{m}_i \neq m_i, \forall i \in \mathcal{S}\} \tag{5.91}$$

where $\hat{m}_i$ is the decoded message for user i and the non-empty set $\mathcal{S}$ is $\mathcal{S} \subseteq \{1, \ldots, L\}$. Thus $\mathcal{E}^{\mathcal{S}}$ is the event that the receiver makes wrong decisions

on the messages of all the users in the set $\mathcal{S}$, and makes correct decisions on the rest.

$$\mathbb{P}[\text{Error}] = \mathbb{P}\left[\bigcup_{\mathcal{S}} \mathcal{E}^{\mathcal{S}}\right] \leq \sum_{\mathcal{S}} \mathbb{P}\left[\mathcal{E}^{\mathcal{S}}\right] \tag{5.92}$$

Assume without loss of generality that $\mathcal{S} = \{1, 2, \ldots, |\mathcal{S}|\}$.

$$\mathbf{C}^T = [\tilde{\mathbf{X}}_1^T \ \tilde{\mathbf{X}}_2^T \ldots \tilde{\mathbf{X}}_{|\mathcal{S}|}^T \ \tilde{\mathbf{X}}_{|\mathcal{S}|+1}^T \ldots \tilde{\mathbf{X}}_L^T]$$

is the set of effective transmit codewords stacked together.

$$\mathbf{E}^T = [\hat{\mathbf{X}}_1^T \ \hat{\mathbf{X}}_2^T \ldots \hat{\mathbf{X}}_{|\mathcal{S}|}^T \ \tilde{\mathbf{X}}_{|\mathcal{S}|+1}^T \ldots \tilde{\mathbf{X}}_L^T]$$

is another set of effective transmit codewords. For a type $\mathcal{S}-$type event we have

$$\mathbf{C}^T = [\tilde{\mathbf{X}}_1^T \ \tilde{\mathbf{X}}_2^T \ldots \tilde{\mathbf{X}}_{|\mathcal{S}|}^T \ \tilde{\mathbf{X}}_{|\mathcal{S}|+1}^T \ldots \tilde{\mathbf{X}}_L^T] \tag{5.93}$$

$$\mathbf{E}^T = [\hat{\mathbf{X}}_1^T \ \hat{\mathbf{X}}_2^T \ldots \hat{\mathbf{X}}_{|\mathcal{S}|}^T \ \tilde{\mathbf{X}}_{|\mathcal{S}|+1}^T \ldots \tilde{\mathbf{X}}_L^T] \tag{5.94}$$

where $\tilde{\mathbf{X}}_i \neq \hat{\mathbf{X}}_i$ for all $i = 1, 2, \ldots, |\mathcal{S}|$. Now this event occurs if the receiver makes a wrong decision in favor of one such effective codeword $\mathbf{E}$

$$\mathbb{P}\left[\mathbf{C} \to \mathbf{E}|\tilde{\mathbf{H}}\right] = Q\left(\sqrt{\frac{\tilde{\rho}}{2}\|\tilde{\mathbf{H}}(\mathbf{C} - \mathbf{E})\|^2}\right) \tag{5.95}$$

$$= Q\left(\sqrt{\frac{\tilde{\rho}}{2}\|\tilde{\mathbf{H}}_{\mathcal{S}}(\mathbf{C}_{\mathcal{S}} - \mathbf{E}_{\mathcal{S}})\|^2}\right) \tag{5.96}$$

$$= Q\left(\sqrt{\frac{\tilde{\rho}}{2}\|\tilde{\mathbf{H}}_{\mathcal{S}}\mathbf{\Phi}_{\mathcal{S}}\|^2}\right) \tag{5.97}$$

where $\tilde{\mathbf{H}}_{\mathcal{S}} = [\tilde{\mathbf{H}}_1 \ \tilde{\mathbf{H}}_2 \cdots \ \tilde{\mathbf{H}}_{|\mathcal{S}|}]$ and $\mathbf{\Phi}_{\mathcal{S}} = \mathbf{C}_{\mathcal{S}} - \mathbf{E}_{\mathcal{S}}$. Now if for every pair of codewords $\mathbf{E}$ and $\mathbf{C}$ in every codebook

$$\prod_{j=1}^{\min(N, |\mathcal{S}|MK)} |\lambda_j(\mathbf{\Phi}_{\mathcal{S}})|^2 \dot{>} \rho^{-2\sum_{i\in\mathcal{S}} r_i} \tag{5.98}$$

where are the smallest $\min(N, |\mathcal{S}|MK)$ singular values of the normalized codeword difference matrix for every subset $\mathcal{S}$, then analysis from the proof Theorem 2 guarantees that in the limit of high SNR each $\mathbb{P}\left[\mathcal{E}^{\mathcal{S}}\right]$ decays according to

$$\mathbb{P}\left[\mathcal{E}^{\mathcal{S}}\right] \doteq \rho^{-d_{|\mathcal{S}|M,K,N}(\sum_{i\in\mathcal{S}} r_i)} \tag{5.99}$$

which when put together with (5.92) gives

$$\mathbb{P}[\text{Error}] = \mathbb{P}\left[\bigcup_{\mathcal{S}} \mathcal{E}^{\mathcal{S}}\right] \leq \sum_{\mathcal{S}} \mathbb{P}\left[\mathcal{E}^{\mathcal{S}}\right] \doteq \sum_{\mathcal{S}} \rho^{-d_{|\mathcal{S}|M,K,N}\left(\sum_{i\in\mathcal{S}} r_i\right)} \tag{5.100}$$

$$\mathbb{P}[\text{Error}] \dot{\leq} \rho^{-d_{|\mathcal{S}^*|M,K,N}\left(\sum_{i\in\mathcal{S}^*} r_i\right)} \tag{5.101}$$

where $\mathcal{S}^*$ is given by

$$\mathcal{S}^* = \arg\min_{\mathcal{S}} d^*_{|S|M,K,N}\left(\sum_{i\in\mathcal{S}} r_i\right). \tag{5.102}$$

Putting it together we have

$$\mathbb{P}[\mathcal{O}] \leq \mathbb{P}[\text{Error}] \leq \sum_{\mathcal{S}} \mathbb{P}\left[\mathcal{E}^{\mathcal{S}}\right] \tag{5.103}$$

$$\rho^{-d_{|\mathcal{S}^*|M,K,N}\left(\sum_{i\in\mathcal{S}^*} r_i\right)} \dot{\leq} \mathbb{P}[\text{Error}] \dot{\leq} \rho^{-d_{|\mathcal{S}^*|M,K,N}\left(\sum_{i\in\mathcal{S}^*} r_i\right)} \tag{5.104}$$

$$\mathbb{P}[\text{Error}] \doteq \rho^{-d_{|\mathcal{S}^*|M,K,N}\left(\sum_{i\in\mathcal{S}^*} r_i\right)} \tag{5.105}$$

where $\mathcal{S}^*$ is as defined in (5.102). □

We explain the implications of the theorem using a two-user example. The subsets of $\{1,2\}$ are $\{1\},\{2\},\{1,2\}$, meaning that the following codeword difference matrices should obey the condition in the theorem for achieving the DMT *simultaneously*, that is,

$$\prod_{j=1}^{\min(N,MK)} \left|\lambda_j\left(\mathbf{\Phi}^{(1)}\right)\right|^2 \dot{>} \rho^{-2r_1} \tag{5.106}$$

$$\prod_{j=1}^{\min(N,MK)} \left|\lambda_j\left(\mathbf{\Phi}^{(2)}\right)\right|^2 \dot{>} \rho^{-2r_2} \tag{5.107}$$

$$\prod_{j=1}^{\min(N,2MK)} \left|\lambda_j\left(\mathbf{\Phi}^{(1,2)}\right)\right|^2 \dot{>} \rho^{-2(r_1+r_2)} \tag{5.108}$$

where

$$\mathbf{\Phi}^{(1)} = \begin{bmatrix} \Delta\mathbf{S}_1{}^T\mathbf{G}^T_{(1)} \\ \Delta\mathbf{S}_1{}^T\mathbf{G}^T_{(2)} \\ \vdots \\ \Delta\mathbf{S}_1{}^T\mathbf{G}^T_{(K)} \end{bmatrix} \tag{5.109}$$

$$\mathbf{\Phi}^{(2)} = \begin{bmatrix} \Delta\mathbf{S}_2{}^T\mathbf{G}_{(1)}^T \\ \Delta\mathbf{S}_2{}^T\mathbf{G}_{(2)}^T \\ \vdots \\ \Delta\mathbf{S}_2{}^T\mathbf{G}_{(K)}^T \end{bmatrix} \tag{5.110}$$

$$\mathbf{\Phi}^{(1,2)} = \begin{bmatrix} \mathbf{\Phi}^{(1)} \\ \mathbf{\Phi}^{(2)} \end{bmatrix}. \tag{5.111}$$

