CORRELATION DIVERSITY IN MIMO BROADCAST AND COHERENCE DIVERSITY IN THE MIMO RELAY CHANNEL

by

Fan Zhang

APPROVED BY SUPERVISORY COMMITTEE:

Aria Nosratinia, Chair

Naofal Al-Dhahir

Hlaing Minn

Mohammad Saquib

To my family and friends.

CORRELATION DIVERSITY IN MIMO BROADCAST AND COHERENCE DIVERSITY IN THE MIMO RELAY CHANNEL

by

FAN ZHANG, BS, MS

DISSERTATION

Presented to the Faculty of

The University of Texas at Dallas

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY IN

ELECTRICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT DALLAS

August 2021

ACKNOWLEDGMENTS

It has been my great pleasure to have been one of the graduate students of Dr. Aria Nosratinia. This dissertation would have never been completed without his knowledge and support. He not only is my academic advisor but also gives me tremendous help in my personal life. I also appreciate the valuable remarks from my committee members, Dr. Naofal Al-Dhahir, Dr. Hlaing Minn, and Dr. Mohammad Saquib. I gratefully acknowledge the National Science Foundation that made my PhD work possible.

I am thankful to my colleagues and friends who always encouraged and supported me during my studies: Ahmed Hindy, Mohamed Fadel, Ahmed Abotabl, Hussein Saad, Noha Hehal, Hassan Zivari-Fard, Heping Wan, Mohammad Esmaeili, Mehdi Karbalayghareh, Pinky Kapoor and Jiawei Gu. I have learned a lot from our valuable discussions and interactions. I want to express my gratitude to my aunt's family in Houston. Despite the long distance, they always give me timely help.

Last but not least, I am grateful to my family: my grandfather, grandmother, father, and mother, who always give me the strength and patience to complete my studies. Their love always sustained me.

June 2021

CORRELATION DIVERSITY IN MIMO BROADCAST AND COHERENCE DIVERSITY IN THE MIMO RELAY CHANNEL

Fan Zhang, PhD
The University of Texas at Dallas, 2021

Supervising Professor: Aria Nosratinia, Chair

The links in wireless networks may have non-identical spatial correlation or coherence times when nodes have different scatters around them or they have unequal mobility. This phenomenon can occur in massive MIMO and high-mobility scenarios. This dissertation investigates the multiuser broadcast channel when the links have non-identical spatial correlations and relay channel with unequal coherence times. It is found that exploiting the disparity in these different fading conditions for multiple users can lead to gains over techniques that do not take advantage of this disparity.

For the MIMO broadcast channel with non-identical spatial transmit correlation, we broaden the scope of transmit correlation diversity to the case of partially and fully overlapping eigenspaces and introduce techniques to harvest these generalized gains. We derive achievable degrees of freedom regions and achievable rate regions and then extend the degrees of freedom results to the K-user case by analyzing the interference graph that characterizes the overlapping structure of the eigenspaces.

For the massive MIMO experiencing different spatial transmit correlation, we propose a strategy combining product superposition and beamforming that applies to any configuration of transmit correlation eigenspaces, leveraging of the statistical characteristics of massive MIMO channels to reconcile the incompatibility of product superposition and beamforming.

To demonstrate the characteristics of this technique, we calculate the sum rate of K-user downlink under two correlation models.

For the MIMO relay with non-identical link coherence times, we calculate the achievable degrees of freedom under this condition. Product superposition technique is employed at the source which allows a more efficient usage of degrees of freedom when the relay and the destination have different training requirements. Analysis is provided and varying configurations of coherence times are studied, including unaligned coherence blocks and arbitrary length of coherence times.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Correlation and Coherence Diversity

In wireless networks, variations in node mobility and scattering environment lead to non-identical delay, Doppler spread, and correlation conditions. Different links may experience widely varying coherence conditions [1, 2, 3]. This dissertation focuses on the non-identical fading conditions in space and time domain.

The effect of spatial correlation on the capacity of MIMO links has been a subject of long-standing interest. Spatial correlation arises in part from propagation environments producing stronger signal gains in some spatial directions than others, and in part from spatially dependent patterns of the antennas. The interest in spatial correlation was sharpened by its experimental validation [4, 5], and more recently by the increasing attention to higher microwave frequencies and larger number of antennas. Shiu et al. [6] proposed an abstract "one-ring" model for the spatial fading correlation and its effect on the MIMO capacity. In single-user channels with channel state information at the receiver (CSIR) but no channel state information at the transmitter (CSIT), channel correlation can boost power but may reduce the degrees of freedom (DoF) [7, 8], thus it can be detrimental at high signal-to-noise ratio (SNR) but a boon at low SNR. Tulino et al. [9] derived analytical characterizations of the capacity of correlated MIMO channels for the large antenna array regime. Chang et al. [10] showed that channel rank deficiency due to spatial correlation lowers the diversity-multiplexing tradeoff curves from that of uncorrelated channel. Capacity bounds subject to channel estimation errors in correlated fading have been characterized [11, 12]. For multiuser systems, at higher frequencies or with large number of antennas, when spatial correlation is unavoidable, comparing capacity against a hypothetically uncorrelated channel may have limited operational impact. Instead, a more immediate question could be: how to maximize performance in the presence of spatial correlation? A useful tool for that purpose is *transmit correlation diversity*, i.e., leveraging the difference between the spatial correlations observed by different users in the system.

There has been a surge of interest in high-mobility wireless communications [13, 14, 15], wherein the co-existence of low-mobility and high-mobility users has been an accepted fact [16]. Naturally, faster nodes lead to links with shorter coherence intervals, and slower nodes experience links with longer coherence intervals. A deeper understanding of relay performance under unequal link coherence times provides new tools and techniques for high-mobility wireless communications. Relaying in high-mobility scenarios has been acknowledged as an important topic [17, 18, 19, 20, 21], but the implications of relaying under unequal coherence intervals is an open problem.

In the dissertation, correlation and coherence disparity in wireless communication networks are studied. It is found that when there are non-identical coherence conditions, gains exist, and these gains are obtained via product superposition and rate splitting.

1.2 Related Works

A review of the literature is as follows. Under the assumption that all users experience *identical* correlation, Al-Naffouri et al. [22] showed that correlation is detrimental to the sum-rate scaling of the MIMO broadcast channels under certain transmission schemes. It is also known that transmit correlation benefits the sum rate in the downlink performance of a heterogeneous cellular network (HetNet) where both macro and small cells share the same spectrum [23]. In practice, however, users may have non-identical correlation matrices because they are not co-located [24], making it difficult to draw conclusions based on [22]. For non-identical spatial correlation, a joint spatial division multiplexing (JSDM) transmission scheme was proposed [25, 26, 27, 28] that reduces the overhead needed for channel estimation. For multiuser networks with orthogonal eigenspace correlation matrices, Adhikary and Caire [29] showed that transmit correlation helps in multi-cell network by partitioning the user spaces into clusters according to correlation. Non-overlapping transmit correlation eigenspaces has also been exploited in a two-tier system where a large number of small cells are deployed under a macro cell [30]. The sum-rate under user-specific transmit correlations with CSIR was studied in [31, 32]. Broadcast channel with spatial correlation was studied in [33, 34, 35] under CSI and feedback. Jiang et al. [36] proposed a scheme for massive MIMO in which the users received the same number of pilots, optimized according to a mutual information metric. In several works [37, 38, 39], the gains in pilots and training were pushed beyond finding pairs of users with non-overlapping transmit-side correlations.

The performance of fading relay channel under *equal* coherence intervals has been extensively studied [40, 41, 42, 43, 44, 45, 46, 47, 48, 49]. Networks with non-identical coherence times have been studied for several scenarios. Broadcast channels with various configurations of coherence times are investigated in [50, 51, 52]. The effect of coherence disparity is also considered in multiple access channel is also considered in [52]. The impact of hybrid channel state information on MISO broadcast channel with unequal coherence times was studied in [53]. For OFDM systems, when the coherence bandwidth varies in different users, the performance is calculated in [54].

1.3 Contributions

Chapter 2 derives the achievable DoF regions for the two-user broadcast channel in spatially correlated fading under the CSIR and no free CSIR assumptions. We analyze the two-user broadcast channel with spatial correlation. We propose an achievable rate region for arbitrary input distributions satisfying the power constraint. We characterize this rate region with an explicit input distribution based on orthogonal pilots and Gaussian data symbols. We also derive the rate achieved with product superposition and a hybrid of pre-beamforming and product superposition. As a by-product, we find the rate achieved with pilot-based schemes for the point-to-point channel, which generalizes the result of Hassibi and Hochwald [55] to correlated fading. We derive achievable DoF regions for the K-user broadcast channel in spatially correlated fading in the presence of CSIR. The results of this chapter were published in [37, 38].

Chapter 3 analyzes the sum rate of a massive MIMO system operating in FDD mode by investigating the pilot reduction and opportunistic additional data transmission that is made possible by spatial correlation. For the achievability results above, we employ pre-beamforming, product superposition, or a combination thereof, in the process demonstrating that these transmission techniques can harvest transmit correlation diversity gains under partially-overlapping eigenspaces.

Chapter 4 analyzes the MIMO relay with coherence diversity. It is assumed that there is no free channel state information (CSI) at the receiver, since unequal coherence times impact channel training and assuming free CSI will distort and obscure important features of the problem. In addition, no channel state information is assumed at transmitters. We propose a product superposition transmission strategy at the source, which was first introduced in [50] for two-user broadcast channels. Product superposition is a technique that allows efficient utilization of channel degrees of freedom under coherence disparity. It is used when the links from the source to the relay and the destination have unequal coherence times. We begin by proving that under identical coherence times, the relay cannot provide any DoF gains over the direct link alone. This result is used as a reference. When the coherence times are unequal, we start with a representative example to show the disparity in coherence times enables DoF gains over conventional transmission. Then we extend the result to the case where the coherence blocks are not aligned and then show that the DoF region is not changed compared to the aligned coherence blocks. Finally, an achievable DoF strategy is proposed where the coherence times are arbitrary times. The results of this chapter were published in [56, 57].

1.4 Notation

Bold lower-case letters, e.g. $\mathbf{x}$, denote column vectors. Bold upper-case letters, e.g. $\mathbf{M}$, denote matrices. The Euclidean norm is denoted by $\|\mathbf{x}\|$ and the Frobenius norm $\|\mathbf{M}\|_F$. The trace, conjugate, transpose and conjugated transpose of $\mathbf{M}$ are denoted $\operatorname{tr}(\mathbf{M})$, $\mathbf{M}^*$, $\mathbf{M}^{\mathsf{T}}$ and $\mathbf{M}^{\mathsf{H}}$, respectively. $\mathbf{M}^{-\mathsf{T}} \triangleq (\mathbf{M}^{-1})^{\mathsf{T}}$ and $\mathbf{M}^{-\mathsf{H}} \triangleq (\mathbf{M}^{-1})^{\mathsf{H}}$. $\mathbf{I}_m$ and $\mathbf{0}_{m\times n}$ denote the $m \times m$ identity matrix and $m \times n$ zero matrix, respectively, and the dimensions are omitted if cleared from the context. $\mathbf{M}_{[i:j]}$ denotes the sub-matrix containing columns from i to j of $\mathbf{M}$, and $\mathbf{M}_{[i]}$ denotes the i-th column. $(\mathbf{x})_{[i:j]}$ and $(\mathbf{x}^{\mathsf{T}})_{[i:j]}$ denotes respectively the column vector and row vector containing entries from i to j of a column vector $\mathbf{x}$. $\operatorname{Span}(\mathbf{U})$ denotes the subspace spanned by the columns of a truncated unitary matrix $\mathbf{U}$. $\operatorname{diag}(x_1, \dots, x_n)$ is a diagonal matrix with diagonal entries $x_1, \dots, x_n$. $[n] \triangleq \{1, 2, \dots, n\}$. $(x)^+ \triangleq \max\{x, 0\}$. $\mathbb{1}\{A\}$ is the indicator function of event A. Logarithms are in base 2. All rates are measured in bits per channel use.

CHAPTER 2

TRANSMIT CORRELATION DIVERSITY: GENERALIZATION, NEW TECHNIQUES, AND IMPROVED BOUNDS[1] [2]

Consider a MIMO broadcast channel in which a transmitter (also known as base station) equipped with M antennas transmitting to K receivers (also known as users), where User k is equipped with N_k antennas, $k \in [K]$. The received signal at User k at channel use j is

$$\mathbf{y}_k(j) = \mathbf{H}_k(j)\mathbf{x}(j) + \mathbf{w}_k(j), \qquad \text{for} \;\; k \in [K], \;\; j = 1, 2, \ldots \tag{2.1}$$

where $\mathbf{x}(j) \in \mathbb{C}^M$ is the transmitted signal at channel use j and $\mathbf{w}_k \in \mathbb{C}^{N_k}$ is the white noise with independent and identically distributed (i.i.d.) $\mathcal{CN}(0,1)$ entries. $\mathbf{H}_k(j) \in \mathbb{C}^{N_k \times M}$ is the channel matrix containing the random fading coefficients between M transmit antennas of the base station and N_k receive antennas of User k. We assume that $\frac{1}{MN_k}\mathbb{E}\left[\|\mathbf{H}_k\|^2\right] = 1, k \in [K]$. The transmitted signal is subject to the power constraint

$$\frac{1}{J}\sum_{j=1}^{J} \|\mathbf{x}(j)\|^2 \leq \rho, \tag{2.2}$$

where J is the number of channel uses spanned by a codeword (of a channel code). Therefore, ρ is the ratio between the average transmit power per antenna and the noise power, and is referred to as the SNR of the channel. Hereafter, we omit the channel use index j.

2.1 Channel Correlation

We assume that the channel is spatially correlated according to the Kronecker model, i.e., separable model, and focus on the transmit-side correlation. Thus the channel matrices are

---

expressed as

$$\mathbf{H}_k = \breve{\mathbf{H}}_k \mathbf{R}_k^{\frac{1}{2}}, \quad k \in [K], \tag{2.3}$$

where $\mathbf{R}_k = \frac{1}{N_k}\mathbb{E}\left[\mathbf{H}_k^{\mathsf{H}}\mathbf{H}_k\right] \in \mathbb{C}^{M\times M}$, $\mathrm{tr}\left(\mathbf{R}_k\right) = M$, is the transmit correlation matrix of User k with rank r_k, and $\breve{\mathbf{H}}_k \in \mathbb{C}^{N_k \times M}$ is drawn from a generic distribution satisfying the conditions

$$h(\breve{\mathbf{H}}_k) > -\infty, \quad \mathbb{E}\left[\breve{\mathbf{H}}_k^{\mathsf{H}}\breve{\mathbf{H}}_k\right] = N_k \mathbf{I}_M, \quad k \in [K]. \tag{2.4}$$

Since the correlation matrices might be rank-deficient, $\breve{\mathbf{H}}_k$ is not necessarily a minimal representation of the randomness in $\mathbf{H}_k$. The correlation eigenspace of User k is revealed via eigendecomposition of the correlation matrix:

$$\mathbf{R}_k = \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{U}_k^{\mathsf{H}}, \tag{2.5}$$

where $\mathbf{\Sigma}_k$ is a $r_k \times r_k$ diagonal matrix containing r_k non-zero eigenvalues of $\mathbf{R}_k$, and $\mathbf{U}_k$ is a $M \times r_k$ matrix whose orthonormal unit column vectors are the eigenvectors of $\mathbf{R}_k$ corresponding to the non-zero eigenvalues. The rows of $\mathbf{H}_k$ belong to the r_k-dimensional eigenspace $\mathrm{Span}(\mathbf{U}_k)$ of $\mathbf{R}_k$, also known as the eigenspace of User k.

The channel expression (2.3) can be expanded as

$$\mathbf{H}_k = \breve{\mathbf{H}}_k \mathbf{U}_k \mathbf{\Sigma}_k^{\frac{1}{2}} \mathbf{U}_k^{\mathsf{H}} = \mathbf{G}_k \mathbf{\Sigma}_k^{\frac{1}{2}} \mathbf{U}_k^{\mathsf{H}}, \tag{2.6}$$

where $\mathbf{G}_k \triangleq \breve{\mathbf{H}}_k \mathbf{U}_k$ is equivalently drawn from a generic distribution satisfying $h(\mathbf{G}_k) > -\infty$, $\mathbb{E}\left[\mathbf{G}_k^{\mathsf{H}}\mathbf{G}_k\right] = N_k \mathbf{I}_{r_k}$, $k \in [K]$.