Discussion

Unless we know more about relative eigenvalues of codeword different matrices in (5.109) and (5.110), we cannot further simplify the conditions (5.106) and (5.107). However, condition (5.108) can be written in a simpler manner observing that non-zero eigenvalues of $\mathbf{\Phi}^{(1,2)}\left(\mathbf{\Phi}^{(1,2)}\right)^H$ and $\left(\mathbf{\Phi}^{(1,2)}\right)^H\mathbf{\Phi}^{(1,2)}$ are identical, and are exactly the magnitude squares of singular values of $\mathbf{\Phi}^{(1,2)}$. Assuming that $MK > N$, it is possible to further simplify (5.108) to

$$\prod_{i=1}^{N}\mu_i\left(\sum_{i=1}^{2}\left(\mathbf{\Phi}^{(i)}\right)^H\mathbf{\Phi}^{(i)}\right) \dot{>} \rho^{-2(r_1+r_2)} \tag{5.112}$$

and (5.106) and (5.107) to

$$\prod_{i=1}^{N}\mu_i\left(\left(\mathbf{\Phi}^{(1)}\right)^H\mathbf{\Phi}^{(1)}\right) \dot{>} \rho^{-2r_1} \tag{5.113}$$

$$\prod_{i=1}^{N}\mu_i\left(\left(\mathbf{\Phi}^{(2)}\right)^H\mathbf{\Phi}^{(2)}\right) \dot{>} \rho^{-2r_2} \tag{5.114}$$

where $\mu_i(\cdot)$ is the i^{th} eigenvalue of the corresponding matrix (in ascending order). In comparison to the frequency selective point-to-point MIMO design criterion at high SNR [21], we see that the conditions (5.106-5.108) are more difficult to satisfy. A dominating error event derivation (e.g. [21]) needs to be carried out here to understand the DMT achieving criterion better.

Implications on the Code Design

The code design criterion tells us that it is sufficient to design the user codes to obey the individual rate constraints *and* jointly in order to obey the sum rate constraint. Can we do better? Is it possible to simplify the design criterion so as to design for the dominating error event? The answer to this difficult question is the following: Even though we can design our codebooks

for the dominating user, we cannot ignore the design choices required for the other user.

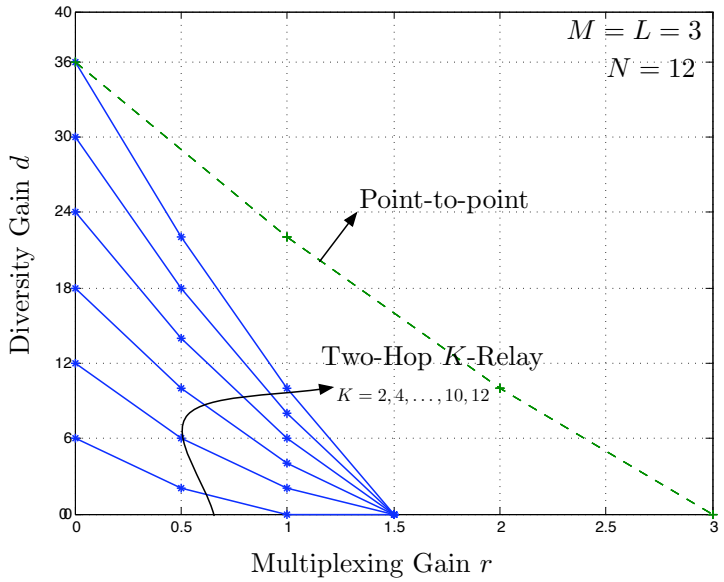


Figure 5.2: Symmetric Tradeoff: DMT curves for a system with 3-users each with 3 transmit antennas, K relays and a 12-antenna destination terminal. For comparison, we include the DMT for a point-to-point system of same dimension (without relays).

Chapter 6

Scaling laws for large ad-hoc wireless networks with Wishart-Poisson fading

Giuseppa Alfano, Maxime Guillaud, and Antonia M. Tulino

In the analysis of large random wireless ad hoc networks, the underlying node distribution is almost ubiquitously assumed to be the homogeneous Poisson point process. Despite the nice analytical properties of such model, the spatial randomness has been, however, mainly exploited for connectivity and interference analysis, but has not yet been taken into account explicitly in the scaling laws evaluation. We move here a first step toward the evaluation of an upper bound on the aggregate throughput when the additional randomness due to the spatial node distribution is taken into account, together with the presence of power attenuation and random phase changes. This could be seen as a first attempt to connect some overoptimistic results based on stochastic channel model to more realistic analysis, relying on electromagnetic propagation arguments.

6.1 Introduction

The question of what is the maximally achievable scaling of the total throughput with the system size in an ad hoc wireless network has been first tackled by Gupta and Kumar in [26] where, considering n nodes, randomly located in the unit disk and wishing to communicate each to a random destination at rate $R(n)$, it is showed that multihop communication with conventional single-user decoding cannot achieve a scaling better than $\mathcal{O}(\sqrt{n})$. This

landmark contribution, which is on its own an interference limited result, has stimulated also the scientific community in providing a quantification of what is the fundamental (i.e. independent on the communication paradigm) limit of throughput scaling in such networks. A substantial amount of works has then appeared, dealing with the scaling evaluation problem under different assumptions on the electromagnetic propagation process, which led to (very) different lower and upper bounds on the overall information rate scaling. The most optimistic results rely on (almost) full cooperation between nodes and/or nodes clusters, and promise scaling of $\mathcal{O}(n^{2/3})$ [51] and even $\mathcal{O}(n^{1-\epsilon})$ [47]. The information-theoretic capacity of a random ad hoc network seems, thus, to well approximate that of a single-user MIMO system, where cooperation between transmit and receive antennas can be fully exploited. The network effect[1] appears as a *panacea* against the interference impairments, and a suitable upper bound on the aggregate rate turns out to be of order $\mathcal{O}(n \log n)$ [47, Th. 3.1]. However, an analysis exploiting spectral behavior of the electromagnetic field showed that, for the assumed geometry, a scaling better than $\mathcal{O}(\sqrt{n} \log n)$ is forbidden by physical limitations [20], since the degrees of freedom of the radiating field, namely, the number of available dimensions in the space whom the radiation operator (i.e. the functional representing the radiated field from the nodes in the *transmit side* of the network) belongs to [12], is bounded by $\sqrt{n} \log n$, as n grows large.

Such a result relies on pure physical (geometric) structure analysis of the communication system under consideration, and shows that more optimistic scenarios are somewhat unrealistic. Usefulness and effectiveness of stochastic channel models cannot be, however, denied at once. The question arise, rather, of how to find a link between the optimistic, stochastic channel based result of linear scaling and the actual square root decay law, dictated by functional analysis.

A first step is made in the present paper by both taking into account some stochastic geometry aspects usually neglected in the common approach, as well as exploiting in our network framework some mutual information expansions, widely adopted in the single-user multiantenna systems analysis.

We first embody spatial randomness, borrowing results from [28], in the characterization of a MIMO communication between geometry-dependent clusters of the networks, providing explicit expressions for some functions of the eigenvalues of the newly defined channel matrix which are useful in the information theoretical analysis. Then, using tools from random matrix theory already proven to be effective in the single-user multiple antenna scenario,

[1] We use the locution to refer to the diversity achievable through the full cooperation among the nodes, as in [47].

we provide the evaluation of the cut-set bound for a large wireless network with Rayleigh faded long-range links, noticing that such an approach allows to recover both the $\mathcal{O}(\sqrt{n}\log n)$ as well as the linear scaling as particular cases.

The paper is organized as follows: Section II contains the system model description, and a list of assumptions on the stochastic propagation process is therein reported. The MIMO Wishart-Poisson model for the wireless channel is defined and analyzed. In Section III a basic SIMO upper bound on the capacity scaling is evaluated under different choices of the path-loss exponent, while a cut-set result is presented and discussed in Section IV.

6.2 System Model and Assumptions

Let $\mathcal{N}$ be a two-dimensional stationary Poisson point process over Re^2, defined on a probability space $(\Omega, \mathcal{F}, P)$.We denote by $E[\cdot]$ the expectation taken with respect to the measure induced by P. The points of the process represent the (fixed) locations of the devices. The process is assumed to have intensity $0 < \lambda < +\infty$, where λ is defined as the expected number of points in the unit square. Given a finite Borel subset $A \in \mathrm{Re}^2$, the number of points of $\mathcal{N}$ in A, denoted by $\mathcal{N}(A)$, is then a Poisson random variable of intensity $\lambda\nu(A)$, where $\nu(A)$ is the Lebesgue measure of the set A. White Gaussian noise of power N_0 is assumed to be present at the receiver, and each node transmits at a prescribed power level P_{tx}.

We assume that the channel gain between any nodes pair is subject to random phase changes, and that each node is in far-field zone with respect to any randomly picked one. The model closely follows the assumptions in [47], and the only difference consists in explicitly taking into account the spatial randomness of the devices locations, whose (ordered) distances probability density function (pdf) is provided in [28].

Our aim is to analytically characterize the communication protocol proposed in [47]; a first step consists of the distribution of the bits to be sent from the source to its first m neighbors, then, a traditional MIMO communication between the two clusters which source and destinations, respectively, belong to, takes place, and finally a joint decoding of the received bits, sent back from its neighbors to the intended destination, is performed. The impact of nodes stochastic geometry on the first and third step will be analyzed, while a deeper analysis of the random distances impact on the inter-cluster

transmission would require exploitation of some results in [11, and references therein] and is subject of ongoing work.

6.2.1 Spatial gain matrix

Let us consider the information sharing between the source and its m neighbors. The point-to-point channel from the source node, sending symbol x_k to its k-th neighbor can be represented as

$$z_k = \sqrt{G} e^{j\theta_k} r_k^{-\alpha/2} x_k + n_k \tag{6.1}$$

where the received signal is indicated as z_k, the gain G can be assumed to be either a deterministic constant (see, e.g., [47]) or a random variable[2], r_k follows the *generalized* Gamma distribution given in [28, Eq. (2)], $\alpha \geq 2$ is the path loss exponent, and θ_k is the random phase on the channel between the source and its k-th neighbor. The random variables r_k and θ_k are assumed to be independent.

In two dimensional random networks, the pdf of the distance from the source to its k-th nearest neighbor[3] can be written as

$$f_{r_k}(r) = e^{-\lambda\pi r^2} \frac{2(\lambda\pi r^2)^k}{r(k-1)!}, \tag{6.2}$$

and, in turn, denoting by $y_k = G r_k^{-\alpha}$ the fading amplitude seen from such neighbor[4],

$$f_{y_k}(y) = \left(\frac{2}{\alpha}\right) \frac{e^{-\lambda\pi (G/y)^{2/\alpha}}}{y(k-1)!} (\lambda\pi)^k \left(\frac{G}{y}\right)^{2k/\alpha}. \tag{6.3}$$

The m-th moment of the random variable y_k has thus the following expression:

$$E[y_k^m] = \frac{G^m (\lambda\pi)^{\frac{\alpha m}{2}}}{(k-1)!} \Gamma\left(k - \frac{\alpha m}{2}\right). \tag{6.4}$$

Such a general model may, however, suffer from the singularity of the path loss function with respect to distance if the network is very dense. A

[2] In this case, we will assume a unit-mean distribution for G, following [28]

[3] There is no loss of generality assuming the source located at the origin *for each* cluster, due to the homogeneous Poisson distribution of the nodes. In fact, it is well known that the spatial nodes distribution conditioned on having a node in the origin is the same as the original one for a homogeneous Poisson point process.

[4] We remark that expression (6.2) refers explicitly to the case of deterministic G, while it can be viewed as a conditional distribution if G is assumed to randomly vary.

way to cope with such problem is to consider a regularized version of the path loss function (see, e.g. [17]), namely to model the channel through

$$z_k = \sqrt{G}e^{j\theta_k} \max\{r_0, r_k\}^{-\alpha/2} x_k + n_k, \tag{6.5}$$

with r_0 a prescribed constant, which regularizes the unbounded behavior of the attenuation function close to the origin. The Cumulative Distribution Function (CDF) of the random variable $u_k = \max\{r_0, r_k\}$ can be written as

$$F_{u_k}(x) = \left[1 - e^{-\lambda\pi x} \sum_{\ell=0}^{n-1} \frac{(\lambda\pi x)^\ell}{l!}\right] \mathbf{1}_{r_0 \leq x},$$

and its pdf as

$$f_{u_k}(x) = \gamma(k, \lambda\pi x)\delta(x - r_0) + \mathbf{1}_{r_0 \leq x} \lambda\pi e^{\lambda\pi x} \frac{(\lambda\pi x)^{k-1}}{(k-1)!}, \tag{6.6}$$

with $\gamma(k, x) = \int_0^x t^{k-1}e^{-t}dt$ the lower incomplete Gamma function. With that,

$$E[u_k^{-\alpha}] = r_0^{-\alpha}\gamma(k, \lambda\pi r_0) + \frac{(\lambda\pi)^\alpha}{(k-1)!}\Gamma(k - \alpha, \lambda\pi r_0),$$

where $\Gamma(k, x) = \int_x^\infty t^{k-1}e^{-t}dt$ is the Incomplete Gamma function. The first step of information sharing among the source and its neighbors can be then modeled through a diagonal random matrix $\mathbf{G}$, which we refer to as the spatial gain matrix from now on, whose squared nonzero entries marginally follow, depending on the chosen attenuation model, either the law (6.3) or (6.6).

To additionally take into account channel randomness due to amplitude fluctuations and not just to stochastic distances, we model G as exponentially distributed with unit mean, i.e. $f_G(x) = e^{-x}$. The random variable $Y_k = Gr_k^{-\alpha}$ pdf can be evaluated through

$$f_{Y_k|r_k}(y) = e^{-yr_k^\alpha} r_k^\alpha,$$

so that

$$f_{Y_k}(y) = \frac{2(\lambda\pi)^k}{(k-1)!} \int_0^\infty e^{-yr_k^\alpha - \lambda\pi r_k^2} r_k^{2k+\alpha-1} dr_k,$$

which for $\alpha = 2$ leads to $f_{Y_k}(y) = \frac{k(\lambda\pi)^k}{(y+\lambda\pi)^{k+1}}$. Notice that this distribution has finite moments up to the order $m = k - 1$, according to previous results in [28].

A general expression for $f_{Y_k}(y)$ when $\alpha > 2$ can be more easily evaluated via $f_{Y_k}(y) = \int_0^\infty f_{Y_k|G}(y|x)e^{-x}dx$, with $f_{Y_k|G}(y|x)$ given by (6.3). Following this way, we finally get

$$f_{Y_k}(y) = \left(\frac{2}{\alpha}\right) \frac{(\lambda\pi)^k}{(k-1)!} \frac{\Gamma(\beta)}{[y(1+y^{-2/\alpha}\lambda\pi)]^\beta}, \tag{6.7}$$

with $\beta = \frac{2k}{\alpha} + 1$.

6.2.2 The MIMO Wishart-Poisson model

The second step of the communication can be modeled as a MIMO communication on the linear channel

$$\mathbf{z} = \mathbf{H}\,\mathbf{G}\,\mathbf{x} + \mathbf{n}$$

where $\mathbf{x}$ and $\mathbf{z}$ are, respectively, the input and output vectors while $\mathbf{n}$ is white Gaussian noise. The channel between the two clusters is represented by the $m \times m$ zero-mean random matrix $\mathbf{H}$, with i.i.d. Gaussian entries, where m is the (common) size of the clusters, while the diagonal matrix $\mathbf{G}$ accounts for the additional randomness due to the intra-cluster information sharing and, in particular, $\mathbf{G}_{k,k} = \sqrt{G}e^{j\theta_k}r_k^{-\alpha/2}$.

The information-theoretic analysis of the above channel subsumes the statistical characterization of the grammian channel matrix $\mathbf{H}\,\mathbf{G}\mathbf{G}^\dagger\,\mathbf{H}^\dagger$, which will be the main subject of the present Section. We will study in particular the positive definite matrix $\mathbf{T} = \mathbf{G}\mathbf{G}^\dagger$, whose main diagonal contains strongly correlated entries, due to the distances ordering. A way to get rid of the correlation and deal with independent entries is to condition the process on having a certain number of nodes in a given interval, or, equivalently, conditioning on the value of $r_{m+1}^\alpha = a$.

With that, following again [28] one can then evaluate the marginal density distribution of an unordered nonzero entry of $\mathbf{T}$ obtaining

$$f_{y_k}^{(a)}(y) = \frac{2}{y\alpha}\left(\frac{1}{ay}\right)^{2/\alpha} \Gamma\left(\beta', \frac{1}{a}\right) \tag{6.8}$$

with $\beta' = \frac{2}{\alpha}+1$ and, as a consequence, the moments of the trace of the matrix related to the spatial randomness. In particular, when the channel between the source and the neighbors is unfaded, one can evaluate for the matrix $\mathbf{T}$ both the Shannon and the η-transforms, whose expressions will be helpful in evaluating the information flow between the two clusters.