The eigenspaces $\mathrm{Span}(\mathbf{U}_k)$ have a prominent role in transmit correlation diversity. For example, methods such as [25, 26, 27, 28] are critically dependent on finding groups of users whose eigenspaces have no intersection. In contrast, in this chapter, we propose transmission schemes that take advantage of both common and noncommon parts of the eigenspaces. To this end, in several instances, we build an equivalent channel $\bar{\mathbf{H}}_k$ that resides in a *subspace* of the eigenspace $\text{Span}(\mathbf{U}_k)$ via the linear transformation

$$\bar{\mathbf{H}}_k = \mathbf{H}_k \mathbf{V}_k, \tag{2.7}$$

for some truncated unitary matrix $\mathbf{V}_k \in \mathbb{C}^{M \times s_k}$, $s_k \leq r_k$, such that $\text{Span}(\mathbf{V}_k) \subset \text{Span}(\mathbf{U}_k)$. Unlike $\mathbf{U}_k$, $k \in [K]$, that characterize the correlation eigenspaces of the links, the subspaces $\text{Span}(\mathbf{V}_k)$ also depend on the proposed transmission scheme and may be customized throughout this chapter.

2.2 Channel Information Availability

We assume throughout this chapter that the distribution of $\mathbf{H}_k$, in particular the second-order statistic $\mathbf{R}_k$ (and thus $\mathbf{\Sigma}_k$ and $\mathbf{U}_k$), is known to both the base station and User k. This is reasonable because $\mathbf{R}_k$ represents long-term behavior of the channel that is stable and can be easily tracked. On the other hand, the realization of $\mathbf{H}_k$ changes much more rapidly. We consider two scenarios:

- CSIR (channel state information at the receiver): User k knows perfectly the realizations of $\mathbf{H}_k$.

- No free CSIR: User k only knows the distribution of $\mathbf{H}_k$. In this case, for a tractable model of the channel variation, we assume a block fading model with equal-length and synchronous coherence interval (across the users) of T channel uses. That is, $\mathbf{H}_k$ remains constant during each block of length T and changes independently across blocks [58]. We assume that $T \geq 2\max(r_k, N_k), \forall k$. Let $\mathbf{X} = [\mathbf{x}(1) \ \dots \ \mathbf{x}(T)]$ be the transmitted signal during a block, the received signal at User k during this block is

$$\mathbf{Y}_k = \mathbf{H}_k \mathbf{X} + \mathbf{W}_k, \tag{2.8}$$

where $\mathbf{Y}_k = [\mathbf{y}_k(1) \ \dots \ \mathbf{y}_k(T)]$, $\mathbf{W}_k = [\mathbf{w}_k(1) \ \dots \ \mathbf{w}_k(T)]$, and the block index is omitted for simplicity. User k might attempt to estimate $\mathbf{H}_k$ with the help of known pilot symbols inserted in $\mathbf{X}$.

2.3 Achievable Rate and DoF

Assuming K independent messages are communicated (no common message), and the corresponding rate tuple $(R_1(\rho), \dots, R_K(\rho))$ is achievable at SNR ρ, $\forall \rho \geq 0$, i.e., lie within the capacity region of the channel, then an achievable DoF tuple $(d_1, \dots, d_K)$ is defined as

$$d_k \triangleq \lim_{\rho \to \infty} \frac{R_k(\rho)}{\log \rho}, \quad k \in [K]. \tag{2.9}$$

The set of achievable rate (resp., DoF) tuples defines an achievable rate (resp., DoF) region of the channel.

For convenience, we denote $N_k^* \triangleq \min(N_k, r_k)$.

2.4 Preliminaries and Useful Results

Lemma 1 (The optimal single-user DoF)**.** *For the correlated MIMO broadcast channel in Section 2.1, the optimal single-user DoF of User k is $d_k = N_k^*$ with CSIR and $d_k = N_k^* \Big(1 - \frac{N_k^*}{T}\Big)$ without free CSIR.*

The result in the CSIR case is well-known (see, e.g., [59]). The no free CSIR case was reported in [37, Thm. 1]. The next lemma is used for the finite-SNR rate analysis.

Lemma 2 (Worst case uncorrelated additive noise [55])**.** *Consider the point-to-point channel*

$$\mathbf{y} = \sqrt{\frac{\rho}{M}} \mathbf{H}\mathbf{x} + \mathbf{w}, \tag{2.10}$$

where the channel $\mathbf{H} \in \mathbb{C}^{N \times M}$ is known to the receiver, and the signal $\mathbf{x} \in \mathbb{C}^{M \times 1}$ and the noise $\mathbf{w} \in \mathbb{C}^{N \times 1}$ satisfy the power constraints $\frac{1}{M}\mathbb{E}\left[\|\mathbf{x}\|^2\right] = 1$ and $\frac{1}{N}\mathbb{E}\left[\|\mathbf{w}\|^2\right] = 1$, are both *complex Gaussian distributed, and are uncorrelated, i.e,* $\mathbb{E}\left[\mathbf{x}\mathbf{w}^{\mathsf{H}}\right] = \mathbf{0}$*. Let* $\mathbf{R}_{\mathbf{x}} \triangleq \mathbb{E}\left[\mathbf{x}\mathbf{x}^{\mathsf{H}}\right]$ *and* $\mathbf{R}_{\mathbf{w}} \triangleq \mathbb{E}\left[\mathbf{w}\mathbf{w}^{\mathsf{H}}\right]$ *and assume* $\mathrm{tr}\left(\mathbf{R}_{\mathbf{x}}\right) = M$ *and* $\mathrm{tr}\left(\mathbf{R}_{\mathbf{w}}\right) = N$*. Then the mutual information* $I(\mathbf{y}; \mathbf{x} \,|\, \mathbf{H})$ *is lower bounded as*

$$I(\mathbf{y}; \mathbf{x} \,|\, \mathbf{H}) \geq \mathbb{E}\left[\log\det\left(\mathbf{I}_N + \frac{\rho}{M}\mathbf{R}_{\mathbf{w}}^{-1}\mathbf{H}\mathbf{R}_{\mathbf{x}}\mathbf{H}^{\mathsf{H}}\right)\right] \tag{2.11}$$

$$\geq \min_{\mathbf{R}_{\mathbf{w}}, \mathrm{tr}(\mathbf{R}_{\mathbf{w}})=N} \mathbb{E}\left[\log\det\left(\mathbf{I}_N + \frac{\rho}{M}\mathbf{R}_{\mathbf{w}}^{-1}\mathbf{H}\mathbf{R}_{\mathbf{x}}\mathbf{H}^{\mathsf{H}}\right)\right], \tag{2.12}$$

If the distribution of $\mathbf{H}$ *is left rotationally invariant, i.e.,* $p(\mathbf{\Theta}\mathbf{H}) = p(\mathbf{H})$ *for any deterministic* $N \times N$ *unitary matrix* $\mathbf{\Theta}$*, then the minimizing noise covariance matrix in* (2.12) *is* $\mathbf{R}_{\mathbf{w},\mathrm{opt}} = \mathbf{I}_N$*.*

Proof. The proof follows from the proof of [55, Thm. 1]. Specifically, the mutual information lower bound (2.11) was stated in [55, Eq.(27)]. To show that $\mathbf{R}_{\mathbf{w},\mathrm{opt}} = \mathbf{I}_N$, we diagonalize $\mathbf{R}_{\mathbf{w}}$ using the left rotational invariance of $\mathbf{H}$, and then use the convexity of $\mathbb{E}\left[\log\det\left(\mathbf{I}_N + \frac{\rho}{M}\mathbf{R}_{\mathbf{w}}^{-1}\mathbf{H}\mathbf{R}_{\mathbf{x}}\mathbf{H}^{\mathsf{H}}\right)\right]$ in the diagonalized $\mathbf{R}_{\mathbf{w}}$. □

The next lemma gives the MMSE estimator used for pilot-based channel estimation without free CSIR.

Lemma 3 (MMSE estimator)**.** *Consider the following linear model*

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{W}, \tag{2.13}$$

where $\mathbf{H} \in \mathbb{C}^{N \times M}$ *has correlation matrix* $\mathbf{R} = \frac{1}{N}\mathbb{E}\left[\mathbf{H}^{\mathsf{H}}\mathbf{H}\right]$*,* $\mathbf{X} \in \mathbb{C}^{M \times M}$ *is known, and* $\mathbf{W} \in \mathbb{C}^{N \times M}$ *has i.i.d.* $\mathcal{CN}(0, 1)$ *entries. The linear MMSE estimator for* $\mathbf{H}$ *is given by*

$$\hat{\mathbf{H}} = \mathbf{Y}(\mathbf{X}^{\mathsf{H}}\mathbf{R}\mathbf{X} + \mathbf{I}_M)^{-1}\mathbf{X}^{\mathsf{H}}\mathbf{R}. \tag{2.14}$$

The MMSE estimate $\hat{\mathbf{H}}$ *is also the conditional mean:* $\hat{\mathbf{H}} = \mathbb{E}[\mathbf{H} \,|\, \mathbf{X}, \mathbf{Y}]$*. The estimate* $\hat{\mathbf{H}}$ *and the estimation error* $\tilde{\mathbf{H}} = \mathbf{H} - \hat{\mathbf{H}}$ *are uncorrelated, which have zero mean and row covariance*

$$\frac{1}{N}\mathbb{E}[\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}] = \mathbf{R}\mathbf{X}(\mathbf{X}^{\mathsf{H}}\mathbf{R}\mathbf{X} + \mathbf{I}_M)^{-1}\mathbf{X}^{\mathsf{H}}\mathbf{R}, \tag{2.15}$$

$$\frac{1}{N}\mathbb{E}[\tilde{\mathbf{H}}^{\mathsf{H}}\tilde{\mathbf{H}}] = \mathbf{R} - \mathbf{R}\mathbf{X}(\mathbf{X}^{\mathsf{H}}\mathbf{R}\mathbf{X} + \mathbf{I}_M)^{-1}\mathbf{X}^{\mathsf{H}}\mathbf{R}. \tag{2.16}$$

Proof. The linear MMSE channel estimator is given by $\hat{\mathbf{H}} = \mathbf{Y}\mathbf{A}$ where $\mathbf{A}$ is the minimizer of the MSE

$$\frac{1}{N}\mathbb{E}[\|\mathbf{H} - \hat{\mathbf{H}}\|_F^2] = \mathrm{tr}\left(\mathbf{R}\right) - \mathrm{tr}\left(\mathbf{R}\mathbf{X}\mathbf{A}\right) - \mathrm{tr}\left(\mathbf{A}^{\mathsf{H}}\mathbf{X}^{\mathsf{H}}\mathbf{R}\right) + \mathrm{tr}\left(\mathbf{A}^{\mathsf{H}}(\mathbf{X}^{\mathsf{H}}\mathbf{R}\mathbf{X} + \mathbf{I}_M)\mathbf{A}\right). \quad (2.17)$$

Solving $\frac{\partial}{\partial \mathbf{A}}\frac{1}{N}\mathbb{E}[\|\mathbf{H}-\hat{\mathbf{H}}\|_F^2] = 0$ yields the optimal $\mathbf{A}_{\mathrm{opt}} = (\mathbf{X}^{\mathsf{H}}\mathbf{R}\mathbf{X}+\mathbf{I}_M)^{-1}\mathbf{X}^{\mathsf{H}}\mathbf{R}$. Some further simple manipulations give (2.15) and (2.16). □

2.5 Two-user Broadcast Channel: DoF Analysis

Both with or without free CSIR assumption, we study first the special case of fully overlapping correlation eigenspaces, then the more general case of partially overlapping correlation eigenspaces.

2.5.1 CSIR

Fully Overlapping Eigenspaces

Consider the case where both users have spatially correlated channels, and User 2's channel eigenspace is a subspace of User 1's, which implies $r_2 \leq r_1 \leq M$.

Proposition 1. *For the two-user broadcast channel with CSIR, when the eigenspace of User 2 is a subspace of User 1's (implying $r_2 \leq r_1 \leq M$), the DoF pairs $(N_1^*, 0), (0, N_2^*)$, and $\left((N_1^* - r_2)^+, N_2^*\right)$ are achievable. Furthermore, if $r_1 \geq N_1 \geq r_1 - r_2$, the DoF pair $\left(r_1 - r_2, \min(N_1 - r_1 + r_2, N_2^*)\right)$ is also achievable. The convex hull of these pairs and the origin $(0, 0)$ is an achievable DoF region.*

Proof. According to Lemma 1, the DoF pairs $(N_1^*, 0)$ and $(0, N_2^*)$ are achievable.

When $N_1^* \geq r_2$, the pair $(N_1^* - r_2, N_2^*)$ can be achieved as follows.

Recall that the eigenspaces of channels $\mathbf{H}_1$ and $\mathbf{H}_2$ are $\mathrm{Span}(\mathbf{U}_1)$ and $\mathrm{Span}(\mathbf{U}_2)$, respectively, and in the present case, $\mathrm{Span}(\mathbf{U}_2) \subset \mathrm{Span}(\mathbf{U}_1)$. There exist transmit eigendirections

$\mathbf{V}_1 \in \mathbb{C}^{M\times(N_1^*-r_2)}, \mathbf{V}_0 \in \mathbb{C}^{M\times N_2^*}$ that are aligned with the common and non-common parts of the two channel eigenspaces such that

$$\mathrm{Span}(\mathbf{V}_0) \subset \mathrm{Span}(\mathbf{U}_2), \tag{2.18}$$

$$\mathrm{Span}(\mathbf{V}_1) \subset \big(\mathrm{Span}(\mathbf{U}_1) \cap \mathrm{Span}(\mathbf{U}_2)^{\perp}\big). \tag{2.19}$$

Define $\mathbf{V} \triangleq [\mathbf{V}_0\ \mathbf{V}_1]$. The proposed transmission scheme is $\mathbf{x} = \mathbf{V}\big[\mathbf{s}_0^{\mathsf{T}}\ \mathbf{s}_1^{\mathsf{T}}\big]^{\mathsf{T}}$ where the signals $\mathbf{s}_1 \in \mathbb{C}^{N_1^*-r_2}, \mathbf{s}_0 \in \mathbb{C}^{N_2^*}$ are intended for User 1 and User 2, respectively. The received signal at User 1 is

$$\mathbf{y}_1 = \mathbf{H}_1\mathbf{x} + \mathbf{w}_1 = \mathbf{H}_1\mathbf{V}\begin{bmatrix}\mathbf{s}_0\\ \mathbf{s}_1\end{bmatrix} + \mathbf{w}_1. \tag{2.20}$$

Since User 1 knows $\mathbf{H}_1\mathbf{V}$, it can decode both $\mathbf{s}_1$ and $\mathbf{s}_0$, achieving respectively $N_1^* - r_2$ and N_2^* DoF. The received signal at User 2 is

$$\mathbf{y}_2 = \mathbf{H}_2\mathbf{x} + \mathbf{w}_2 = \mathbf{H}_2[\mathbf{V}_0\ \mathbf{V}_1]\begin{bmatrix}\mathbf{s}_0\\ \mathbf{s}_1\end{bmatrix} + \mathbf{w}_2 = \mathbf{H}_2\mathbf{V}_0\mathbf{s}_0 + \mathbf{w}_2, \tag{2.21}$$

which uses $\mathbf{H}_2\mathbf{V}_1 = \mathbf{0}$ due to (2.19). Since User 2 knows $\mathbf{H}_2\mathbf{V}_0$, it can decode $\mathbf{s}_2$, achieving N_2^* DoF. By dedicating $\mathbf{s}_2$ to user 2, the DoF pair $(N_1^* - r_2, N_2^*)$ is achieved.