6.2.3 Shannon Transform

Following [60, (2.47) and (2.49)], we evaluate the transforms for the law in eq. (6.8), obtaining,

$$\eta_{\mathbf{T}}(\gamma) = \left(\frac{a}{\gamma}\right) \frac{{}_2F_1\left(1, \beta'; \beta'+1; -\frac{a}{\gamma}\right)}{\beta'}, \tag{6.9}$$

with ${}_2F_1$ the Gauss Hypergeometric function [4], and, respectively,

$$\nu_{\mathbf{T}}(\gamma) = \left(\frac{\alpha}{2\gamma}\right)\left[a^{\beta'}\left(\frac{\gamma}{\alpha}\left(\frac{\gamma}{\alpha}\frac{{}_2F_1\left(1, 2-\beta'; 3-\beta'; -\frac{\gamma}{a}\right)}{\frac{2}{\alpha}-1} + \log\left(1+\frac{\gamma}{a}\right)\right) - \left(\frac{\gamma}{a}\right)^{\beta'} \pi \csc\left(\pi\beta'\right)\right)\right]. \tag{6.10}$$

From the expressions above, moreover, we can evaluate the Shannon transform for the compound channel matrix as [60]

$$\nu_{\mathbf{HTH}^\dagger}(\gamma) = \widetilde{\beta}\nu_{\mathbf{T}}(\eta\gamma) + \log\frac{1}{\eta} + (\eta - 1)\log e, \tag{6.11}$$

with η the solution to

$$\widetilde{\beta} = \frac{1-\eta}{1-\eta_{\mathbf{T}}(\gamma\eta)}.$$

6.2.4 Moment Generating Function

A further tool to investigate the finite-size behavior of the Wishart-Poisson model above proposed is the Moment Generating Function (MGF); referring to [60, eqq. (2.20) and (2.21)] we can provide the conditional (with respect to $\mathbf{T}$) MGF expression

$$E\{\det\left(\mathbf{I} + \mathbf{HTH}^\dagger\right)^\zeta | \mathbf{T}\} = {}_2F_0(-\zeta; m; -\mathbf{T}). \tag{6.12}$$

Notice that one among the hypergeometric coefficients is a negative integer, hence the series in (6.12), for integer ζ, turns out to be a polynomial of degree $m\zeta$ [42, p. 258]. Moreover, writing

$${}_2F_0(-\zeta; m; -\mathbf{T}) = \sum_{k=0}^{m} \zeta \sum_{\kappa} \frac{(-1)^k(\zeta)_\kappa(m)_\kappa}{k!}\mathcal{C}_\kappa(-\mathbf{T}),$$

and recalling that zonal polynomials $\mathcal{C}_\kappa$ are symmetric, homogeneous polynomials of degree $|\kappa|$ in the eigenvalues of $\mathbf{T}$, which are i.i.d. distributed with

law (6.8), the unconditional MGF can be obtained through the evaluation of $E[\mathcal{C}_\kappa(-\mathbf{T})]$. The calculus boils down to the evaluation of the moments of the random variables in (6.8). We note explicitly that, while for the asymptotic analysis through Shannon and η-transforms in the absence of fading all the involved parameters can be evaluated in closed form, the MGF only takes finite values when the regularized attenuation function in (6.5) is exploited.

6.3 Scaling laws analysis: SIMO upper bound

Before evaluating the cut-set rate scaling of the network, we evaluate the scaling achievable in the general case by exploiting the underlying spatial random structure of the node process. We remark that after random source-destination matching and the consideration of the actual power regimes experienced by the nodes, such scaling will strongly change, as shown in Section 6.4.

Theorem 4. *The average aggregate throughput in a network with n nodes distributed according to a two-dimensional homogeneous Poisson process is bounded above by*

$$T(n) \leq Kn \log\log n \tag{6.13}$$

Proof. Along the lines of [47], we can upper bound the transmission rate $R(n)$ from a source node s to the destination d by the capacity of the single-input multiple-output (SIMO) channel between the source and the remaining nodes. In our case, specifically, the bound turns out to be, for $\alpha = 2$ and exploiting (6.4),

$$\begin{aligned}\mathbf{E}\left[\log\left(1+\frac{P}{N_0}\sum_{i\neq s}\frac{G}{r_{i,s}^{\alpha}}\right)\right] &\leq \log\left(1+\frac{P}{N_0}\sum_{i\neq s}\mathbf{E}\left[\frac{G}{r_{i,s}^{\alpha}}\right]\right)\\ &\leq \log\left(1+\frac{PG}{N_0}\lambda\pi\sum_{\ell=2}^{n}\frac{1}{\ell-1}\right),\end{aligned}$$

which for large n tends to $R(n) \leq K' \log\log n$ by [4].

We explicitly note that, for $\alpha > 2$, expression (6.4) is singular at least for the closest neighbors. In this case, we resort to the regularized attenuation function with first moment given by (6.6), noticing that

$$\frac{x^n E_{1+\alpha/2-n}(x) + \Gamma(n) - \Gamma(n,x)}{\Gamma(n)} \leq \frac{1}{n-1}$$

and, then, still

$$\mathbf{E}\left[\log\left(1+\frac{P}{N_0}\sum_{i\neq s}\frac{G}{\max\{r,r_0\}_{i,s}^{\alpha}}\right)\right]\leq \tag{6.14}$$
$$\log\left(1+\frac{P}{N_0}\sum_{i\neq s}\mathbf{E}\left[\frac{G}{\max\{r,r_0\}_{i,s}^{\alpha}}\right]\right)=$$
$$\log\left(1+\frac{PG}{2N_0r_0^{\alpha}}\sum_{\ell=2}^{n}\frac{1}{n-1}\right).$$

□

6.4 Cut-set bound

The elegant derivation in [20] aims at obtaining an upper bound to the mutual information between the nodes contained within a circular region of area proportional to the node number and those placed in a circular crown surrounding the first one. The information flow is split into two contributions, a first one concerning the communication between the inner nodes and those placed within an annulus of width δ, and a second one coming from the information transfer to the remaining outer nodes. While the first contribution analysis can be carried out without resorting to electromagnetic theory, a careful analysis of the degrees of freedom dictating the maximum number of available independent channels (and thus, basically, the achievable scaling) requires tools from functional analysis [12, and references therein],[20]. Specifically, the rank of the radiation operator (i.e. of the channel matrix) is evaluated performing a singular value decomposition of the radiated field, which leads to the remarkable result that the effective number of independent MISO channels between the inner nodes and the farthest nodes is of order of $\sqrt{n}$, rather than proportional to n, thus forbidding a linear scaling.

In this section, we aim at providing an upper bound on aggregate rate sustainable by a large random network through some compact information-theoretic performance indexes already proven to be very effective in the design and analysis of single and multi-user wireless communications systems. To bound the achievable information flow we resort to the classical cut-set argument. Our study relies on the high Signal to Noise Ratio (SNR) affine expansion of the mutual information in coherent MIMO channels [59] as to the near-to-cut region, and on the low-power approximation [58], respectively, for the remaining nodes.

Before proceeding further in the analysis, we briefly recall the low and high-power expansions to be exploited in the following, and their expressions as a function of the involved channel matrices.

6.4.1 High and low-power mutual information analysis

In the high-SNR regime, one can expand the mutual information as a function of the SNR[5] as [59]

$$\mathcal{I}(\mathsf{SNR}) = S_\infty \left[\frac{\mathsf{SNR}|_{\mathrm{dB}}}{3\,\mathrm{dB}} - \mathcal{L}_\infty \right] + o(1),$$

where

$$S_\infty = \lim_{\mathsf{SNR}\to\infty} \frac{\mathcal{I}(\mathsf{SNR})}{\log_2 \mathsf{SNR}} \tag{6.15}$$

and $\mathcal{L}_\infty$, whose value is not of direct interest as long as just the scaling is to be investigated, is defined in [59].

In the low-SNR regime instead [61], the key performance measures are $\frac{E_b}{N_0}_{\min}$ (the minimum energy per information bit required to convey any positive rate reliably) and S_0, the capacity slope in bits/s/Hz/(3 dB), such that

$$\mathsf{C}(\tfrac{E_b}{N_0}) = \frac{S_0}{3\,\mathrm{dB}} \left(\frac{E_b}{N_0}|_{\mathrm{dB}} - \frac{E_b}{N_{0\,\min}}|_{\mathrm{dB}} \right) + \epsilon \tag{6.16}$$

with ϵ vanishing faster than the main term as $\frac{E_b}{N_0}$ approaches $\frac{E_b}{N_0}_{\min}$. In the following, we will be mainly concerned with the evaluation of the low-power slope S_0, since $\frac{E_b^r}{N_0}_{\min} = \ln 2$ for AWGN noise, irrespectively on the adopted transmission scheme.