The pair $\big(r_1 - r_2, \min(N_1 - r_1 + r_2, N_2^*)\big)$ can be achieved similarly when $r_1 \geq N_1 \geq r_1 - r_2$ by setting $\mathbf{V}_1 \in \mathbb{C}^{M\times(r_1-r_2)}, \mathbf{V}_0 \in \mathbb{C}^{M\times\min(N_1-r_1+r_2,N_2^*)}$, and the dimensions of $\mathbf{s}_1, \mathbf{s}_0$ accordingly. □

Partially Overlapping Eigenspaces

Theorem 1. *For the two-user broadcast channel with CSIR,* $\mathrm{rank}(\mathrm{Span}(\mathbf{U}_1) \cap \mathrm{Span}(\mathbf{U}_2)) = r_0 \geq 0$, *the DoF pairs* $(N_1^*, 0), (0, N_2^*)$, $\big((N_1^* - r_0)^+, N_2^*\big)$, *and* $\big(N_1^*, (N_2^* - r_0)^+\big)$ *are achievable. Furthermore, if* $N_1 \leq r_1$ *and* $N_2 \leq r_2$, *the DoF pairs*

$$\Big(\min\big(N_1, r_1 - r_0\big) + \min\big((N_1 - r_1 + r_0)^+, (N_2 - r_2 + r_0)^+\big),\ \min\big(N_2, r_2 - r_0\big)\Big), \tag{2.22}$$

$$\Big(\min\big(N_1, r_1 - r_0\big),\ \min\big(N_2, r_2 - r_0\big) + \min\big((N_1 - r_1 + r_0)^+, (N_2 - r_2 + r_0)^+\big)\Big), \tag{2.23}$$

are also achievable. The convex hull of these pairs and the origin $(0,0)$ *is an achievable DoF region.*

Proof. The DoF pairs $(N_1^*, 0)$ and $(0, N_2^*)$ are achievable according to Lemma 1. The achievable schemes for the other pairs are as follows. For non-negative integers $s_0 \leq r_0$, $s_1 \leq r_1 - r_0$, and $s_2 \leq r_2 - r_0$, there exist transmit eigendirections $\mathbf{V}_0 \in \mathbb{C}^{M \times s_0}$ aligned with the common part of the two channel eigenspaces, and eigendirections $\mathbf{V}_1 \in \mathbb{C}^{M \times s_1}, \mathbf{V}_2 \in \mathbb{C}^{M \times s_2}$ aligned with the two non-common parts, such that

$$\mathrm{Span}(\mathbf{V}_0) \subset \big(\mathrm{Span}(\mathbf{U}_1) \cap \mathrm{Span}(\mathbf{U}_2)\big), \tag{2.24}$$

$$\mathrm{Span}(\mathbf{V}_1) \subset \big(\mathrm{Span}(\mathbf{U}_1) \cap \mathrm{Span}(\mathbf{U}_2)^{\perp}\big), \tag{2.25}$$

$$\mathrm{Span}(\mathbf{V}_2) \subset \big(\mathrm{Span}(\mathbf{U}_2) \cap \mathrm{Span}(\mathbf{U}_1)^{\perp}\big). \tag{2.26}$$

Define $\mathbf{V} \triangleq [\mathbf{V}_0\ \mathbf{V}_1\ \mathbf{V}_2]$. Let the transmitter send the signal $\mathbf{x} = \mathbf{V}\big[\mathbf{s}_0^{\mathsf{T}}\ \mathbf{s}_1^{\mathsf{T}}\ \mathbf{s}_2^{\mathsf{T}}\big]^{\mathsf{T}}$, where $\mathbf{s}_k \in \mathbb{C}^{s_k}$ contains symbols for User k, $k \in \{1, 2\}$, and $\mathbf{s}_0 \in \mathbb{C}^{s_0}$ contains symbols that both users can decode.

The received signal at User 1 and User 2 are respectively

$$\mathbf{y}_1 = \mathbf{H}_1\mathbf{x} + \mathbf{w}_1 = \mathbf{H}_1[\mathbf{V}_0\ \mathbf{V}_1]\begin{bmatrix}\mathbf{s}_0\\ \mathbf{s}_1\end{bmatrix} + \mathbf{w}_1, \tag{2.27}$$

$$\mathbf{y}_2 = \mathbf{H}_2\mathbf{x} + \mathbf{w}_2 = \mathbf{H}_2[\mathbf{V}_0\ \mathbf{V}_2]\begin{bmatrix}\mathbf{s}_0\\ \mathbf{s}_2\end{bmatrix} + \mathbf{w}_2, \tag{2.28}$$

using $\mathbf{H}_2\mathbf{V}_1 = \mathbf{0}$ and $\mathbf{H}_1\mathbf{V}_2 = \mathbf{0}$ due to (2.25) and (2.26), respectively. Then if $s_k + s_0 \leq N_k$, User k can decode both $\mathbf{s}_k$ and $\mathbf{s}_0$, $k \in \{1, 2\}$.

- If $N_1 \geq r_0$ and $N_2 \leq r_0$, set $s_1 = N_1^* - r_0$, $s_2 = 0$, and $s_0 = N_2$. By dedicating $\mathbf{s}_0$ to User 2, the DoF pair $(N_1^* - r_0, N_2)$ can be achieved. Similarly, if $N_1 \leq r_0$ and $N_2 \geq r_0$, the DoF pair $(N_1, N_2^* - r_0)$ can be achieved.

- If $N_1 \geq r_0$ and $N_2 \geq r_0$, set $s_1 = N_1^* - r_0$, $s_2 = N_2^* - r_0$, and $s_0 = r_0$. By dedicating $\mathbf{s}_0$ to one of the users, the DoF pairs $\left(N_1^* - r_0, N_2^*\right)$ and $\left(N_1^*, N_2^* - r_0\right)$ are achievable.

- When $N_1 \leq r_1$ and $N_2 \leq r_2$, by setting $s_1 = \min\left(N_1, r_1 - r_0\right)$, $s_2 = \min\left(N_2, r_2 - r_0\right)$, $s_0 = \min\left((N_1 - r_1 + r_0)^+, (N_2 - r_2 + r_0)^+\right)$, and dedicating $\mathbf{s}_0$ to one of the users, the DoF pairs given in (2.22) and (2.23) are achievable.

Therefore, the proof is completed. $\square$

An outer bound for the achievable DoF region is given as follows.

Theorem 2. *When* $\mathrm{rank}(\mathrm{Span}(\mathbf{U}_1) \cap \mathrm{Span}(\mathbf{U}_2)) = r_0 \geq 0$*, the achievable DoF region is outer bounded by* $d_k \leq N_k^*$*,* $k \in \{1, 2\}$*, and*

$$d_1 + d_2 \leq \min\{r_1 + r_2 - r_0, N_1 + N_2\}. \tag{2.29}$$

When $\{r_1 \leq N_1, r_2 \leq N_2\}$ *or* $\{N_1 \leq r_1 - r_0, N_2 \leq r_2 - r_0\}$*, this outer bound is tight.*

Proof. The single-user bounds $d_k \leq N_k^*$, $k \in \{1, 2\}$, follow from Lemma 1.

Denote by $\mathbf{V}_1 \in \mathbb{C}^{M \times (r_1 - r_0)}, \mathbf{V}_2 \in \mathbb{C}^{M \times (r_2 - r_0)}$ the non-unique transmit eigendirections that are aligned with the non-common parts, i.e., $\mathrm{Span}(\mathbf{V}_1) = \mathrm{Span}(\mathbf{U}_1) \cap \mathrm{Span}(\mathbf{U}_2)^{\perp}$ and $\mathrm{Span}(\mathbf{V}_2) = \mathrm{Span}(\mathbf{U}_2) \cap \mathrm{Span}(\mathbf{U}_1)^{\perp}$, and $\mathbf{V}_0 \in \mathbb{C}^{M \times r_0}$ the common part, i.e., $\mathrm{Span}(\mathbf{V}_0) = \mathrm{Span}(\mathbf{U}_1) \cap \mathrm{Span}(\mathbf{U}_2)$, of the eigenspaces. Let $\mathbf{V}_{\perp} \in \mathbb{C}^{M \times (M - r_1 - r_2 + r_0)}$ denote the orthogonal complement of the total channel eigenspaces, i.e., $\mathbf{V} \triangleq [\mathbf{V}_0 \ \mathbf{V}_1 \ \mathbf{V}_2 \ \mathbf{V}_{\perp}]$ is an unitary matrix. For a transmit vector $\mathbf{x} \in \mathbb{C}^M$, define $[\mathbf{x}_0^{\mathsf{T}} \ \mathbf{x}_1^{\mathsf{T}} \ \mathbf{x}_2^{\mathsf{T}} \ \mathbf{x}_{\perp}^{\mathsf{T}}]^{\mathsf{T}} \triangleq \mathbf{V}\mathbf{x}$, where $\mathbf{x}_0 \in \mathbb{C}^{r_0}$, $\mathbf{x}_1 \in \mathbb{C}^{r_1 - r_0}$, $\mathbf{x}_2 \in \mathbb{C}^{r_2 - r_0}$ and $\mathbf{x}_{\perp} \in \mathbb{C}^{M - r_1 - r_2 + r_0}$.

A cooperative cut-set upper bound is as follows, using invertibility of $\mathbf{V}$:

$$R_1 + R_2 \leq I(\mathbf{y}_1, \mathbf{y}_2; \mathbf{x}) = I(\mathbf{y}_1, \mathbf{y}_2; \mathbf{V}\mathbf{x}). \tag{2.30}$$

The next step is to bound the right-hand side in (2.30). To extract a full-rank representation of $\mathbf{H}_1$ and $\mathbf{H}_2$,

$$\begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{G}_1\mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{U}_1^{\mathsf{H}} \\ \mathbf{G}_2\mathbf{\Sigma}_2^{\frac{1}{2}}\mathbf{U}_2^{\mathsf{H}} \end{bmatrix} = \begin{bmatrix} \mathbf{G}_1\mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{T}_{1c} & \mathbf{G}_1\mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{T}_{1p} & \mathbf{0} \\ \mathbf{G}_2\mathbf{\Sigma}_2^{\frac{1}{2}}\mathbf{T}_{2c} & \mathbf{0} & \mathbf{G}_2\mathbf{\Sigma}_2^{\frac{1}{2}}\mathbf{T}_{2p} \end{bmatrix} [\mathbf{V}_0\ \mathbf{V}_1\ \mathbf{V}_2]^{\mathsf{H}}, \tag{2.31}$$

where $\mathbf{T}_{1c}, \mathbf{T}_{1p}, \mathbf{T}_{2c}$, and $\mathbf{T}_{2p}$ are matrices such that $[\mathbf{T}_{ic}\ \mathbf{T}_{ip}]$ is non-singular and $\mathbf{U}_k^{\mathsf{H}} = [\mathbf{T}_{ic}\ \mathbf{T}_{ip}][\mathbf{V}_0\ \mathbf{V}_i]^{\mathsf{H}},\ i \in \{1, 2\}$. Replacing $\mathbf{x}$ by $\mathbf{V}\mathbf{x}$, the concatenated received signal is

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \end{bmatrix} \mathbf{V}\mathbf{x} + \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{bmatrix} = \tilde{\mathbf{H}} \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{bmatrix}, \tag{2.32}$$

where

$$\tilde{\mathbf{H}} \triangleq \begin{bmatrix} \mathbf{G}_1\mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{T}_{1c} & \mathbf{G}_1\mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{T}_{1p} & \mathbf{0} \\ \mathbf{G}_2\mathbf{\Sigma}_2^{\frac{1}{2}}\mathbf{T}_{2c} & \mathbf{0} & \mathbf{G}_2\mathbf{\Sigma}_2^{\frac{1}{2}}\mathbf{T}_{2p} \end{bmatrix} \in \mathbb{C}^{(N_1+N_2)\times(r_1+r_2-r_0)}. \tag{2.33}$$

Because $\tilde{\mathbf{H}}$ is known at the receivers,

$$\begin{aligned} I(\mathbf{y}_1, \mathbf{y}_2; \mathbf{V}\mathbf{x}) &= I(\mathbf{y}_1, \mathbf{y}_2; \mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2) && (2.34) \\ &\leq \min\{r_1 + r_2 - r_0, N_1 + N_2\} \log \rho + o(\log \rho). && (2.35) \end{aligned}$$

This yields the sum DoF bound $d_1 + d_2 \leq \min\{r_1 + r_2 - r_0, N_1 + N_2\}$. This outer bound is tight against the achievable region in Theorem 1. □

Figure 2.1 shows the regions where the outer bound in Theorem 2 is tight.

Figure 2.2 compares the achievable region proposed in Theorem1 and the achievable region achieved with TDMA (time sharing between $(N_1^*, 0)$ and $(0, N_2^*)$) for $r_1 = 12$, $r_2 = 10$, $r_0 \in \{0, 3, 6, 9\}$ and $N_1 \geq r_1$, $N_2 \geq r_2$. The proposed achievable region is much larger than the TDMA region, especially when r_0 is small. In this setting, according to Theorem 2, the proposed region is optimal.

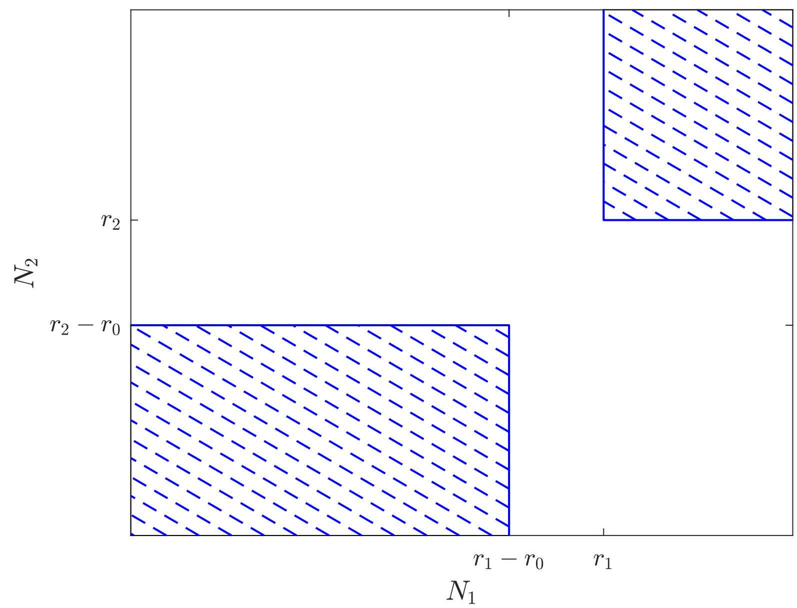

Figure 2.1. Regions (the hashed part) where the outer bound for the DoF region with CSIR in Theorem 2 is tight.

2.5.2 No free CSIR

In this case, CSIR is not available a priori and must be acquired via pilot transmission. On the one hand, one needs to take into account the cost of CSI acquisition in both energy and DoF. On the other hand, pilot transmission enables product superposition [50] that can improve upon rate splitting.

Fully Overlapping Eigenspaces

Consider the case where User 2's eigenspace is a subspace of User 1's, which implies $r_2 \le r_1 \le M$. The following proposition presents achievable DoF with product superposition in this case.

Proposition 2. *In a two-user broadcast channel without free CSIR, when the eigenspace of User 2 is a subspace of User 1's (implying $r_2 \le r_1 \le M$), the DoF pair $\left(N_1^*\big(1-\frac{r_1}{T}\big), N_2^*\frac{r_1-r_2}{T}\right)$ is achievable with product superposition.*

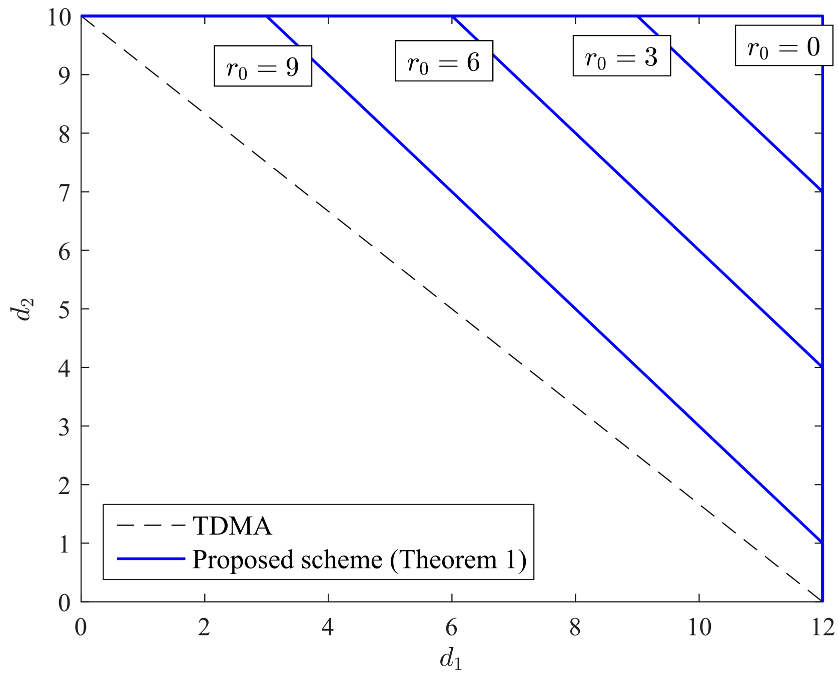

Figure 2.2. The achievable DoF region for two-users with CSIR, under TDMA and the proposed scheme (Theorem 1) for $r_1 = 12$, $r_2 = 10$, $r_0 \in \{0, 3, 6, 9\}$ and $N_1 \geq r_1$, $N_2 \geq r_2$. In this case, the latter region is optimal.