6.4.2 Upper bounding the information flow

We start from a square portion of the plane of linear dimension $\sqrt{n}$, where n nodes are randomly deployed, and consider a cut dividing the area into two equal halves. The maximum achievable sum-rate between the randomly chosen source-destination pairs is bounded above by the capacity of the MIMO channel between nodes S located to the left of the cut and nodes D located to its right. Under fast fading assumption, such sum-rate can be upper bounded by $E[\log\det\left(\mathbf{I}+\mathbf{H}\mathbf{H}^\dagger\right)]$, with $\mathbf{H}$ the channel matrix of the overall MIMO system[6].

[5] In our case of fixed transmit power and thermal noise at the receiver, $\mathsf{SNR} = P_{tx}/N_0$.

[6] We remark that a careful analysis of the network would require to adopt a (conditionally on the fading) euclidean random matrix model for $\mathbf{H}$ [11]. However, such discussion is

We further consider the domain D to be sub-divided into a rectangular strip of area $1 \times \sqrt{n}$ close to the border of the cut, which contains no more than $\sqrt{n}\log n$ nodes [47, Lemma (5.1.a)]. By Hadamard's inequality we can write

$$\log\det\left(\mathbf{I}+\mathbf{H}\mathbf{H}^{\dagger}\right) \leq \\ \log\det\left(\mathbf{I}+\mathbf{H_1}\mathbf{H_1}^{\dagger}\right)+\log\det\left(\mathbf{I}+\mathbf{H_2}\mathbf{H_2}^{\dagger}\right), \tag{6.17}$$

where $\mathbf{H_1}$ is the channel matrix modeling the communication between the nodes on the left of the cut, and those in the little strip, say the region V_D and, in turn, $\mathbf{H_2}$ is the matrix modeling the communication between the nodes on the left and the remaining ones on the right $D-V_D$. Nodes placed in the strip V_D, thus, are experiencing a communication in high-SNR regime, while the remaining nodes communication takes place in the low-power regime. From [59], $S_\infty \leq \min\{n_{tx}, n_{rx}\}$, with n_{tx} the number of transmit antennas and, respectively, n_{rx} the number of receiving antennas. For our channel, $n_{tx} = O(n)$ and $n_{rx} = O(\sqrt{n}\log n)$, $S_\infty \leq \sqrt{n}\log n$, confirming the $\sqrt{n}$ law for the near-to-cut region to which both purely stochastic and purely electromagnetic approach end up [20][47]. As to $\mathcal{L}_\infty$, its value can be approximated by [59, Eq. (15)] (this is tantamount to assume the entries of $\mathbf{H_1}$ be i.i.d.). Finally, by [58, eq. (19)],

$$S_0 = \frac{2n_{eqtx}n_{eqrx}}{n_{eqtx}+n_{eqrx}},$$

where n_{eqtx} and n_{eqrx} are *equivalent* number of transmit and receive antennas, which take into account a reduction in the effective number of spatial degrees of freedom due to phenomena impairing the transmit or, respectively, the receive side of the channel. The work in [20], in particular, has revealed by inspection of the radiation operator that the number of independent channels crossing the cut toward the low-power region[7] is $O(\sqrt{n}\log n)$, and that this phenomenon can be interpreted as an effective reduction of the number of independent receiving nodes within the abovementioned region. A more qualitative analysis [3, Proposition I], comprehensive of both stochastic as well as electromagnetic aspects, still shows such an impact on the spatial degrees of freedom, reducible at the definition of n_{eqtx} and n_{eqrx} on the basis of electromagnetic considerations in the region close to the transmitter and the receiver, separately. Finally, we notice that if a separable correlation model

beyond the scope of the paper. We will, instead, make a simplifying assumption keeping $\mathbf{H}$ zero-mean Gaussian with possibly correlated entries.

[7] The analysis is carried out in a region with circular symmetry, however as far as asymptotic scaling is concerned the conclusions can be transferred also to a square geometry.

can be identified for the channel under consideration, due to [58, Property I], $O(1) \leq n_{eqrx} \leq O(n)$, and thus $1 \leq S_0 \leq n$ (in order sense), encompassing different scenarios, ranging from constant to linear throughput scaling with n, and comprehensive of the $\sqrt{n}$ law.

6.5 Conclusion

A closer look to random spatial features of large wireless networks has led to the formulation of a new (matrix-variate) channel model, whose moments and transforms have been in part evaluated, paving the way to future information-theoretic analysis. The traditional cut-set argument has then been exploited using an affine expansion of the mutual information approach, which shows that the ultimate scaling law strongly depends on the spatial impairments model. Future work will comprehend analysis of some commonly adopted spatial correlation models and the discussion on their (eventual) fitting with physically dictated scaling laws.

Acknowledgment

This work was supported by the STREP projects Mascot and Newcom++ of the European Commission, and by the Austrian Government and the City of Vienna within the competence center programme COMET through ftw project I0. This work was partially funded by the Italian Ministry for University and Research through the PRIN CARTOON 2006 project.

Chapter 7

High and low-SNR regimes for stochastic networks

Giuseppa Alfano, Maxime Guillaud, and Antonia Tulino

A compact model for the evaluation of scaling laws in random wireless networks is proposed and studied. The model allows the information-theoretic characterization of both point-to-point as well as distributed communications. It is analyzed under several assumptions about spatial correlation and the utilized channel state information and transmission schemes.

7.1 Introduction

The promise of linear scaling of the aggregate throughput[1] with the number of users has been for over a decade under the exclusive ownership of multiantenna systems [55] and/or of fully coordinated multiuser schemes [31, and references therein]. Far less promising were the so called *ad hoc* networks, whose seminal investigation [26] led to the conclusion that the sum-rate therein could scale only up to become proportional to the square root of the number of users. More recent studies, however [46], exhibited indeed some transmission protocols able to achieve up to $\mathcal{O}(n^{\frac{h}{h+1}})$ aggregate capacity scaling, where n is the number of users in a two-dimensional, planar network of area proportional to n (*extended network* scenario) and h is any positive integer. It seemed, then, that the loss in co-location and coordination with respect to single-user multiantenna and/or multiuser schemes could be

[1] Throughout this paper, aggregate throughput denotes the maximum achievable sum-rate between randomly chosen source-destination pairs in the network, following the cut-set argument hypotheses in [14, Ch. 14].

reduced or eliminated by paying a proper delay price. These results are, however, in contrast with the physically achievable scaling [19], based on pure geometric and electromagnetic arguments, which turns out to be of order $\sqrt{n}$ in a spatial region of area of $\mathcal{O}(n)$, regardless of the adopted communication protocol.

The gap between the information-theoretic and the electromagnetic laws needs thus to be closed, in order to provide an actual and tight upper bound on the capacity scaling, to serve as a protocol design guideline. The exploitation of the cut-set upper bound [14] for the network under exam offers a supremum to the achievable aggregate rate in terms of the mutual information conveyed by an equivalent Multiple-Input-Multiple-Output (MIMO) channel. A particular care should be devoted then to adequately take into account the several physical impairments [58, and ref. therein][59] already arising in the point-to-point MIMO case, without neglecting the additional randomness given by stochastic nodes locations (see e.g. [27]).

In this contribution, we will extend the approach proposed in [8] to evaluate the scaling law of the aggregate throughput to a spatially correlated channel, and then refine the analysis moving from an equivalent point-to-point to a more effective equivalent broadcast channel. The paper is organized as follows: Section 7.2 recalls the cut-set bound derivation, and contains some further background material on high and low-SNR scaling analysis. The case of spatially correlated point-to-point channel is treated in Section 7.3. Conclusion and hints for future work are contained in the final section.