Proof. There exist transmit eigendirections $\mathbf{V}_1 \in \mathbb{C}^{M \times (r_1 - r_2)}$ and $\mathbf{V}_0 \in \mathbb{C}^{M \times r_2}$ that are aligned with the noncommon and common parts, respectively, of the two channel eigenspaces such that

$$\mathrm{Span}(\mathbf{V}_0) = \mathrm{Span}(\mathbf{U}_2), \tag{2.36}$$

$$\mathrm{Span}(\mathbf{V}_1) = \mathrm{Span}(\mathbf{U}_1) \cap \mathrm{Span}(\mathbf{U}_2)^{\perp}. \tag{2.37}$$

Define $\mathbf{V} \triangleq [\mathbf{V}_0 \ \mathbf{V}_1]$. Let the transmitter send the signal $\mathbf{X} = \mathbf{V}\mathbf{X}_2\mathbf{X}_1$ during a coherence block, with $\mathbf{X}_1 = [\mathbf{I}_{r_1} \ \mathbf{S}_1] \in \mathbb{C}^{r_1 \times T}$ and $\mathbf{X}_2 = \begin{bmatrix} \mathbf{I}_{r_2} & \mathbf{S}_2 \\ \mathbf{0} & \mathbf{I}_{r_1 - r_2} \end{bmatrix} \in \mathbb{C}^{r_1 \times r_1}$, where $\mathbf{S}_1 \in \mathbb{C}^{r_1 \times (T - r_1)}$ contains symbols for User 1 and $\mathbf{S}_2 \in \mathbb{C}^{r_2 \times (r_1 - s_0)}$ contains symbols for User 2. The received signal at User 1 is

$$\mathbf{Y}_1 = \mathbf{H}_1\mathbf{V}\mathbf{X}_2\mathbf{X}_1 + \mathbf{W}_1 = \mathbf{H}_1\mathbf{V}\mathbf{X}_2[\mathbf{I}_{r_1} \ \mathbf{S}_1] + \mathbf{W}_1. \tag{2.38}$$

User 1 estimates the equivalent channel $\mathbf{H}_1\mathbf{V}\mathbf{X}_2$ and then decodes $\mathbf{S}_1$, achieving $N_1^*(T-r_1)$ DoF. The received signal at User 2 during the first r_1 channel uses is

$$\mathbf{Y}_{2[1:r_1]} = \mathbf{H}_2\mathbf{V}\begin{bmatrix}\mathbf{I}_{r_2} & \mathbf{S}_2 \\ \mathbf{0} & \mathbf{I}_{r_1-r_2}\end{bmatrix}\mathbf{I}_{r_1} + \mathbf{W}_{2[1:r_1]} = \mathbf{H}_2\mathbf{V}_0[\mathbf{I}_{r_2} \quad \mathbf{S}_2] + \mathbf{W}_{2[1:r_1]}, \tag{2.39}$$

using $\mathbf{H}_2\mathbf{V}_1 = \mathbf{0}$ due to (2.37). User 2 estimates the equivalent channel $\mathbf{H}_2\mathbf{V}_0$, and then decodes $\mathbf{S}_2$, achieving $N_2^*(r_1 - r_2)$ DoF. Therefore, the normalized DoF pair $\Big(N_1^*\big(1-\frac{r_1}{T}\big), N_2^*\frac{r_1-r_2}{T}\Big)$ is achievable. $\square$

Partially Overlapping Eigenspaces

Theorem 3. *For the two-user broadcast channel without free CSIR and* $\mathrm{rank}(\mathrm{Span}(\mathbf{U}_1) \cap \mathrm{Span}(\mathbf{U}_2)) = r_0 \geq 0$*, the DoF pairs* $\Big(N_1^*\big(1-\frac{N_1^*}{T}\big), 0\Big)$ *and* $\Big(0, N_2^*\big(1-\frac{N_2^*}{T}\big)\Big)$ *are achievable. Furthermore, for any integers* (s_1, s_2, s_0) *such that* $0 \leq s_1 \leq r_1 - r_0$*,* $0 \leq s_2 \leq r_2 - r_0$*, and* $0 \leq s_0 \leq r_0$*, the DoF pairs*

$$\mathcal{D}_1 = \bigg(\min(s_0, N_1)\frac{s_2}{T}, \min(s_2 + s_0, N_2)\Big(1 - \frac{s_2 + s_0}{T}\Big)\bigg), \tag{2.40}$$

$$\mathcal{D}_2 = \bigg(\min(s_1 + s_0, N_1)\Big(1 - \frac{s_1 + s_0}{T}\Big), \min(s_0, N_2)\frac{s_1}{T}\bigg) \tag{2.41}$$

are achievable. On top of that, if $s_1 \geq s_2$*, the DoF pairs*

$$\begin{aligned}\mathcal{D}_3 = \bigg(&\min(s_1 + s_0, N_1)\Big(1 - \frac{s_1 + s_0}{T}\Big), \\ &\min(s_2, N_2)\frac{s_1 - s_2}{T} + \min(s_2, (N_2 - s_0)^+)\Big(1 - \frac{s_1 + s_0}{T}\Big)\bigg),\end{aligned} \tag{2.42}$$

$$\begin{aligned}\mathcal{D}_4 = \bigg(&\min(s_1, (N_1 - s_0)^+)\Big(1 - \frac{s_1 + s_0}{T}\Big), \\ &\min(s_2, N_2)\frac{s_1 - s_2}{T} + \min(s_2 + s_0, N_2)\Big(1 - \frac{s_1 + s_0}{T}\Big)\bigg),\end{aligned} \tag{2.43}$$

$$\begin{aligned}\mathcal{D}_5 = \bigg(&\min(s_1 + s_0, N_1)\Big(1 - \frac{s_1 + s_0}{T}\Big), \\ &\min(s_2 + s_0, N_2)\frac{s_1 - s_2}{T} + \min(s_2, N_2)\Big(1 - \frac{s_1 + s_0}{T}\Big)\bigg)\end{aligned} \tag{2.44}$$

are achievable; if $s_1 \le s_2$, the DoF pairs

$$\mathcal{D}_3 = \Big(\min(s_1, N_1)\frac{s_2 - s_1}{T} + \min(s_1, (N_1 - s_0)^+)\Big(1 - \frac{s_2 + s_0}{T}\Big), \\ \min(s_2 + s_0, N_2)\Big(1 - \frac{s_2 + s_0}{T}\Big)\Big), \tag{2.45}$$

$$\mathcal{D}_4 = \Big(\min(s_1, N_1)\frac{s_2 - s_1}{T} + \min(s_1 + s_0, N_1)\Big(1 - \frac{s_2 + s_0}{T}\Big), \\ \min(s_2, (N_2 - s_0)^+)\Big(1 - \frac{s_2 + s_0}{T}\Big)\Big), \tag{2.46}$$

$$\mathcal{D}_5 = \Big(\min(s_1 + s_0, N_1)\frac{s_2 - s_1}{T} + \min(s_1, N_1)\Big(1 - \frac{s_2 + s_0}{T}\Big), \\ \min(s_2 + s_0, N_2)\Big(1 - \frac{s_2 + s_0}{T}\Big)\Big) \tag{2.47}$$

are achievable. The convex hull of these DoF pairs (over all feasible values of s_1, s_2, and s_0) and the origin $(0, 0)$ is achievable.

Remark 1. *The parameters s_0, s_1, s_2 represent the allocation of available dimensions to the encoding of messages for the two users. By tuning these parameters, we explore the trade-off between the number of data dimensions (indicating the amount of channel uses needed for pilot transmission) and the amount of channel uses for data transmission within each section of the eigenspaces.*

Proof of Theorem 3. The DoF pairs $\Big(N_1^*\Big(1 - \frac{N_1^*}{T}\Big), 0\Big)$ and $\Big(0, N_2^*\Big(1 - \frac{N_2^*}{T}\Big)\Big)$ are achieved by activating only one user according to Lemma 1.

For any non-negative integers s_0, s_1, s_2 satisfying $s_0 \le r_0$, $s_1 \le r_1 - r_0$ and $s_2 \le r_2 - r_0$, there exist eigendirections $\mathbf{V}_0 \in \mathbb{C}^{M \times s_0}, \mathbf{V}_1 \in \mathbb{C}^{M \times s_1}, \mathbf{V}_2 \in \mathbb{C}^{M \times s_2}$ that are aligned with part of the common and non-common sections of the two channel eigenspaces such that[1]

$$\mathrm{Span}(\mathbf{V}_0) \subset \big(\mathrm{Span}(\mathbf{U}_1) \cap \mathrm{Span}(\mathbf{U}_2)\big), \tag{2.48}$$

$$\mathrm{Span}(\mathbf{V}_1) \subset \big(\mathrm{Span}(\mathbf{U}_1) \cap \mathrm{Span}(\mathbf{U}_2)^{\perp}\big), \tag{2.49}$$

$$\mathrm{Span}(\mathbf{V}_2) \subset \big(\mathrm{Span}(\mathbf{U}_2) \cap \mathrm{Span}(\mathbf{U}_1)^{\perp}\big). \tag{2.50}$$

To achieve $\mathcal{D}_1$, the base station employs product superposition and transmits

$$\mathbf{X} = [\mathbf{V}_0 \ \mathbf{V}_1]\mathbf{X}_2\mathbf{X}_1, \tag{2.51}$$

with $\mathbf{X}_1 = [\mathbf{I}_{s_1+s_0} \ \mathbf{S}_1]$ and $\mathbf{X}_2 = \begin{bmatrix} \mathbf{I}_{s_0} & \mathbf{S}_2 \\ \mathbf{0} & \mathbf{I}_{s_1} \end{bmatrix}$, where $\mathbf{S}_1 \in \mathbb{C}^{(s_1+s_0)\times(T-s_1-s_0)}$ and $\mathbf{S}_2 \in \mathbb{C}^{s_0\times s_1}$ contain symbols for User 1 and User 2, respectively. Following steps similar to the proof of Proposition 2, it can be verified that this achieves the DoF pair $\mathcal{D}_1$. The DoF pair $\mathcal{D}_2$ can be achieved similarly by switching the users' roles.

When $s_1 \geq s_2$, the pairs $\mathcal{D}_3$ and $\mathcal{D}_4$ are achieved with rate splitting as follows. Let the transmitter send

$$\mathbf{X} = [\mathbf{V}_0 \ \mathbf{V}_1 \ \mathbf{V}_2] \begin{bmatrix} \mathbf{I}_{s_0} & [\mathbf{0}_{s_0\times s_1} \quad \mathbf{S}_0] \\ \mathbf{0}_{s_1\times s_0} & [\mathbf{I}_{s_1} \quad \mathbf{S}_1] \\ \mathbf{0}_{s_2\times s_0} & [\mathbf{I}_{s_2} \quad \mathbf{S}_2] \end{bmatrix}, \tag{2.52}$$

where $\mathbf{S}_0 \in \mathbb{C}^{s_0\times(T-s_1-s_0)}$ is a common signal to both users while $\mathbf{S}_1 \in \mathbb{C}^{s_1\times(T-s_1-s_0)}$ and $\mathbf{S}_2 \in \mathbb{C}^{s_2\times(T-s_2-s_0)}$ are private signals to User 1 and User 2, respectively.

---

[1] $\mathbf{V}_0$ can be calculated from $\mathbf{U}_1$ and $\mathbf{U}_2$ using, e.g., the Zassenhaus algorithm [60]. Specifically, this algorithm uses elementary row operations to transform the $(r_1 + r_2) \times 2M$ matrix $\begin{bmatrix} \mathbf{U}_1^{\intercal} & \mathbf{U}_1^{\intercal} \\ \mathbf{U}_2^{\intercal} & \mathbf{0}_{r_2\times M} \end{bmatrix}$ (or $\begin{bmatrix} \mathbf{U}_2^{\intercal} & \mathbf{U}_2^{\intercal} \\ \mathbf{U}_1^{\intercal} & \mathbf{0}_{r_1\times M} \end{bmatrix}$) to the row echelon form $\begin{bmatrix} \mathbf{V}_0^{\intercal} & * \\ \mathbf{0} & \mathbf{V}_0^{\intercal} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, where $*$ stands for a matrix which is not of interest. $\mathbf{V}_1$ and $\mathbf{V}_2$ can be found similarly by applying the Zassenhaus algorithm to $\mathbf{U}_1$ and $\mathrm{null}(\mathbf{U}_2)$, and $\mathrm{null}(\mathbf{U}_1)$ and $\mathbf{U}_2$, respectively, where $\mathrm{null}(\mathbf{U}_k)$ is the matrix such that $[\mathbf{U}_k \ \mathrm{null}(\mathbf{U}_k)]$ is unitary.

The received signal at User 1 is

$$\begin{aligned}\mathbf{Y}_1 &= \mathbf{H}_1[\mathbf{V}_0\ \mathbf{V}_1\ \mathbf{V}_2]\mathbf{X} + \mathbf{W}_1 \\ &= \mathbf{H}_1[\mathbf{V}_0\ \mathbf{V}_1]\begin{bmatrix}\mathbf{I}_{s_0} & \mathbf{0} & \mathbf{S}_0 \\ \mathbf{0} & \mathbf{I}_{s_1} & \mathbf{S}_1\end{bmatrix} + \mathbf{W}_1.\end{aligned} \tag{2.53}$$

User 1 estimates the equivalent channel $\mathbf{H}_1[\mathbf{V}_0\ \mathbf{V}_1]$ during the first $s_1 + s_0$ channel uses and decodes both $\mathbf{S}_1$ and $\mathbf{S}_0$ during the remaining $T - s_1 - s_0$ channel uses, achieving $\min(s_1 + s_0, N_1)\frac{T-s_1-s_0}{T}$ DoF. The received signal at User 2 is

$$\begin{aligned}\mathbf{Y}_2 &= \mathbf{H}_2[\mathbf{V}_0\ \mathbf{V}_1\ \mathbf{V}_2]\mathbf{X} + \mathbf{W}_2 \\ &= \mathbf{H}_2[\mathbf{V}_0\ \mathbf{V}_2]\begin{bmatrix}\mathbf{I}_{s_0} & \mathbf{0} & [\mathbf{0}_{s_0\times(s_1-s_2)}\ \mathbf{S}_0] \\ \mathbf{0} & \mathbf{I}_{s_2} & \mathbf{S}_2\end{bmatrix} + \mathbf{W}_2.\end{aligned} \tag{2.54}$$

User 2 estimates the equivalent channel $\mathbf{H}_2[\mathbf{V}_0\ \mathbf{V}_2]$ and then decodes $\mathbf{S}_0$ and $\mathbf{S}_2$, achieving $\min(s_2, N_2)\frac{s_1-s_2}{T} + \min(s_2 + s_0, N_2)\frac{T-s_1-s_0}{T}$ DoF. By dedicating $\mathbf{S}_0$ to only User 1 or User 2, DoF pairs $\mathcal{D}_3$ and $\mathcal{D}_4$ are achieved, respectively.

$\mathcal{D}_5$ can be achieved (still assuming $s_1 \geq s_2$), via a combination of rate splitting and product superposition as follows. The transmitted signal is

$$\mathbf{X} = [\mathbf{V}_0\ \mathbf{V}_1]\mathbf{X}_2'\mathbf{X}_1 + \mathbf{V}_2\mathbf{X}_2, \tag{2.55}$$

with $\mathbf{X}_2 = [\mathbf{0}_{s_0\times s_0}\ \mathbf{I}_{s_2}\ \mathbf{S}_2]$, $\mathbf{X}_1 = [\mathbf{I}_{s_1+s_0}\ \mathbf{S}_1]$, and $\mathbf{X}_2' = \begin{bmatrix}\mathbf{I}_{s_0} & [\mathbf{0}_{s_0\times s_2}\ \mathbf{S}_2'] \\ \mathbf{0}_{s_1\times s_0} & \mathbf{I}_{s_1}\end{bmatrix}$, where $\mathbf{S}_1 \in \mathbb{C}^{(s_1+s_0)\times(T-s_1-s_0)}$ contains symbols intended for User 1 while $\mathbf{S} \in \mathbb{C}^{s_2\times(T-s_2-s_0)}$ and $\mathbf{S}_2' \in \mathbb{C}^{s_0\times(s_1-s_2)}$ contain symbols intended for User 2. The received signal at User 1 is

$$\begin{aligned}\mathbf{Y}_1 &= \mathbf{H}_1[\mathbf{V}_0\ \mathbf{V}_1]\mathbf{X}_2'\mathbf{X}_1 + \mathbf{W}_1 \\ &= \mathbf{H}_1[\mathbf{V}_0\ \mathbf{V}_1]\mathbf{X}_2'[\mathbf{I}_{s_1+s_0}\ \mathbf{S}_1] + \mathbf{W}_1.\end{aligned} \tag{2.56}$$

User 1 estimates the equivalent channel $\mathbf{H}_1[\mathbf{V}_0\ \mathbf{V}_1]\mathbf{X}_2'$, and then decodes $\mathbf{S}_1$ to achieve $\min(s_0+s_1, N_1)\frac{T-s_0-s_1}{T}$ DoF. The received signal at User 2 is

$$\begin{aligned}\mathbf{Y}_2 &= \mathbf{H}_2[\mathbf{V}_0\ \mathbf{V}_1\ \mathbf{V}_2]\mathbf{X} + \mathbf{W}_2 \\ &= \mathbf{H}_2[\mathbf{V}_0\ \mathbf{V}_2]\begin{bmatrix} \mathbf{I}_{s_0} & \mathbf{0}_{s_2\times s_2} & [\mathbf{S}_2'\ \mathbf{A}] \\ \mathbf{0}_{s_0\times s_0} & \mathbf{I}_{s_2} & \mathbf{S}_2 \end{bmatrix} + \mathbf{W}_2,\end{aligned} \tag{2.57}$$

where $\mathbf{A} \triangleq [\mathbf{I}_{s_0}\ \mathbf{0}_{s_0\times s_2}\ \mathbf{S}_2']\mathbf{S}_1$. User 2 estimates its equivalent channel $\mathbf{H}_2[\mathbf{V}_0\ \mathbf{V}_2]$ in the first s_2+s_0 channel uses, and then decodes $\mathbf{S}_2'$ and $\mathbf{S}_2$, achieving $\min(s_2+s_0, N_2)\frac{s_1-s_2}{T} + \min(s_2, N_2)\frac{T-s_1-s_0}{T}$ DoF in total. Therefore, $\mathcal{D}_5$ is achieved.

Therefore, the proof for the case where $s_1 \geq s_2$ is completed. A similar analysis applies to the case $s_2 \geq s_1$ and completes the proof of Theorem 3. $\square$

In Figure 2.5.2,2.5.2, the achievable DoF region in Theorem 3 is demonstrated for the scenario where $T = 24$, $N_1 = 12$, $N_2 = 12$, $(r_1, r_2) \in \{(12, 10), (12, 12)\}$, and $r_0 \in \{0, 3, 6, 9\}$. Similar to the CSIR case, exploiting the channel correlation improves significantly the DoF region upon TDMA, especially for small r_0.

This completes the DoF analysis for the two-user case. By using both product superposition and rate splitting, achievable DoF regions were calculated for a variety of correlation structures and antenna configurations. Also, an outer bound was calculated under perfect CSIR.

2.6 Two-User Broadcast Channel: Rate Analysis

We assume no free CSIR under partially overlapping eigenspaces, and assume that $r_k \leq N_k$, $k \in \{1, 2\}$. In addition, without loss of generality $r_1 \geq r_2$.

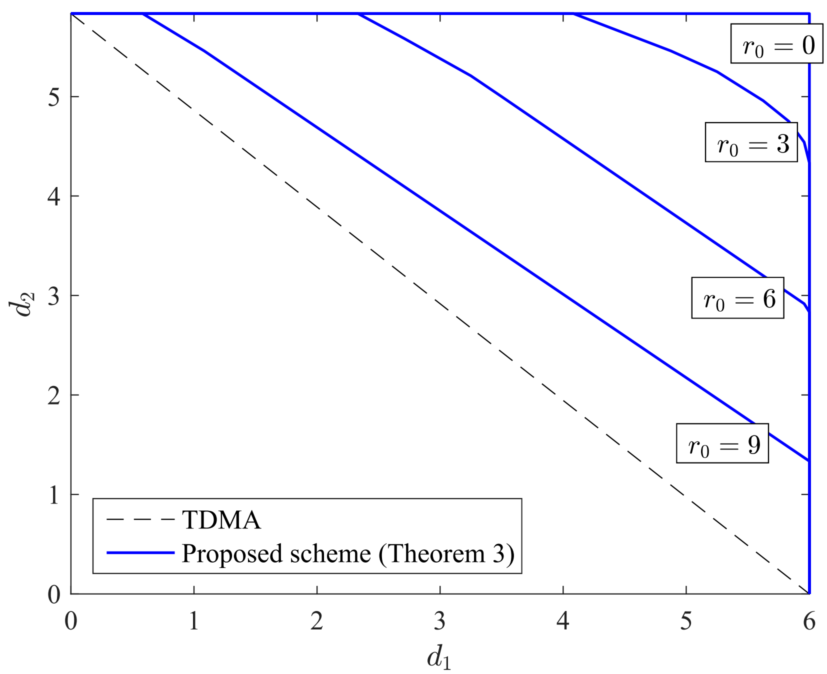

Figure 2.3. DoF region of a two-user broadcast channel without free CSIR with $T = 24$, $N_1 = 12$, $N_2 = 12$, $r_1 = 12$, $r_2 = 10$, $r_0 \in \{0, 3, 6, 9\}$

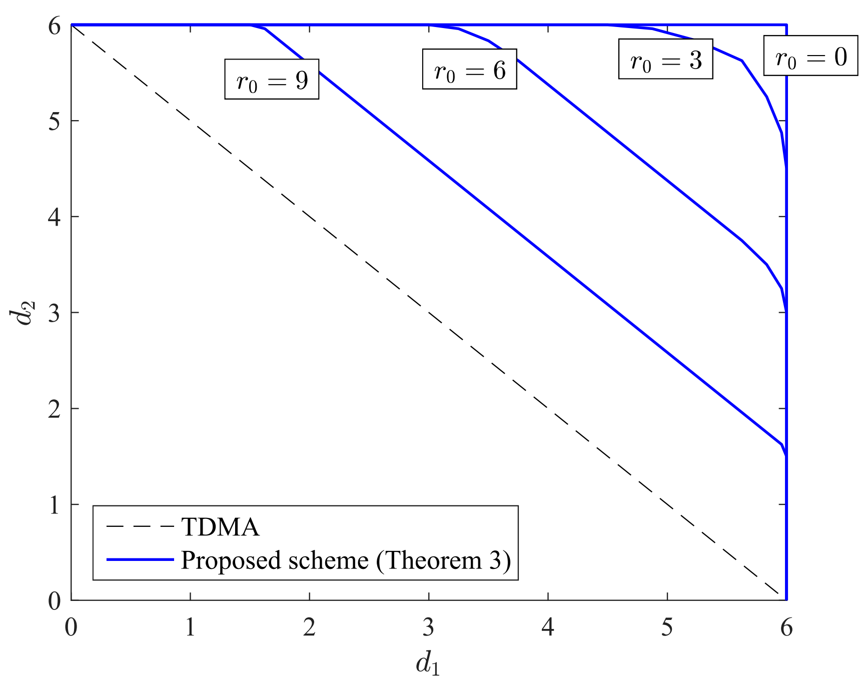

Figure 2.4. DoF region of a two-user broadcast channel without free CSIR with $T = 24$, $N_1 = 12$, $N_2 = 12$, $r_1 = 12$, $r_2 = 12$, $r_0 \in \{0, 3, 6, 9\}$.

2.6.1 The Single-User Case

Let us first consider the single-user case where, for simplicity, we omit the user's index. The received signal is

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{W}, \tag{2.58}$$

where the assumptions for the transmitted signal $\mathbf{X}$, the Gaussian noise $\mathbf{W}$, and the channel $\mathbf{H}$ are as before. In particular, $\mathbf{H}$ is block fading with coherence time T, and has correlation matrix $\mathbf{R} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^{\mathsf{H}}$, thus can be written as $\mathbf{H} = \mathbf{G}\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{U}^{\mathsf{H}}$ with $\mathbf{G} \in \mathbb{C}^{N\times r}$ drawn from a generic distribution. The following theorem states the achievable rate (in bits/channel use) for this channel.

Theorem 4. *For the single-user spatially correlated channel without free CSIR,*

1. *if the transmitter does not exploit* $\mathbf{R}$*, the following rate is achievable with a pilot-based scheme*

$$R = \Big(1 - \frac{M}{T}\Big)\mathbb{E}\bigg[\log\det\bigg(\mathbf{I}_N + \frac{\rho_\delta\rho_\tau}{\rho_\delta \mathrm{tr}\big((\mathbf{\Sigma}^{-1} + \rho_\tau\mathbf{I}_r)^{-1}\big) + M}\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathsf{H}}\bigg)\bigg], \tag{2.59}$$

where rows of $\hat{\mathbf{H}}$ *obey* $\mathcal{CN}\big(\mathbf{0}^t, \mathbf{R}(\mathbf{I}_M + \rho_\tau\mathbf{R})^{-1}\mathbf{R}\big)$ *and are independent of each other, for powers* ρ_τ *and* ρ_δ *satisfying* $\rho_\tau M + \rho_\delta(T - M) \le \rho T$*;*

2. *if the transmitter exploits* $\mathbf{R}$*, the following rate is achievable with a pilot-based scheme by transmitting in the eigenspace of* $\mathbf{R}$*:*

   - *if the transmitter uses orthogonal pilot:*

$$R = \Big(1 - \frac{r}{T}\Big)\mathbb{E}\bigg[\log\det\bigg(\mathbf{I}_N + \frac{\rho_\delta\rho_\tau}{\rho_\delta \mathrm{tr}\big((\bar{\mathbf{R}}^{-1} + \rho_\tau\mathbf{I}_r\big)^{-1}\big) + r}\hat{\mathbf{\Omega}}\hat{\mathbf{\Omega}}^{\mathsf{H}}\bigg)\bigg], \tag{2.60}$$

   where rows of $\hat{\mathbf{\Omega}}$ *obey* $\mathcal{CN}\big(\mathbf{0}^t, \bar{\mathbf{R}}(\mathbf{I}_r + \rho_\tau\bar{\mathbf{R}})^{-1}\bar{\mathbf{R}}\big)$ *and are independent of each other.*

- *if the transmitter optimizes the pilot:*

$$R = \Big(1 - \frac{r}{T}\Big)\mathbb{E}\Bigg[\log\det\bigg(\mathbf{I}_N + \frac{\rho_\delta}{r\rho_\delta\big(\rho_\tau + \frac{1}{r}\mathrm{tr}(\bar{\mathbf{R}}^{-1})\big)^{-1} + r}\hat{\mathbf{\Omega}}\hat{\mathbf{\Omega}}^{\mathsf{H}}\bigg)\Bigg], \tag{2.61}$$

where rows of $\hat{\mathbf{\Omega}}$ *obey* $\mathcal{CN}\Big(\mathbf{0}^t, \bar{\mathbf{R}} - \Big(\rho_\tau + \frac{1}{r}\mathrm{tr}\left(\bar{\mathbf{R}}^{-1}\right)\Big)^{-1}\mathbf{I}_r\Big)$ *and are independent of each other.*

The correlation matrix $\bar{\mathbf{R}} \triangleq \mathbf{V}^{\mathsf{H}}\mathbf{R}\mathbf{V}$ *for a truncated unitary matrix* $\mathbf{V} \in \mathbb{C}^{M\times r}$ *such that* $\mathrm{Span}\,(\mathbf{V}) = \mathrm{Span}\,(\mathbf{U})$*. Powers* ρ_τ *and* ρ_δ *satisfy* $\rho_\tau r + \rho_\delta(T-r) \le \rho T$*, and the optimal power allocation maximizing the rate in* (2.61) *is characterized by* $\rho_\tau = \frac{(1-\alpha)\rho T}{r}$ *and* $\rho_\delta = \frac{\alpha\rho T}{T-r}$ *with*

$$\alpha = \begin{cases} \frac{1}{2}, & \text{if } T = 2r, \\ b - \sqrt{b(b-a)}, & \text{if } T > 2r, \end{cases} \tag{2.62}$$

where $a \triangleq 1 + \frac{\mathrm{tr}(\bar{\mathbf{R}}^{-1})}{\rho T} - \frac{r^2}{\rho T\mathrm{tr}(\bar{\mathbf{R}})}$ *and* $b \triangleq \frac{T-r}{T-2r}\Big(1 + \frac{\mathrm{tr}(\bar{\mathbf{R}}^{-1})}{\rho T}\Big)$.

Proof. See Appendix. □

2.6.2 The Baseline TDMA Schemes

We consider TDMA without free CSIR. If only User k is activated and the base station does not exploit $\mathbf{R}_k$, according to Theorem 4, the following corollary demonstrates the achievable rate:

Corollary 1. *For 2-user broadcast channel, when the base station does not exploit* $\mathbf{R}_k$*, the following rate is achievable by activating only one user:*

$$R_k = \Big(1 - \frac{M}{T}\Big)\mathbb{E}\Bigg[\log\det\bigg(\mathbf{I}_{N_k} + \frac{\rho_\delta\rho_\tau}{\rho_\delta\mathrm{tr}\big((\mathbf{\Sigma}_k^{-1} + \rho_\tau\mathbf{I}_{r_k})^{-1}\big) + M}\hat{\mathbf{\Omega}}_k\hat{\mathbf{\Omega}}_k^{\mathsf{H}}\bigg)\Bigg], \tag{2.63}$$

where rows of $\hat{\mathbf{\Omega}}_k$ *obey* $\mathcal{CN}\big(\mathbf{0}^t, \mathbf{R}_k(\mathbf{I}_M + \rho_\tau\mathbf{R}_k)^{-1}\mathbf{R}_k\big)$ *and are independent of each other, for powers* ρ_τ *and* ρ_δ *satisfying* $\rho_\tau M + \rho_\delta(T-M) \le \rho T$*;*

If the base station transmits in the eigenspace of $\mathbf{R}_k$ using precoder $\mathbf{V}_k = \mathbf{U}_k$, i.e., $\mathbf{U}_k^{\mathsf{H}}\mathbf{V}_k = \mathbf{I}_{r_k}$, and optimize the pilot, the following corollary demonstrates the achievable rate:

Corollary 2. *For 2-user broadcast channel, when the base station transmits in the eigenspace of $\mathbf{R}_k$, the following rate is achievable by activating only one user:*

$$R_k = \left(1 - \frac{r_k}{T}\right)\mathbb{E}\left[\log\det\left(\mathbf{I}_{N_k} + \frac{\rho_\delta}{r\rho_\delta\left(\rho_\tau + \frac{1}{r}\mathrm{tr}(\mathbf{\Sigma}_k^{-1})\right)^{-1} + r_k}\hat{\mathbf{\Omega}}_k\hat{\mathbf{\Omega}}_k^{\mathsf{H}}\right)\right], \tag{2.64}$$

where rows of $\hat{\mathbf{\Omega}}_k$ obey $\mathcal{CN}\left(\mathbf{0}^t, \mathbf{\Sigma}_k - \left(\rho_\tau + \frac{1}{r_k}\mathrm{tr}(\mathbf{\Sigma}_k^{-1})\right)^{-1}\mathbf{I}_{r_k}\right)$ and are independent of each other, for powers ρ_τ and ρ_δ satisfying $\rho_\tau r + \rho_\delta(T-r) \leq \rho T$. Furthermore, the optimal power allocation for the rate in (2.64) *is given by $\rho_\tau r_k = (1-\alpha)\rho T$ and $\rho_\delta(T-r_k) = \alpha\rho T$ with*

$$\alpha = \begin{cases} \frac{1}{2}, & \text{if } T = 2r_k, \\ b - \sqrt{b(b-a)}, & \text{if } T > 2r_k, \end{cases} \tag{2.65}$$

where $a \triangleq 1 + \frac{\mathrm{tr}(\mathbf{\Sigma}_k^{-1})}{\rho T} - \frac{r_k^2}{\rho T \mathrm{tr}(\mathbf{\Sigma}_k)}$ and $b \triangleq \frac{T-r_k}{T-2r_k}\left(1 + \frac{\mathrm{tr}(\mathbf{\Sigma}_k^{-1})}{\rho T}\right)$.

The convex hull of $(0,0)$, $(R_1, 0)$, and $(0, R_2)$ is achievable by TDMA.