7.2 Cut-set bound

Let us consider a network whose n nodes are randomly placed in a square region of edge length $\sqrt{n}$. Without loss of generality, we will assume throughout the paper the nodes to be located according to an Homogeneous Poisson Point Process (HPPP) of unit intensity, and, unless otherwise stated, to be single antenna equipped. By considering a cut which divides the area into two equal halves, the maximum achievable sum-rate between randomly chosen source-destination pairs is bounded above by the capacity of the single-user MIMO channel between nodes S located to the left of the cut and nodes D located to its right [14, Ch. 14]. Notice that, typically, there will be approximately half of the nodes in the region S and the remaining ones in the region D. Under fast fading assumption, such sum-rate can be upper bounded by $E[\log\det\left(\mathbf{I}+\mathbf{H}\mathbf{H}^{\dagger}\right)]$, with $\mathbf{H}$ the channel matrix of the overall MIMO system, and $\mathbf{I}$ the identity matrix. According to [46][19], we further consider the domain D to be sub-divided into a rectangular strip,

say V_D, of area $1 \times \sqrt{n}$ close to the border of the cut, which contains no more than $\sqrt{n}\log n$ nodes [46, Lemma (5.1.a)], and a remaining region to which we refer as $D - V_D$. The idea behind this separation is that, due to the randomness of the node positions, some nodes in V_D can be arbitrarily close to the boundary with S, and therefore (under the usual assumption of a path-loss geometric in the distance between transmitter and receiver) enjoy an arbitrarily high signal power, for which a separate analysis is required. By Hadamard's inequality, we can write

$$\log\det\left(\mathbf{I}+\mathbf{H}\mathbf{H}^{\dagger}\right) \leq \log\det\left(\mathbf{I}+\mathbf{H_1}\mathbf{H_1}^{\dagger}\right) + \log\det\left(\mathbf{I}+\mathbf{H_2}\mathbf{H_2}^{\dagger}\right) \quad (7.1)$$

where $\mathbf{H_1}$ is the channel matrix modeling the communication between the nodes on the left of the cut and those in the little strip, and, in turn, $\mathbf{H_2}$ is the matrix modeling the communication between the nodes on the left and those in $D - V_D$. Nodes placed in the strip V_D, thus, are experiencing a communication in high-SNR regime, while the remaining nodes communication takes place in the low-power regime. Following [59][8], $S_\infty \leq \sqrt{n}\log n$, where S_∞ denotes the sum-rate slope in the high-SNR region. As to the low-power slope[2], since its evaluation involves the correlation structure of the receive and/or transmit side, as well as the channel condition, the adopted protocol and the amount of channel state information available at both receive and transmit ends, we will make some further assumptions on the system, and detail the results in the following Sections.

7.3 Spatially correlated MIMO

The low-power capacity slope for a Rayleigh fading MIMO channel can be written as [58]

$$S_0 = \frac{2n_{\mathrm{T}}^{\mathrm{eq}} n_{\mathrm{R}}^{\mathrm{eq}}}{n_{\mathrm{T}}^{\mathrm{eq}} + n_{\mathrm{R}}^{\mathrm{eq}}},$$

where $n_{\mathrm{T}}^{\mathrm{eq}}$ and $n_{\mathrm{R}}^{\mathrm{eq}}$ are *equivalent* number of transmit and receive antennas, which take into account a reduction in the effective number of spatial degrees of freedom due to phenomena impairing the transmit or, respectively, the receive side of the channel.

In [8] indeed, it is shown that, under the hypothesis of independent signalling, $1 \leq S_0 \leq n$ (in order sense), encompassing both linear (as a limiting case) as well as $\sqrt{n}$ law as a particular case. Linear scaling could be, in principle, achieved only in case $n_{\mathrm{T}}^{\mathrm{eq}}$ and $n_{\mathrm{R}}^{\mathrm{eq}}$ are of the same order of

[2] We refer the reader to [58, Sec. I.B and III] for the introduction and an extensive discussion on the low-power capacity slope in MIMO channels.

magnitude. This is in contradiction with the electromagnetic argument of counting independent channels in the low-power region [19], and the aim of this Section is to detail a framework where stochastic and electromagnetic arguments offer closer results.

A simplistic, though suitable, way to enhance channel non-ideality effects is to express the random channel matrix following the Kronecker or *separable correlation* model, i.e.

$$\mathbf{H}_2 = \mathbf{\Theta}_{\mathrm{R}}^{1/2}\mathbf{W}\mathbf{\Theta}_{\mathrm{T}}^{1/2} \tag{7.2}$$

where $\mathbf{W}$ is composed of independent unit-variance complex Gaussian random entries. This model implies that the covariance matrix of each row of $\mathbf{H}$ is thus given by $\mathbf{\Theta}_{\mathrm{T}}$ while the covariance matrix of each column is given by $\mathbf{\Theta}_{\mathrm{R}}$. From a physical point of view, in single-user MIMO systems this means that the immediate surroundings to each array are responsible for the correlation between its antennas but have no impact on the correlation between the antennas at the other end of the link [40][22]. Such a model could be questionable in the case of spatially distributed nodes, due to the further geometric aspects to be managed in order to verify whether the spatial distribution of the nodes in a region of the network does impact on the correlation structure on the other side, but we choose as a starting point of the analysis, and for ease of exposition, to keep it as far as no multiple-polarizations and/or other signalling strategies, usually known as breaking the Kronecker model, are not exploited.

In terms of the spatial correlation matrices, the equivalent number of antennas can be expressed as [58, Corollary 2]

$$n_{\mathrm{T}}^{\mathrm{eq}} = \frac{n_{\mathrm{T}}}{\zeta(\mathbf{\Theta}_{\mathrm{T}})} \qquad n_{\mathrm{R}}^{\mathrm{eq}} = \frac{n_{\mathrm{R}}}{\zeta(\mathbf{\Theta}_{\mathrm{R}})}. \tag{7.3}$$

with n_{T} and n_{R} the actual number of transmit (respectively, receive) antennas, and the function $\zeta(\cdot)$, often referred to as the *dispersion* of the (random) square matrix of dimension m in the argument, defined as [58, Eq.(8)],

$$\zeta(\mathbf{A}) = m\frac{E[\mathrm{Tr}\{\mathbf{A}^2\}]}{E^2[\mathrm{Tr}\{\mathbf{A}\}]}. \tag{7.4}$$

Notice that, since a correlation matrix has unit diagonal elements, expression (7.4) particularizes to

$$\zeta(\mathbf{\Theta}) = \frac{E[\mathrm{Tr}\{\mathbf{\Theta}^2\}]}{n}. \tag{7.5}$$

which is usually referred to as *correlation number*, and whose value is bounded, being

$$1 \leq \zeta(\mathbf{\Theta}) \leq n \tag{7.6}$$

with the lower bound achieved if and only if the sensor nodes are uncorrelated and the upper bound achieved if and only if they are fully correlated.

The evaluation of S_0 in our framework essentially boils down to the evaluation of $E[\mathrm{Tr}\{\mathbf{\Theta}^2\}]$ for $\mathbf{\Theta}$ a suitable (random) spatial correlation matrix in a wireless network.

As a first step to deeper understanding the random network behavior, we analyze some among the most commonly adopted spatial correlation models, namely

- *Constant correlation model* [52, Example IV]

  Given $\rho \in (0,1)$, $\mathbf{\Theta}_{\mathrm{R}i,j}(\rho) = \rho \, \forall i,j$ such that $i \neq j$.

- *Jakes' model*[33]

  $\mathbf{\Theta}_{\mathrm{R}i,j} = \mathcal{J}_0\left(\frac{2\pi}{\lambda} d_{i,j}\right)$, with $d_{i,j}$ the inter-node distances for an HPPP and λ the carrier wavelength.

- *Random Toeplitz*

  The n elements of the first row of $\mathbf{\Theta}_{\mathrm{R}}$, say $\vartheta_{1,j}$, $j = 1, \ldots, n$ are independent random variables which form a Toeplitz matrix.

For simplicity, and in order to allow comparison with previous results such as [46][19], we make the optimistic assumption w.r.t. transmit correlation that $\mathbf{\Theta}_{\mathrm{T}} = \mathbf{I}$, hence $\zeta(\mathbf{\Theta}_{\mathrm{T}}) = 1$ and $n_T^{eq} = n_T$.

7.3.1 Spatial correlation: analysis

Proposition 1. *The correlation number of a constant correlation matrix* $\mathbf{\Theta}_{\mathrm{R}}(\rho)$ *can be written as* $\zeta(\mathbf{\Theta}_{\mathrm{R}}(\rho)) = 1 + (n-1)\rho^2$.

Proposition 2. *The correlation number for the random Jakes' model when nodes locations follow a HPPP pattern can be written as*

$$\zeta(\mathbf{\Theta}_{\mathrm{R}}) = 1 + \sum_{\ell=2}^{n} \binom{n}{\ell} \sum_{t=0}^{n-\ell} \frac{(-1)^t}{\ell + t} \binom{n-\ell}{t} \cdot {}_2F_3\left(\ell + t, 1/2; 1, 1, \ell + t + 1; -\left(\frac{2\pi\sqrt{n}}{\lambda}\right)^2\right), \tag{7.7}$$

where ${}_pF_q$ *is the generalized hypergeometric function [24].*
Proof: *See Appendix.*

Proposition 3. *The correlation number when the entries of* $\mathbf{\Theta}_{\mathrm{R}}$ *are independent, nonzero mean*[3]*, subgaussian random variables*[4]*, with common mean* μ*, forming a Toeplitz matrix, can be lower bounded by*

$$\zeta(\mathbf{\Theta}_{\mathrm{R}}) \geq \mu.$$

Proof: *See Appendix.*

7.3.2 Spatial correlation: discussion

The listed results are worth some comments; notice that, since the number of nodes in S and that in $D - V_D$ are of the same order, one can write as for the single-user MIMO channel with both receiver and transmitter equipped with the same number of antennas [58, Corollary 3]

$$S_0 = \frac{2\,n}{1 + \zeta(\mathbf{\Theta}_{\mathrm{R}})}. \tag{7.8}$$

Particularizing the above equation to the constant correlation model, it is evident that it offers a very poor low power scaling, namely a per-node decay of the throughput as fast as $O(n^{-1})$. Constant correlation suitably models a scenario where receive nodes are so closely spaced as to not to experience any anisotropy in the received signal behavior. One should then be very careful in stating that colocated nodes are most likely to well cooperate and offer higher throughput, since also such spatial impairments, like in the single-user MIMO case, have to be taken into account.

Things strongly change when putting into (7.8) the Jakes based correlation number. Indeed, one can numerically verify that the summation of hypergeometric functions does not cause a substantial increase of the correlation number. However, such a parameter has been evaluated under overoptimistic hypotheses, namely assuming that inter-nodes distances are scattered as randomly as a HPPP dictates, and that the sensors are omnidirectional. A numerical investigation of a generalized Jakes model, keeping fixed the angular beam of interest for the received signal, and relying on

[3]We cannot assume the mean of all elements to be zero since at least for the diagonal elements we should ensure them to have unit mean, in order to comply with the definition of correlation number.

[4]A random variable X is called *subgaussian* if

$$\mathbf{P}[|X| \leq t] \leq 2e^{-at^2} \qquad \forall t \geq 0$$

and for some constant $a > 0$.

distance statistics tailored on a non-isotropic node radiation pattern (see e.g.[28, Corollary 3]) would offer more significant results.

The Toeplitz scenario is the more challenging from a mathematical point of view. Indeed, the given bound is trivial when $\mu = 1$, which indeed should be at least on the main diagonal. Consider, however, that an actual evaluation of the correlation number would involve nonlinear statistics of the noncentral Toeplitz matrix entries, while we are herein exploiting results on the spectral norm, i.e. the square of the maximum eigenvalue of the matrix under exam. Notice further that the given result is based on approximating the elements of the random correlation matrix to have all the same first moment, which clearly does not apply to a real world scenario.

The *brute-force* generalization (and, even, adoption without any adaptation) of the correlation models of the point-to-point scenario is questionable, tough, it suggests further lines of investigations. Mostly, the question arises of how to refine the stochastic channel model in order to embody the system geometry, while keeping it still analytically tractable.

While a good alternative to separable correlation models, actually subject of ongoing work, is the adoption of Vandermonde-like channel matrices ([43][45]) in order to accounting for the randomness of phases relationships between the receive sensors, we believe, nevertheless, that the inherent broadcast setting of the cut-set strategy in our framework (a single multiple-antenna user, whose role is played by the nodes in S, which sends data to two users in very different SNR regimes, namely the nodes in V_D and in $D - V_D$) could be also characterized resorting to tools indeed developed in the broadcast framework.

7.4 Conclusion

Upper bounds for the scaling laws of the aggregate throughput in random networks are provided in a compact way, using well assessed and effective tools from multiple-antenna literature. Both point-to-point as well as broadcast frameworks are exploited, leading to the remarkable result that the overall scaling, up to first order analysis, is strongly limited by the information flow among nodes experiencing communications in the low-power regime.

APPENDIX

A: Proof of Proposition 2

The assumed system geometry implies that $\mathbf{\Theta}_{\mathrm{R}i,j} = \mathcal{J}_0\left(\frac{2\pi}{\lambda}d_{i,j}\right)$ with $d_{i,j}$ the random distance between the i-th and the j-th node in $D - V_D$. The k-th diagonal entry of $\mathbf{\Theta}_{\mathrm{R}}$, since $d_{i,j} = d_{j,i}$, can be written as $\mathbf{\Theta}^2_{\mathrm{R}k,k} = 1 + \sum_{\ell\neq k}^{n} \mathcal{J}_0^2\left(\frac{2\pi}{\lambda}d_{\ell,k}\right)$, from which

$$E[\mathrm{Tr}\{\mathbf{\Theta}_{\mathrm{R}}\}^2] = n + \sum_{k=1}^{n}\sum_{\ell\neq k}^{n} E[\mathcal{J}_0^2\left(\frac{2\pi}{\lambda}d_{\ell,k}\right)].$$

Distance statistics of random pairs in a network are to the best of the authors knowledge not available in handy form; rather simplified expressions have been provided instead in very recent years as long as ordered distances in HPPP are concerned. Since the value of the mean is not altered by considering for each k the sum ranging over the ordered distances from node k, we can write the probability density function of the ℓ-th ordered distance from node k as in [53, Th. 2.1]

$$f_{d_\ell}(r) = \frac{2}{L}\binom{n}{\ell}\left(\frac{r}{L}\right)^{2\ell-1}\sum_{t=0}^{n-\ell}(-1)^t\binom{n-\ell}{t}\left(\frac{r}{L}\right)^{2t} \qquad r\in(0,L)$$

where L is the typical linear dimension of the finite region where nodes are placed and herein is of the order of $\sqrt{n}$. Since the $n-1$ ordered distances are identically distributed whatever choice of the origin node is made due to the homogeneity assumption, we can then sum up over k and write

$$\zeta(\mathbf{\Theta}_{\mathrm{R}}) = 1 + 2\binom{n}{\ell}\sum_{\ell\neq k}^{n}\sum_{t=0}^{n-\ell}(-1)^t\binom{n-\ell}{t} \int_0^L \left(\frac{r}{L}\right)^{2\ell+2t-1}\mathcal{J}_0^2\left(\frac{2\pi}{\lambda}r\right)\frac{dr}{L}, \tag{7.9}$$

from which (7.7) follows.

B: Proof of Proposition 3

Follows from $E[\mathrm{Tr}\{\mathbf{\Theta}^2_{\mathrm{R}}\}] \geq \lambda^2_{max}$, with λ_{max} the maximum eigenvalue of $\mathbf{\Theta}_{\mathrm{R}}$, and from [39, Eq.(8)].

Bibliography

[1] http://www.research.att.com/∼njas/sequences/A001147. 35

[2] http://www.research.att.com/∼njas/sequences/A000085. 37

[3] T.D. Abhayapala, T.S. Pollock, and R.A. Kennedy. Spatial decomposition of the MIMO wireless channels. In *Proc. of 7th ISSPA*, Paris, F, July 2003. 80

[4] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover Publications, New York, 1995. 76, 77

[5] W. Ajib and D. Hoccoun. An overview of scheduling algorithms in mimo-based fourth-generation wireless systems. *IEEE Network*, Sep.-Oct. 2005. 33

[6] Cemal Akçaba, Patrick Kuppinger, and Helmut Bölcskei. Distributed transmit diversity in relay networks. In *IEEE Information Theory Workshop (ITW)*, pages 233–237, July 2007. 63

[7] Cemal Akçaba and H. Bölcskei. Distributed transmit diversity in relay networks. *To be decided*, 2008. 56

[8] Giuseppa Alfano, Maxime Guillaud, and Antonia Tulino. Scaling laws for large ad-hoc wireless networks with wishart-poisson fading. In *Proc. International Symposium on Spread Spectrum Techniques and Applications (ISSSTA)*, Bologna, Italy, August 2008. 43, 48, 83, 84

[9] B. C. Banister and J. R. Zeidler. A simple gradient sign algorithm for transmit antenna weight adaptation with feedback. *IEEE Trans. Signal Processing*, 51(5):1156–1171, May 2003. 9, 11, 14, 23, 25

[10] H. Bölcskei, R. U. Nabar, Ö. Oyman, and A. J. Paulraj. Capacity scaling laws in MIMO relay networks. *IEEE Trans. Wireless Comm.*, 5(6):1433–1444, Jun. 2006. 63

[11] Charles Bordenave. Eigenvalues of Euclidean random matrices, 2006. 73, 79

[12] Ovidio M. Bucci and Giorgio Franceschetti. On the degrees of freedom of scattered fields. *IEEE Trans. on Antennas and Prop.*, 37(7):918–926, Jul. 1989. 71, 78

[13] Pedro Coronel and H. Bölcskei. Diversity-multiplexing tradeoff in selective-fading MIMO channels. In *Proc. IEEE ISIT*, Nice, France, June 2007. 65

[14] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 1991. 82, 83

[15] A. F. Dana and B. Hassibi. On the power efficiency of sensory and ad-hoc wireless networks. In *Proc. IEEE ISIT*, page 412, Yokohama, Japan, June/July 2003. 8, 15

[16] J.D. Dixon and B. Mortimer. *Permutation Groups*. Number 163 in Graduate Texts in Mathematics. Springer-Verlag, 1996. 38, 48