2.6.3 Rate Splitting

In the following, we analyze the rate achievable with the schemes achieving the DoF region in Theorem 3. Recall that for a set of non-negative integers $s_0 \leq r_0$, $s_1 \leq r_1 - r_0$, and $s_2 \leq r_2 - r_0$, the precoding matrices $\mathbf{V}_0, \mathbf{V}_1, \mathbf{V}_2$, are defined in (2.48)-(2.50). For $k \in \{1,2\}$, define

- $\mathbf{\Phi}_k \triangleq \mathbf{U}_k^{\mathsf{H}}[\mathbf{V}_0 \ \mathbf{V}_k]$, $\mathbf{\Phi}_{k0} \triangleq \mathbf{U}_k^{\mathsf{H}}\mathbf{V}_0$, $\mathbf{\Phi}_{kk} \triangleq \mathbf{U}_k^{\mathsf{H}}\mathbf{V}_k$ (so $\mathbf{\Phi}_k = [\mathbf{\Phi}_{k0} \ \mathbf{\Phi}_{kk}]$);
- $\bar{\mathbf{R}}_k \triangleq \mathbf{\Phi}_k^{\mathsf{H}}\mathbf{\Sigma}_k\mathbf{\Phi}_k$, $\bar{\mathbf{R}}_{k0} \triangleq \mathbf{\Phi}_k^{\mathsf{H}}\mathbf{\Sigma}_k\mathbf{\Phi}_{k0}$, $\bar{\mathbf{R}}_{kk} \triangleq \mathbf{\Phi}_k^{\mathsf{H}}\mathbf{\Sigma}_k\mathbf{\Phi}_{kk}$ (so $\bar{\mathbf{R}}_k = [\bar{\mathbf{R}}_{k0} \ \bar{\mathbf{R}}_{kk}]$);

- $\breve{\mathbf{R}}_{k0} \triangleq \mathbf{\Phi}_{k0}^{\mathsf{H}} \mathbf{\Sigma}_k \mathbf{\Phi}_{k0}$, $\breve{\mathbf{R}}_{kk} \triangleq \mathbf{\Phi}_{kk}^{\mathsf{H}} \mathbf{\Sigma}_k \mathbf{\Phi}_{kk}$.

Let the base station transmit

$$\mathbf{X} = \mathbf{V}_0 \mathbf{X}_0 + \mathbf{V}_1 \mathbf{X}_1 + \mathbf{V}_2 \mathbf{X}_2, \tag{2.66}$$

where $\mathbf{X}_0$, $\mathbf{X}_1$, and $\mathbf{X}_2$ are independent and satisfy the power constraint

$$\mathbb{E}\left[\|\mathbf{X}_0\|_F^2 + \|\mathbf{X}_1\|_F^2 + \|\mathbf{X}_2\|_F^2\right] \leq \rho T. \tag{2.67}$$

Thanks to the precoders, the private signal $\mathbf{X}_k$ is seen by User k only, while the common signal $\mathbf{X}_0$ is seen by both users. The received signals become

$$\mathbf{Y}_1 = \mathbf{G}_1 \mathbf{\Sigma}_1^{\frac{1}{2}} \mathbf{\Phi}_{10} \mathbf{X}_0 + \mathbf{G}_1 \mathbf{\Sigma}_1^{\frac{1}{2}} \mathbf{\Phi}_{11} \mathbf{X}_1 + \mathbf{W}_1, \tag{2.68}$$

$$\mathbf{Y}_2 = \mathbf{G}_2 \mathbf{\Sigma}_2^{\frac{1}{2}} \mathbf{\Phi}_{20} \mathbf{X}_0 + \mathbf{G}_2 \mathbf{\Sigma}_2^{\frac{1}{2}} \mathbf{\Phi}_{22} \mathbf{X}_2 + \mathbf{W}_2, \tag{2.69}$$

where the equivalent channels $\mathbf{G}_k \mathbf{\Sigma}_k^{\frac{1}{2}} \mathbf{\Phi}_{k0} \in \mathbb{C}^{N_k \times s_0}$ and $\mathbf{G}_k \mathbf{\Sigma}_k^{\frac{1}{2}} \mathbf{\Phi}_{kk} \in \mathbb{C}^{N_k \times s_k}$, $k \in \{1, 2\}$, are correlated and unknown. It can be observed that the received signal at each user is similar to a non-coherent two-user MAC: (2.68) as the MAC 1 with (s_0, s_1) equivalent transmit antennas and N_1 receive antennas, (2.69) as the MAC 2 with (s_0, s_2) equivalent transmit antennas and N_2 receive antennas. The two MACs share a common signal $\mathbf{X}_0$.

From the capacity region of multiple access channels [61], we know that the rate pairs (R_0, R_1^p) and (R_0, R_2^p) are simultaneously achievable for the MAC 1 and MAC 2, respectively, if the rates $R_0 \geq 0, R_1^p \geq 0, R_2^p \geq 0$ satisfy

$$R_0 \leq \frac{1}{T} I(\mathbf{Y}_1; \mathbf{X}_0 | \mathbf{X}_1), \tag{2.70}$$

$$R_1^p \leq \frac{1}{T} I(\mathbf{Y}_1; \mathbf{X}_1 | \mathbf{X}_0), \tag{2.71}$$

$$R_0 + R_1^p \leq \frac{1}{T} I(\mathbf{Y}_1; \mathbf{X}_0, \mathbf{X}_1), \tag{2.72}$$

$$R_0 \leq \frac{1}{T} I(\mathbf{Y}_2; \mathbf{X}_0 | \mathbf{X}_2), \tag{2.73}$$

$$R_2^p \leq \frac{1}{T} I(\mathbf{Y}_2; \mathbf{X}_2 | \mathbf{X}_0), \tag{2.74}$$

$$R_0 + R_1^p \leq \frac{1}{T} I(\mathbf{Y}_2; \mathbf{X}_0, \mathbf{X}_2). \tag{2.75}$$

Then, User 1 achieves rate R_1^p with private signal $\mathbf{X}_1$, user 2 achieves rate R_2^p with private signal $\mathbf{X}_2$, and both users can achieve rate R_0 with common signal $\mathbf{X}_0$. Let R_{0k} be the User k's share in R_0, then the rate pair $(R_1, R_2) = (R_{01} + R_1^p, R_{02} + R_2^p)$ is achievable. Replacing $R_0 = R_{01} + R_{02}$, $R_1^p = R_1 - R_{01}$, and $R_2^p = R_2 - R_{02}$ in (2.70)-(2.75) and applying Fourier-Motzkin elimination leads to the following result.

Lemma 4. *With rate splitting and without free CSIR, the rate pairs (R_1, R_2) are achievable with:*

$$R_1 \leq \frac{1}{T} \min\{I(\mathbf{Y}_1; \mathbf{X}_1, \mathbf{X}_0), I(\mathbf{Y}_1; \mathbf{X}_1|\mathbf{X}_0) + I(\mathbf{Y}_2; \mathbf{X}_0|\mathbf{X}_2)\}, \tag{2.76}$$

$$R_2 \leq \frac{1}{T} \min\{I(\mathbf{Y}_2; \mathbf{X}_2, \mathbf{X}_0), I(\mathbf{Y}_2; \mathbf{X}_2|\mathbf{X}_0) + I(\mathbf{Y}_1; \mathbf{X}_0|\mathbf{X}_1)\}, \tag{2.77}$$

$$R_1 + R_2 \leq \frac{1}{T} \min\{I(\mathbf{Y}_1; \mathbf{X}_1|\mathbf{X}_0) + I(\mathbf{Y}_2; \mathbf{X}_2, \mathbf{X}_0), I(\mathbf{Y}_1; \mathbf{X}_1, \mathbf{X}_0) + I(\mathbf{Y}_2; \mathbf{X}_2|\mathbf{X}_0)\}, \tag{2.78}$$

for input distributions $p(\mathbf{X}_0)$, $p(\mathbf{X}_1)$, and $p(\mathbf{X}_2)$ satisfying $\mathbb{E}\left[\|\mathbf{X}_0\|_F^2 + \|\mathbf{X}_1\|_F^2 + \|\mathbf{X}_2\|_F^2\right] \leq \rho T$.

By bounding the mutual information terms in Lemma 4, we have the following theorem:

Theorem 5. *Under rate splitting, the following rate region can be achieved in the two-user correlated broadcast channel with partially overlapped eigenspaces:*

$$R_1 \leq \min\{R_1', R_1^p + R_0''\}, \tag{2.79}$$

$$R_2 \leq \min\{R_2', R_2^p + R_0'\}, \tag{2.80}$$

$$R_1 + R_2 \leq \min\{R_1^p + R_2', R_1' + R_2^p\}, \tag{2.81}$$

where

$$R_1' = \Big(1 - \frac{s_1 + s_0}{T}\Big) \mathbb{E}\left[\log\det\left(\mathbf{I}_{N_1} + \frac{1}{\mathrm{tr}\big(\big(\bar{\mathbf{R}}_1^{-1} + \mathbf{P}_{1\tau}\big)^{-1}\mathbf{P}_{1\delta}\big) + 1} \bar{\mathbf{\Omega}}_1 \bar{\mathbf{R}}_1 \mathbf{P}_{1\delta} \bar{\mathbf{R}}_1^{\mathsf{H}} \bar{\mathbf{\Omega}}_1^{\mathsf{H}}\right)\right], \tag{2.82}$$

$$R_1^p = \Big(1 - \frac{s_1 + s_0}{T}\Big)\mathbb{E}\Big[\log\det\Big(\mathbf{I}_{N_1} + \frac{\rho_{1\delta}}{s_1\big[\mathrm{tr}\big((\bar{\mathbf{R}}_1^{-1} + \mathbf{P}_{1\tau})^{-1}\mathbf{P}_{1\delta}\big) + 1\big]}\bar{\mathbf{\Omega}}_1\bar{\mathbf{R}}_{11}\bar{\mathbf{R}}_{11}^{\mathsf{H}}\bar{\mathbf{\Omega}}_1^{\mathsf{H}}\Big)\Big], \tag{2.83}$$

$$R_1'' = \Big(1 - \frac{s_1 + s_0}{T}\Big)\mathbb{E}\Big[\log\det\Big(\mathbf{I}_{N_1} + \frac{\rho_{0\delta}}{s_0\big[\mathrm{tr}\big((\bar{\mathbf{R}}_1^{-1} + \mathbf{P}_{1\tau}\big)^{-1}\mathbf{P}_{1\delta}\big) + 1\big]}\bar{\mathbf{\Omega}}_1\bar{\mathbf{R}}_{10}\bar{\mathbf{R}}_{10}^{\mathsf{H}}\bar{\mathbf{\Omega}}_1^{\mathsf{H}}\Big)\Big], \tag{2.84}$$

where rows of $\bar{\mathbf{\Omega}}_1$ *obey* $\mathcal{CN}\big(\mathbf{0}^t, \mathbf{P}_{1\tau}^{\frac{1}{2}}(\mathbf{P}_{1\tau}^{\frac{1}{2}}\bar{\mathbf{R}}_1\mathbf{P}_{1\tau}^{\frac{1}{2}} + \mathbf{I}_{s_1+s_0})^{-1}\mathbf{P}_{1\tau}^{\frac{1}{2}}\big)$ *and are independent of each other.*

$$\begin{aligned} R_2' = {}& \frac{s_1 - s_2}{T}\mathbb{E}\Big[\log\det\Big(\mathbf{I}_{N_2} + \frac{\rho_{2\delta}}{\rho_{2\delta}\mathrm{tr}\big(\bar{\mathbf{R}}_{22}^{\mathsf{H}}(\bar{\mathbf{R}}_2 + \bar{\mathbf{R}}_2\mathbf{P}_{2\tau}\bar{\mathbf{R}}_2)^{-1}\bar{\mathbf{R}}_{22}\big) + s_2}\bar{\mathbf{\Omega}}_2\bar{\mathbf{R}}_{22}\bar{\mathbf{R}}_{22}^{\mathsf{H}}\bar{\mathbf{\Omega}}_2^{\mathsf{H}}\Big)\Big] \\ & + \Big(1 - \frac{s_1 + s_0}{T}\Big)\mathbb{E}\Big[\log\det\Big(\mathbf{I}_{N_2} + \frac{1}{\mathrm{tr}\big((\bar{\mathbf{R}}_2^{-1} + \mathbf{P}_{2\tau})^{-1}\mathbf{P}_{2\delta}\big) + 1}\bar{\mathbf{\Omega}}_2\bar{\mathbf{R}}_2\mathbf{P}_{2\delta}\bar{\mathbf{R}}_2^{\mathsf{H}}\bar{\mathbf{\Omega}}_2^{\mathsf{H}}\Big)\Big], \end{aligned} \tag{2.85}$$

$$\begin{aligned} R_2^p = {}& \frac{s_1 - s_2}{T}\mathbb{E}\Big[\log\det\Big(\mathbf{I}_{N_2} + \frac{\rho_{2\delta}}{\rho_{2\delta}\mathrm{tr}\big(\bar{\mathbf{R}}_{22}^{\mathsf{H}}(\bar{\mathbf{R}}_2 + \bar{\mathbf{R}}_2\mathbf{P}_{2\tau}\bar{\mathbf{R}}_2)^{-1}\bar{\mathbf{R}}_{22}\big) + s_2}\bar{\mathbf{\Omega}}_2\bar{\mathbf{R}}_{22}\bar{\mathbf{R}}_2\bar{\mathbf{R}}_2^{\mathsf{H}}\bar{\mathbf{R}}_{22}^{\mathsf{H}}\bar{\mathbf{\Omega}}_2^{\mathsf{H}}\Big)\Big] \\ & + \Big(1 - \frac{s_1 + s_0}{T}\Big)\mathbb{E}\Big[\log\det\Big(\mathbf{I}_{N_2} + \frac{\rho_{2\delta}}{s_2\big[\mathrm{tr}\big((\bar{\mathbf{R}}_2^{-1} + \mathbf{P}_{2\tau})^{-1}\mathbf{P}_{2\delta}\big) + 1\big]}\bar{\mathbf{\Omega}}_2\bar{\mathbf{R}}_{22}\bar{\mathbf{R}}_2\bar{\mathbf{R}}_2^{\mathsf{H}}\bar{\mathbf{R}}_{22}^{\mathsf{H}}\bar{\mathbf{\Omega}}_2^{\mathsf{H}}\Big)\Big], \end{aligned} \tag{2.86}$$

$$R_0'' = \Big(1 - \frac{s_1 + s_0}{T}\Big)\mathbb{E}\Big[\log\det\Big(\mathbf{I}_{N_2} + \frac{\rho_{0\delta}}{s_0\big[\mathrm{tr}\big((\bar{\mathbf{R}}_2^{-1} + \mathbf{P}_{2\tau})^{-1}\mathbf{P}_{2\delta}\big) + 1\big]}\bar{\mathbf{\Omega}}_2\bar{\mathbf{R}}_{20}\bar{\mathbf{R}}_2\bar{\mathbf{R}}_2^{\mathsf{H}}\bar{\mathbf{R}}_{20}^{\mathsf{H}}\bar{\mathbf{\Omega}}_2^{\mathsf{H}}\Big)\Big], \tag{2.87}$$

where rows of $\bar{\mathbf{\Omega}}_2$ *obey* $\mathcal{CN}\big(\mathbf{0}^t, \mathbf{P}_{2\tau}^{\frac{1}{2}}(\mathbf{P}_{2\tau}^{\frac{1}{2}}\bar{\mathbf{R}}_2\mathbf{P}_{2\tau}^{\frac{1}{2}} + \mathbf{I}_{s_2+s_0})^{-1}\mathbf{P}_{2\tau}^{\frac{1}{2}}\big)$ *and are independent of each other.* s_0, s_1, s_2 *are designed to allocate transmit dimensions to the components of product* *superposition, and take values in the range* $s_0 \le r_0$*,* $s_1 \le r_1 - r_0$ *and* $s_2 \le r_2 - r_0$*. The component powers* $\rho_{\tau 0}, \rho_{\tau 1}, \rho_{\tau 2}, \rho_\delta$ *satisfy the power constraint*

$$\rho_{0\tau} s_0 + \rho_{0\delta}(T - s_1 - s_0) + \sum_{i=1}^{2} \left[\rho_{i\tau} s_i + \rho_{i\delta}(T - s_i - s_0)\right] \le \rho T. \tag{2.88}$$

The overall achievable rate region is the convex hull of (2.79),(2.80),(2.81) *over all feasible values of* s_0, s_1, s_2 *and power allocations* (2.88).