[17] O. Dousse, F. Baccelli, and P. Thiran. Impact of interferences on connectivity in ad–hoc networks. *IEEE/ACM Trans. Networking*, 13(2):425–436, Jul 2005. 74

[18] P. Fertl, A. Hottinen, and G. Matz. Perturbation-based distributed beamforming for wireless relay networks. In *Proc. IEEE GLOBECOM-2008*, New Orleans, USA, Nov./Dec. 2008. 23, 25, 29

[19] M. Franceschetti, P. Minero, and M. D. Migliore. The capacity of wireless networks: Information-theoretic and physical limits. In *Proc. Allerton Conference*, 2007. 83, 85, 86

[20] Massimo Franceschetti, Paolo Minero, and Marco D. Migliore. The capacity of wireless networks: Information-theoretic and physical limits. In *Proc. of 45-th Allerton Conference*, Urbana-Champaign, IL, 2007. 71, 78, 80

[21] Markus Gärtner and Helmut Bölcskei. Multiuser space-time/frequency code design. In *Proc. IEEE Int. Symposium on Information Theory (ISIT)*, pages 2819–2823, Jul. 2006. 68

[22] D. Gesbert, H. Bolcskei, D. Gore, and A. Paulraj. Outdoor MIMO wireless channels: Models and performance prediction. *IEEE Trans. on Communications*, December 2002. 85

[23] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, 3rd edition, 1996. 23, 27

[24] I.S. Gradshteyn and I.M. Ryzhik. *Table of Integrals, Series and Products*. 7th edition, 2007. 86

[25] Larry C. Grove. *Classical Groups and Geometric Algebra*, volume 39 of *Graduate studies in mathematics*. American Mathematical Society, Providence, Rhode Island, 2002. 27

[26] P. Gupta and P. R. Kumar. The capacity of wireless networks. *IEEE Trans. on Inf. Th.*, February 2000. 70, 82

[27] M. Haenggi. A geometric interpretation of fading in wireless networks: Theory and applications. *IEEE Trans. on Inf. Th.* submitted. 83

[28] M. Haenggi. On distances in uniformly random networks. *IEEE Trans. on Information Theory*, October 2005. 71, 72, 73, 74, 75, 88

[29] I. Hammerström, M. Kuhn, and A. Wittneben. Impact of relay gain allocation on the performance of cooperative diversity networks. In *Proc. IEEE VTC-2004 (fall)*, pages 1815–1819, Los Angeles, USA, Sept. 2004. 8, 16, 22, 25

[30] T. Heikkinen and A. Hottinen. Delay-differentiated scheduling in a fading channel. *IEEE Tr. Wireless Comm.*, 7(3):848–856, 2008. 43

[31] B. Hochwald and S. Vishwanath. Space-time multiple access: linear growth in the sumrate. In *Proc. Allerton Conference*, 2002. 82

[32] A. Hottinen and T. Heikkinen. Delay-differentiated scheduling in a randomized MIMO relay-network. In *European Signal Processing Conference (EUSIPCO)*, Poznán, Poland, September 3–7 2007. 43

[33] W. C. Jakes. *Microwave Mobile Communications*. IEEE Press, New York, 1974. 86

[34] Y. Jing and H. Jafarkhani. Network beamforming using relays with perfect channel information. In *Proc. IEEE ICASSP-2007*, pages III–473–III–476, 2007. 9, 22

[35] H.W. Kuhn. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2:83–97, 1955. 34, 39, 45, 46

[36] H.W. Kuhn. Variants of the hungarian method for assignment problems. *Naval Research Logistics Quarterly*, 3:253–258, 1956. 34, 39

[37] P. Larsson. Large-scale cooperative relaying network with optimal coherent combining under aggregate relay power constraints. In *Proc. Future Telecommunication Conference (FTC)*, 2003. 8, 16, 22, 25

[38] C. Li and X. Wang. Cooperative multibeamforming in ad hoc networks. *EURASIP Journal on Advances in Signal Processing*, 2008(310247), 2008. 9

[39] M. W. Meckes. On the spectral norm of a random Toeplitz matrix. *Electron. Commun. Probab.*, December 2007. 89

[40] A. L. Moustakas, H. U. Baranger, A. M. Sengupta, L. Balents, and S. H. Simon. Communication through diffusive media: Coherence and capacity. *Science*, January 2000. 85

[41] R. Mudumbai, J. Hespanha, U. Madhow, and G. Barriac. Distributed transmit beamforming using feedback control. *Arxiv preprint cs.IT/0603072*, 2006. 9, 11

[42] R. J. Muirhead. *Aspects of Multivariate Statistical Theory*. Wiley, 1982. 76

[43] R. Muller. A random matrix model of communication via antenna arrays. *IEEE Trans. on Inform. Theory*, September 2002. 88

[44] J. Munkres. Algorithms for the assignment and transportation problems. *Journal of the Society of Industrial and Applied Mathematics*, 5(1):32–38, 1957. 34, 39

[45] M. Debbah Ø. Ryan. Random vandermonde matrices - part I: Fundamental results. http://arxiv.org/abs/0802.3570v1. 88

[46] A. Özgür, O. Lévêque, and D. Tse. Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks. *IEEE Trans. on Inf. Th.*, October 2007. 82, 83, 84, 86

[47] Ayfer Özgür, Olivier Lévêque, and David Tse. Exact Capacity Scaling of Extended Wireless Networks. In *2007 IEEE Int. Symp. on Inf. Theory*, 2007. 71, 72, 73, 77, 80

[48] B. Raghothaman. Deterministic perturbation gradient approximation for transmission subspace tracking in FDD-CDMA. In *Proc. IEEE ICC-2003*, volume 4, pages 2450–2454, May 2003. 9, 11, 14, 23, 25

[49] C. Rao and B. Hassibi. Diversity-multiplexing gain trade-off of a MIMO system with relays. In *Proc. of the 2007 IEEE Information Theory Workshop on Information Theory for Wireless Networks*, pages 47–52, 2007. 56, 57

[50] A. Roumy and D. Gesbert. Optimal matching in wireless sensor networks. *IEEE Journal on Selected Topics in Signal Processing*, 1(4):725–735, December 2007. 43

[51] Aeron S. and Saligrama V. Wireless ad-hoc networks: Strategies and scaling laws for the fixed SNR regime. *IEEE Trans. on Inf. Th.*, 53(6), 2007. 71

[52] Hyundong Shin and Jae Hong Lee. Capacity of multiple-antenna fading channels: Spatial fading correlation, double scattering and keyhole. *IEEE Trans. on Inform. Theory*, 49(10):2636–2647, October 2003. 86

[53] S. Srinivasa and M. Haenggi. Distance distributions in finite uniformly random networks: Theory and applications. *IEEE Transactions on Vehicular Technology*, 2008. Submitted. 89

[54] S. Tavildar and P. Viswanath. Approximately universal codes over slow-fading channels. *IEEE Trans. Inf. Theory*, 52(7):3233–3258, Jul. 2006. 57, 58, 63, 65

[55] İ.E. Telatar. Capacity of multi-antenna gaussian channels. *Eur. Trans. Telecom.*, Nov.-Dec. 1999. 82

[56] J. Thukral and H. Bölcskei. Distributed spatial multiplexing with 1-bit feedback. In *Proc. 45th Allerton Conference on Communication, Control, and Computing*, Sept. 2007. 9

[57] D. N. C. Tse, P. Viswanath, and L. Zheng. Diversity-multiplexing tradeoff in multiple access channels. *IEEE Trans. Inf. Theory*, 50(9):1859–1874, Sep. 2004. 64

[58] A. Tulino, A. Lozano, and S. Verdú. Multiple antenna capacity in the low-power regime. 49(10):1019–1030, October 2003. 78, 80, 81, 83, 84, 85, 87

[59] A. Tulino, A. Lozano, and S. Verdú. High-SNR power offset of multi-antenna communication. 51(12):1019–1030, December 2005. 78, 79, 80, 83, 84

[60] A. M. Tulino and S. Verdú. Random matrices and wireless communications. *Foundations and Trends in Communications and Information Theory*, 1(1), June 2004. 76

[61] S. Verdú. Spectral efficiency in the wideband regime. *IEEE Trans. Inform. Theory*, 48(6):1319 –1343, 2002. 79

[62] Jack E. Volder. The CORDIC trigonometric computing technique. *IRE Trans. Electron. Comput.*, EC-8:330–334, Sept. 1959. 28

[63] Armin Wittneben and Boris Rankov. Impact of cooperative relays on the capacity of rank-deficient MIMO channels. In *Proceedings of the 12th IST Summit on Mobile and Wireless Communications*, pages 421–425, June 2003. 44

[64] L. Zheng and D. N. C. Tse. Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels. *IEEE Trans. Inf. Theory*, 49(5):1073–1096, May 2003. 56, 63