Proof. Please see the Appendix. □

2.6.4 Product Superposition

Theorem 6. *With product superposition, the following rate pair* (R_1, R_2) *can be achieved:*

$$R_1 = \frac{s_2}{T} \mathbb{E}\left[\log\det\left(\mathbf{I}_{N_1} + \frac{\nu_{1\delta}\rho_{2\tau}}{s_0 + \nu_{1\delta}\rho_{2\tau}\mathrm{tr}\big((\breve{\mathbf{R}}_{10}^{-1} + \nu_{1\tau}\rho_{2\tau}\mathbf{I}_{s_0})^{-1}\big)} \hat{\mathbf{\Omega}}_{10}\hat{\mathbf{\Omega}}_{10}^{\mathsf{H}}\right)\right], \tag{2.89}$$

where rows of $\hat{\mathbf{\Omega}}_{10}$ *obey* $\mathcal{CN}\big(\mathbf{0}^t, \nu_{1\tau}\rho_{2\tau}\breve{\mathbf{R}}_{10}(\nu_{1\tau}\rho_{2\tau}\breve{\mathbf{R}}_{10} + \mathbf{I}_{s_0})^{-1}\breve{\mathbf{R}}_{10}\big)$ *and are independent of each other.*

$$R_2 = \left(1 - \frac{s_2 + s_0}{T}\right)\mathbb{E}\left[\log\det\left(\mathbf{I}_{N_2} + \frac{\rho_{2\delta}}{s_2 + s_0 + \rho_{2\delta}\mathrm{tr}\big((\mathbf{R}_{2e}^{-1} + \rho_{2\tau}\mathbf{I}_{s_2+s_0})^{-1}\big)} \hat{\mathbf{G}}_{2e}\hat{\mathbf{G}}_{2e}^{\mathsf{H}}\right)\right], \tag{2.90}$$

where rows of $\hat{\mathbf{G}}_{2e}$ *obey* $\mathcal{CN}\big(\mathbf{0}^t, \rho_{2\tau}\mathbf{R}_{2e}\big(\rho_{2\tau}\mathbf{R}_{2e} + \mathbf{I}_{s_2+s_0}\big)^{-1}\mathbf{R}_{2e}\big)$ *and are independent of each other, with*

$$\mathbf{R}_{2e} \triangleq \begin{bmatrix} \nu_{1\tau}\breve{\mathbf{R}}_{20} & \sqrt{\nu_{1\tau}\nu_{1a}}\mathbf{\Phi}_{20}^{\mathsf{H}}\mathbf{\Sigma}_2\mathbf{\Phi}_{22} \\ \sqrt{\nu_{1\tau}\nu_{1a}}\mathbf{\Phi}_{22}^{\mathsf{H}}\mathbf{\Sigma}_2\mathbf{\Phi}_{20} & \frac{\nu_{1\delta}}{s_0}\mathrm{tr}\left(\breve{\mathbf{R}}_{20}\right)\mathbf{I}_{s_2} + \nu_{1a}\breve{\mathbf{R}}_{22} \end{bmatrix}. \tag{2.91}$$

s_0, s_1, s_2 *are designed to allocate transmit dimensions to the components of product superposition, and take values in the range* $s_0 \le r_0$ *and* $s_2 \le r_2 - r_0$ *with the power constraint*

$$\big(s_0\nu_{1\tau} + s_2(\nu_{1\delta} + \nu_{1a})\big)\left(\rho_{2\tau} + \frac{T - s_2 - s_0}{s_2 + s_0}\rho_{2\delta}\right) \le \rho T. \tag{2.92}$$

By swapping the users' roles, another achievable rate pair is obtained. The overall achievable rate region is the convex hull of these pairs over all feasible values of s_0, s_1, s_2 *and feasible power allocations* (2.92).

Proof. Please see the Appendix. $\square$

2.6.5 Hybrid Superposition

Hybrid superposition in this chapter refers to a composite scheme that involves both rate splitting and product superposition.

Theorem 7. *With hybrid superposition, the following rate pair* (R_1, R_2) *can be achieved:*

$$R_1 = \Big(1 - \frac{s_1 + s_0}{T}\Big)\mathbb{E}\Bigg[\log\det\Bigg(\mathbf{I}_{N_1} + \frac{\rho_{1\delta}}{s_1 + s_0 + \rho_{1\delta}\mathrm{tr}\big((\mathbf{R}_{1e}^{-1} + \rho_{1\tau}\mathbf{I}_{s_1+s_0})^{-1}\big)}\hat{\mathbf{G}}_{1e}\hat{\mathbf{G}}_{1e}^{\mathsf{H}}\Bigg)\Bigg], \tag{2.93}$$

where rows of $\hat{\mathbf{G}}_{1e}$ *obey* $\mathcal{CN}\big(\mathbf{0}^t, \rho_{1\tau}\mathbf{R}_{1e}\big(\rho_{1\tau}\mathbf{R}_{1e} + \mathbf{I}_{s_1+s_0}\big)^{-1}\mathbf{R}_{1e}\big)$ *and are independent of each other, with*

$$\mathbf{R}_{1e} \triangleq \begin{bmatrix} \nu_{2\tau}\breve{\mathbf{R}}_{10} & \sqrt{\nu_{2\tau}\nu_{2a}}\boldsymbol{\Phi}_{10}^{\mathsf{H}}\boldsymbol{\Sigma}_1\boldsymbol{\Phi}_{11} \\ \sqrt{\nu_{2\tau}\nu_{2a}}\boldsymbol{\Phi}_{11}^{\mathsf{H}}\boldsymbol{\Sigma}_1\boldsymbol{\Phi}_{10} & \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{\nu_{2\delta}}{s_0}\mathrm{tr}\left(\breve{\mathbf{R}}_{10}\right)\mathbf{I}_{s_1-s_2} \end{bmatrix} + \nu_{2a}\breve{\mathbf{R}}_{22} \end{bmatrix}, \tag{2.94}$$

$$\begin{aligned} R_2 = {} & \frac{s_1 - s_2}{T}\mathbb{E}\Bigg[\log\det\Bigg(\mathbf{I}_{N_2} + \frac{1}{\mathrm{tr}\left(\left(\bar{\mathbf{R}}_2^{-1} + \mathbf{P}_{2\tau}\right)^{-1}\mathbf{P}_{2\delta a}\right) + 1}\bar{\boldsymbol{\Omega}}_2\mathbf{P}_{2\delta a}\bar{\boldsymbol{\Omega}}_2^{\mathsf{H}}\Bigg)\Bigg] \\ & + \Big(1 - \frac{s_1 + s_0}{T}\Big)\mathbb{E}\Bigg[\log\det\Bigg(\mathbf{I}_{N_2} + \frac{1}{\mathrm{tr}\left(\left(\bar{\mathbf{R}}_2^{-1} + \mathbf{P}_{2\tau}\right)^{-1}\mathbf{P}_{2\delta b}\right) + 1}\bar{\boldsymbol{\Omega}}_2\mathbf{P}_{2\delta b}\bar{\boldsymbol{\Omega}}_2^{\mathsf{H}}\Bigg)\Bigg] \\ & - \Big(1 - \frac{s_1 + s_0}{T}\Big)\mathbb{E}\Bigg[\log\det\Bigg(\mathbf{I}_{N_2} + \rho_{1\delta}\Big(\nu_{2\tau} + \nu_{2\delta}\frac{s_1 - s_2}{s_0}\Big)\bar{\boldsymbol{\Omega}}_{20}\bar{\boldsymbol{\Omega}}_{20}^{\mathsf{H}}\Bigg)\Bigg], \end{aligned} \tag{2.95}$$

where rows of $\bar{\boldsymbol{\Omega}}_2$ *obey* $\mathcal{CN}\big(\mathbf{0}^t, \bar{\mathbf{R}}_2^{\mathsf{H}}\big(\bar{\mathbf{R}}_2 + \mathbf{P}_{2\tau}^{-1}\big)^{-1}\bar{\mathbf{R}}_2\big)$ *and* $\bar{\boldsymbol{\Omega}}_{20}$ *obey* $\mathcal{CN}\big(\mathbf{0}^t, \breve{\mathbf{R}}_{20}\big)$, *and are independent of each other.* s_0, s_1, s_2 *are designed to allocate transmit dimensions to the* *components of product superposition, and take values in the range* $s_0 \leq r_0$, $s_1 \leq r_1 - r_0$ *and* $s_2 \leq r_2 - r_0$ *with the power constraint*

$$\left(s_0\nu_{2\tau} + s_1\nu_{2a} + (s_1 - s_2)\nu_{2\delta}\right)\left(\rho_{1\tau} + \frac{T - s_1 - s_0}{s_1 + s_0}\rho_{1\delta}\right) + s_2\rho_{2\tau} + (T - s_2 - s_0)\rho_{2\delta} \leq \rho T. \tag{2.96}$$

The overall achievable rate region is the convex hull of these pairs over all possible power allocations satisfying the power constraint and all feasible values of s_0, s_1, s_2.

Proof. Please see the Appendix. □

Remark 2. *Hybrid superposition utilizes both rate splitting and product superposition but is not a generalization, in the sense that the results of pure rate splitting and product superposition cannot be recovered from the hybrid scheme. At very high SNR under partially overlapped eigenspaces, hybrid superposition can improve over rate splitting and product superposition, but in other channel conditions, the hybrid superposition may in fact perform worse than the individual schemes.*

2.6.6 Numerical Results

Simulations in this section assume Rayleigh fading, i.e., $\mathbf{G}_k$ has independent $\mathcal{CN}(0,1)$ entries. The correlation matrix $\mathbf{R}_k = \mathbf{U}_k\mathbf{\Sigma}_k\mathbf{U}_k^{\mathsf{H}}$, $k \in \{1,2\}$, is generated by assuming the same magnitude along all eigendirections, i.e., $\mathbf{\Sigma}_k = \mathbf{I}$. Furthermore, we assume the eigendirections of transmit correlation matrices of the two users are either the same or orthogonal to each other. The simplicity of this configuration makes it suitable for a representative example. Assuming a constant magnitude along different eigendirections allows us to concentrate on gains that are *purely* due to correlation diversity rather than, e.g., water-filling.

When the eigenspaces of the two users are partially overlapped, in Figure 2.5 and 2.6, we plot the rate regions achieved with these schemes in a setting of $T = 24$, $M = 16$,

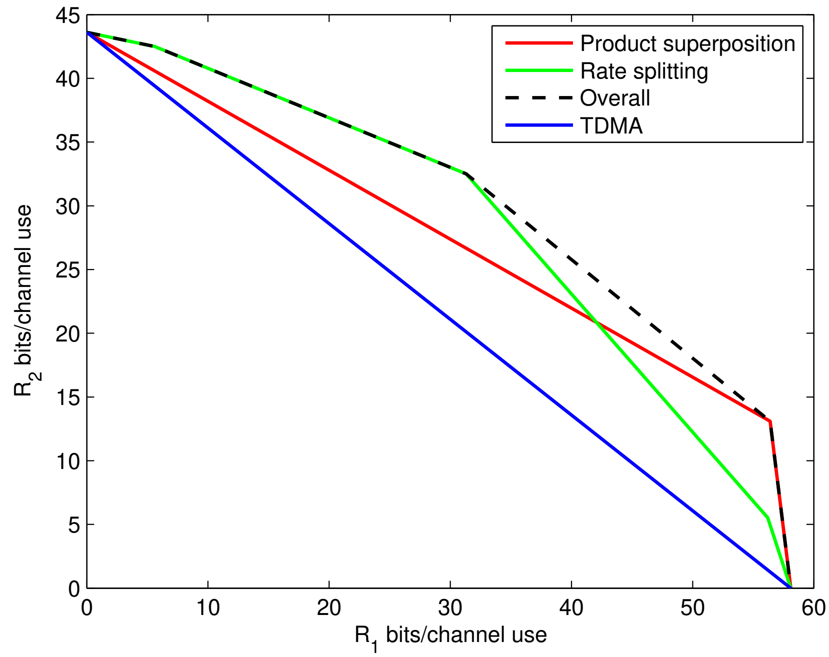

Figure 2.5. Rate region broadcast channel $\rho = 30$ dB, $r_1 = 15$, $r_2 = 8$, $r_0 = 7$

$N_1 = N_2 = 12$, $r_1 = 16$, $r_0 = 6$, $r_2 = 10$ and $T = 32$, $M = N_1 = N_2 = 16$, $r_1 = 15$, $r_0 = 7$, $r_2 = 8$, at power constraint $\rho = 30$ dB. We observe that the performance of rate splitting and product superposition depends strongly on the rank of the eigenspaces. When the rank of the two individual eigenspaces is close to each other, rate splitting will obtain a better rate region since the gains achieved by product superposition come from the difference between the ranks of the two eigenspaces. In the channel configuration in Figure 2.5 and 2.6, the hybrid superposition scheme produced rates that are inferior to both product superposition and to rate splitting, therefore they are not displayed. Hybrid superposition becomes competitive at very high SNR, while the results of this section focus on moderate SNR.

When one of the users' eigenspaces is strictly a subspace of the other, rate splitting performs no better than TDMA. In Figure 2.7, we plot the rate region for this scenario achieved via product superposition in the setting of $T = 20$, $M = N_1 = N_2 = 10$, $r_1 = 10, r_2 = 5, r_0 = 5$ and at power constraint $\rho = 30$ dB .

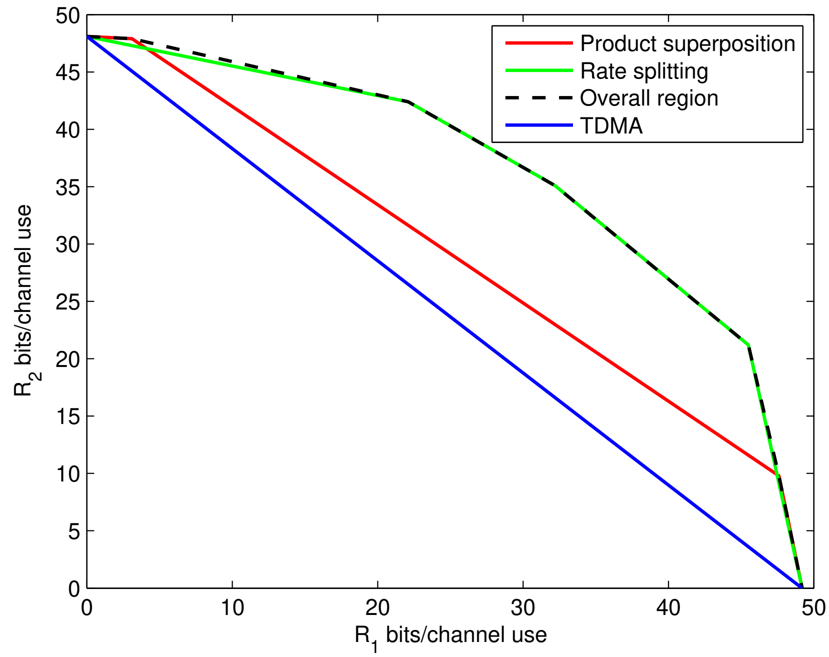

Figure 2.6. Rate region broadcast channel $\rho = 30$ dB, $r_1 = 16$, $r_2 = 10$, $r_0 = 6$

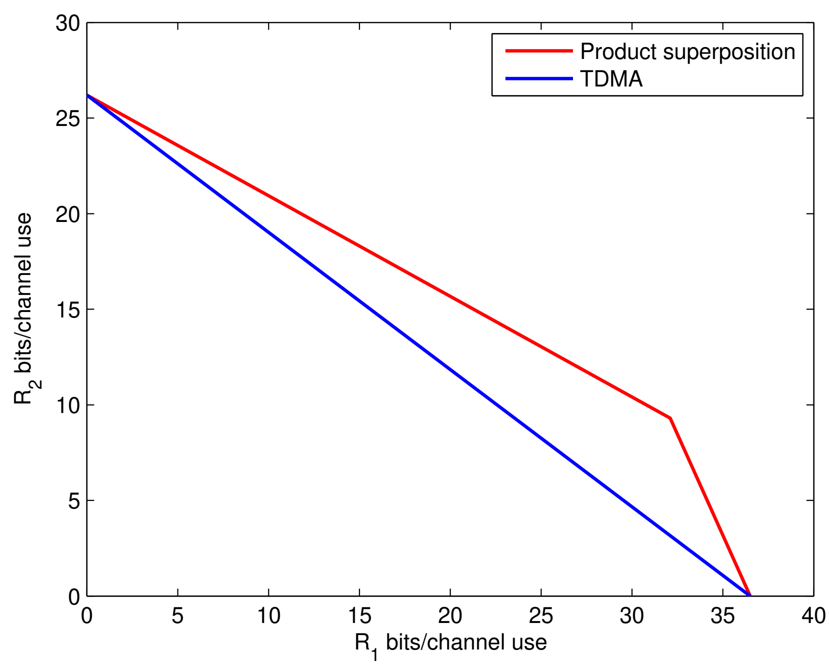

Figure 2.7. Rate region of a broadcast channel where $\rho = 30dB, M = N_1 = N_2 = 10, T = 20, r_1 = 10, r_0 = 5, r_2 = 5$

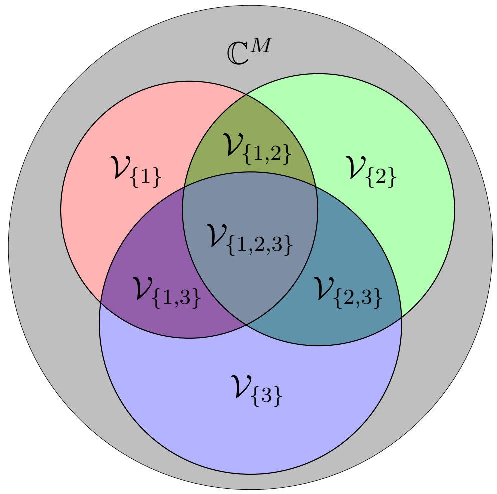

Figure 2.8. The channel eigenspace overlapping structure of the three-user broadcast channel

2.7 K-user Broadcast Channel: DoF Analysis

To extend the study to the K-user scenario, some further assumptions on the correlation model are made as follows. Recall that the rows of $\mathbf{H}_k$ belong to the eigenspace $\mathrm{Span}\left(\mathbf{U}_k\right)$ of $\mathbf{R}_k$. Denote the union of all channel eigenspaces as

$$\mathcal{V} = \bigcup_{k\in[K]} \mathrm{Span}\left(\mathbf{U}_k\right). \tag{2.97}$$

$\mathcal{V}$ can be partitioned into $2^K - 1$ subspaces $\mathcal{V}_{\mathcal{J}}$ of $r_{\mathcal{J}}$ dimensional whose $r_{\mathcal{J}}$ basis vectors span the channel of every user in a non-empty group $\mathcal{J} \subset [K]$ and are orthogonal to all vectors in $\mathrm{Span}\left(\mathbf{U}_k\right)$ for $k \in \{[K] \setminus \mathcal{J}\}$. In other words, $\mathcal{V}_{\mathcal{J}} = \bigcap_{k\in\mathcal{J}} \mathrm{Span}\left(\mathbf{U}_k\right)$. Obviously, $\sum_{\mathcal{J}\subset[K]} r_{\mathcal{J}} = \mathrm{rank}\left(\mathcal{V}\right) \leq M$ and $\sum_{\mathcal{J}\subset[K]:\ k\in\mathcal{J}} r_{\mathcal{J}} = \mathrm{rank}\left(\mathrm{Span}\left(\mathbf{U}_k\right)\right) = r_k$. An example of the correlation structure for the case of three-user broadcast channel is presented in Figure 2.8.

In this way, the signal transmitted in the subspace $\mathcal{V}_{\mathcal{J}}$ can be seen by every user in $\mathcal{J}$ and is vague to all other users. On the other hand, the signals transmitted in $\mathcal{V}_{\mathcal{J}}$ and $\mathcal{V}_{\mathcal{K}}$ interfere each other at every user in $\mathcal{J} \cap \mathcal{K}$. To characterize the interfering relation between signals transmitted in different subspaces, we introduce the concept of *interference graph* as follows:

Definition 1. *For $k \in [K]$, the* interference graph *of order k, denoted by $G(K,k)$, is an undirected graph for which:*

- *the set of vertices is the set of unordered subsets of cardinality k of $[K]$, i.e., $\mathcal{J} \subset [K] : |\mathcal{J}| = k$, hence a vertex is also denoted by a subset $\mathcal{J}$;*
- *there exists an edge between two vertices $\mathcal{J}$ and $\mathcal{K}$ if and only if $\mathcal{J} \cap \mathcal{K} \neq \emptyset$.*

The interference graph $G(K,k)$ has $\binom{K}{k}$ vertices. It is a regular graph [62, Sec. 1.2] of degree $\binom{K}{k} - \binom{K-k}{k} - 1$, with the convention $\binom{m}{n} = 0$ if $m < n$. Let $\chi(G(K,k))$ denote the chromatic number of $G(K,k)$, i.e., the minimum number of colors to color all the vertices such that adjacent vertices have different colors. We have the following property.

Property 1 (The chromatic number of the interference graph)**.** $\chi\big(G(K,1)\big) = 1$ *and when* $1 < k \leq \lfloor K/2 \rfloor$, $\chi\big(G(K,k)\big) \leq \binom{K}{k} - \binom{K-k}{k} - 1$, *and when* $k > \lfloor K/2 \rfloor$, $\chi\big(G(K,k)\big) = \binom{K}{k}$.

Proof. $\chi\big(G(K,1)\big) = 1$ since $G(K,1)$ is edgeless. $\chi\big(G(K,k)\big) = \binom{K}{k}$ when $k > \lfloor K/2 \rfloor$ because in this case, $G(K,k)$ is complete. The results for the case $1 < k \leq \lfloor K/2 \rfloor$ follows from Brook's theorem [62, Thm. 5.2.4]. □

In this section, we assume the users have perfect CSIR.

Theorem 8. *For the K-user broadcast channel with CSIR, for any integers $d_{\mathcal{J}}$ satisfy*

$$d_{\mathcal{J}} \leq r_{\mathcal{J}}, \quad \forall \mathcal{J} \subset [K], \tag{2.98}$$

$$\sum_{\mathcal{J}\subset[K]:\ k\in\mathcal{J}} d_{\mathcal{J}} \leq \min\big(r_k, N_k\big), \quad \forall k \in [K], \tag{2.99}$$

the DoF tuple $(d_1, \dots, d_K)$ given by

$$d_k = \sum_{\mathcal{J}\subset[K]:\ k\in\mathcal{J}} \tau_{k,\mathcal{J}} d_{\mathcal{J}}, \quad k \in [K], \tag{2.100}$$

for some time-sharing coefficients $\tau_{k,\mathcal{J}} \geq 0$ *satisfying* $\tau_{k,\mathcal{J}} = 0, \forall k \in \{[K]\backslash\mathcal{J}\}$ *and* $\sum_{i=1}^{K} \tau_{k,\mathcal{J}} = 1, \forall \mathcal{J} \subset [K]$, *is achievable.*

Proof. For $\mathcal{J} \subset [K]$, let $\mathbf{V}_{\mathcal{J}} \in \mathbb{C}^{M \times d_{\mathcal{J}}}$ be a matrix with orthonormal columns such that $\operatorname{Span}(\mathbf{V}_{\mathcal{J}}) \subset \mathcal{V}_{\mathcal{J}}$. Then $\mathbf{U}_k^{\mathsf{H}} \mathbf{V}_{\mathcal{J}} = \mathbf{0}$, $\forall k \notin \mathcal{J}$, and $\operatorname{rank}(\mathbf{U}_k^{\mathsf{H}} \mathbf{V}_{\mathcal{J}}) = d_{\mathcal{J}}$, $\forall k \in \mathcal{J}$. Let the transmitter send the signal

$$\mathbf{X} = \sum_{\mathcal{J} \subset [K]} \mathbf{V}_{\mathcal{J}} \mathbf{s}_{\mathcal{J}}, \tag{2.101}$$

where $\mathbf{s}_{\mathcal{J}} \in \mathbb{C}^{d_{\mathcal{J}}}$ contains data symbols. Let us consider User k and label the subsets in $\{\mathcal{J} \subset [K] : k \in \mathcal{J}\}$ as $\{\mathcal{J}_1, \ldots, \mathcal{J}_l\}$. The received signal at User k is

$$\begin{aligned} \mathbf{Y}_k &= \mathbf{H}_k \sum_{\mathcal{J} \subset [K]} \mathbf{V}_{\mathcal{J}} \mathbf{S}_{\mathcal{J}} + \mathbf{W}_k \\ &= \mathbf{G}_k \mathbf{\Sigma}_k^{\frac{1}{2}} \mathbf{U}_k^{\mathsf{H}} \sum_{\mathcal{J} \subset [K]:\ 1 \in \mathcal{J}} \mathbf{V}_{\mathcal{J}} \mathbf{s}_{\mathcal{J}} + \mathbf{W}_k \\ &= \mathbf{G}_k \mathbf{\Sigma}_k^{\frac{1}{2}} \mathbf{U}_k^{\mathsf{H}} [\mathbf{V}_{\mathcal{J}_1} \ \ldots \ \mathbf{V}_{\mathcal{J}_l}] \begin{bmatrix} \mathbf{s}_{\mathcal{J}_1} \\ \vdots \\ \mathbf{s}_{\mathcal{J}_l} \end{bmatrix} + \mathbf{W}_k. \end{aligned} \tag{2.102}$$

Because $\sum_{i=1}^{l} d_{\mathcal{J}_i} \leq \min(r_k, N_k)$, User k can decode $\mathbf{s}_{\mathcal{J}_1}, \ldots, \mathbf{s}_{\mathcal{J}_l}$, that is, $\{\mathbf{s}_{\mathcal{J}} \subset [K] :\ k \in \mathcal{J}\}$, where the signal $\mathbf{s}_{\mathcal{J}}$ provides $d_{\mathcal{J}}$ DoF. Signal $\mathbf{s}_{\mathcal{J}}$ can be decoded by all the users in $\mathcal{J}$. By dedicating $\mathbf{s}_{\mathcal{J}}$ to user $k \in \mathcal{J}$ in a fraction $\tau_{k,\mathcal{J}}$ of time, User k can achieve $\sum_{\mathcal{J} \in [K]: k \in \mathcal{J}} \tau_{k,\mathcal{J}} d_{\mathcal{J}}$ DoF. This completes the proof. $\square$

2.8 Appendix

2.8.1 Proof of Theorem 4

We prove by constructing pilot-based schemes that can achieve (2.59), (2.60), and (2.61).

Transmitter Ignores Correlation

The transmitter can ignore $\mathbf{R}$ and form the transmitted signal as if the channel is uncorrelated, but the performance still depends on correlation. Within each coherence block, the transmitter first sends an orthogonal pilot matrix $\mathbf{X}_\tau \in \mathbb{C}^{M\times M}$ such that $\mathbf{X}_\tau \mathbf{X}_\tau^{\mathsf{H}} = M\mathbf{I}_M$ during the first M channel uses (this is optimal for uncorrelated fading [55, Sec. III-A]), and then sends i.i.d. $\mathcal{CN}(0,1)$ data matrix $\mathbf{X}_\delta \in \mathbb{C}^{M\times(T-M)}$ during the remaining $T-M$ channel uses. That is,

$$\mathbf{X} = \left[\sqrt{\frac{\rho_\tau}{M}}\mathbf{X}_\tau \ \sqrt{\frac{\rho_\delta}{M}}\mathbf{X}_\delta\right], \tag{2.103}$$

where ρ_τ and ρ_δ are the average power used for training and data phases, respectively, and satisfy the power constraint $\rho_\tau M + \rho_\delta(T-M) \le \rho T$.

In the training phase, the receiver observes $\mathbf{Y}_\tau \triangleq \mathbf{Y}_{[1:M]} = \sqrt{\frac{\rho_\tau}{M}}\mathbf{H}\mathbf{X}_\tau + \mathbf{W}_{[1:M]}$. Following Lemma 3, it performs a linear MMSE channel estimator as

$$\hat{\mathbf{H}} = \sqrt{\frac{\rho_\tau}{M}}\mathbf{Y}_\tau\left(\frac{\rho_\tau}{M}\mathbf{X}_\tau^{\mathsf{H}}\mathbf{R}\mathbf{X}_\tau + \mathbf{I}_M\right)^{-1}\mathbf{X}_\tau^{\mathsf{H}}\mathbf{R}. \tag{2.104}$$

The estimate $\hat{\mathbf{H}}$ and the estimation error $\tilde{\mathbf{H}} = \mathbf{H} - \hat{\mathbf{H}}$ have zero mean and row covariance

$$\frac{1}{N}\mathbb{E}[\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}] = \frac{\rho_\tau}{M}\mathbf{R}\mathbf{X}_\tau\left(\frac{\rho_\tau}{M}\mathbf{X}_\tau^{\mathsf{H}}\mathbf{R}\mathbf{X}_\tau + \mathbf{I}_M\right)^{-1}\mathbf{X}_\tau^{\mathsf{H}}\mathbf{R} = \rho_\tau\mathbf{R}(\mathbf{I}_M + \rho_\tau\mathbf{R})^{-1}\mathbf{R}, \tag{2.105}$$

$$\frac{1}{N}\mathbb{E}[\tilde{\mathbf{H}}^{\mathsf{H}}\tilde{\mathbf{H}}] = \mathbf{R} - \rho_\tau\mathbf{R}(\mathbf{I}_M + \rho_\tau\mathbf{R})^{-1}\mathbf{R}. \tag{2.106}$$

In the data transmission phase, the received signal is

$$\mathbf{Y}_\delta \triangleq \mathbf{Y}_{[M+1:T]} = \sqrt{\frac{\rho_\delta}{M}}\mathbf{H}\mathbf{X}_\delta + \mathbf{W}_{[M+1:T]} = \sqrt{\frac{\rho_\delta}{M}}\hat{\mathbf{H}}\mathbf{X}_\delta + \mathbf{W}_\delta, \tag{2.107}$$

where $\mathbf{W}_\delta \triangleq \sqrt{\frac{\rho_\delta}{M}}\tilde{\mathbf{H}}\mathbf{X}_\delta + \mathbf{W}_{[M+1:T]}$ is the combined noise consisting of additive noise and channel estimation error. With MMSE estimator, $\mathbf{W}_\delta$ and $\mathbf{X}_\delta$ are uncorrelated because

$$\mathbb{E}[\mathbf{X}_\delta\mathbf{W}_\delta^{\mathsf{H}}|\mathbf{X}_\tau, \mathbf{Y}_\tau] = \mathbb{E}\left[\mathbf{X}_\delta\left(\sqrt{\frac{\rho_\delta}{M}}\mathbf{X}_\delta^{\mathsf{H}}\tilde{\mathbf{H}}^{\mathsf{H}} + \mathbf{W}_\delta^{\mathsf{H}}\right)\Big|\mathbf{X}_\tau, \mathbf{Y}_\tau\right] \tag{2.108}$$

$$= \sqrt{\frac{\rho_\delta}{M}}\mathbb{E}\left[\mathbf{X}_\delta\mathbf{X}_\delta^{\mathsf{H}}(\mathbf{H} - \hat{\mathbf{H}})\Big|\mathbf{X}_\tau, \mathbf{Y}_\tau\right] \tag{2.109}$$

$$= \mathbf{0}, \tag{2.110}$$

since $\mathbb{E}\big[\mathbf{H}-\hat{\mathbf{H}}\big|\mathbf{X}_\tau,\mathbf{Y}_\tau\big]=0$. From Lemma 2, a lower bound on the achievable rate is obtained by replacing $\mathbf{W}_\delta$ by i.i.d. Gaussian noise with the same variance

$$\sigma^2_{\mathbf{W}_\delta}=\frac{1}{N(T-M)}\mathrm{tr}\big(\mathbb{E}\big[\mathbf{W}_\delta^{\mathsf{H}}\mathbf{W}_\delta\big]\big)=\frac{\rho_\delta}{M}\mathrm{tr}\left(\mathbf{R}-\rho_\tau\mathbf{R}(\mathbf{I}_M+\rho_\tau\mathbf{R})^{-1}\mathbf{R}\right)+1 \tag{2.111}$$

$$=\frac{\rho_\delta}{M}\mathrm{tr}\left((\mathbf{\Sigma}^{-1}+\rho_\tau\mathbf{I}_r)^{-1}\right)+1. \tag{2.112}$$

Thus, the achievable rate is lower bounded by

$$R=\frac{T-M}{T}\mathbb{E}\left[\log\det\left(\mathbf{I}_N+\frac{\rho_\delta}{M\sigma^2_{\mathbf{W}_\delta}}\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathsf{H}}\right)\right]. \tag{2.113}$$

From (2.105), $\hat{\mathbf{H}}$ has correlation matrix $\rho_\tau\mathbf{R}(\mathbf{I}_M+\rho_\tau\mathbf{R})^{-1}\mathbf{R}$. This shows (2.59).

Transmitter Exploits Correlation

By exploiting $\mathbf{R}$, the transmitter can project the signal onto the eigenspace of $\mathbf{R}$ and can also adapt the pilot symbols. The transmitter builds a precoder $\mathbf{V}\in\mathbb{C}^{M\times r}$ with r orthonormal columns such that $\mathrm{Span}\,(\mathbf{V})=\mathrm{Span}\,(\mathbf{U})$. Let $\mathbf{\Phi}=\mathbf{U}^{\mathsf{H}}\mathbf{V}$. The transmitted signal is

$$\mathbf{X}=\mathbf{V}\left[\sqrt{\frac{\rho_\tau}{r}}\mathbf{X}_\tau\;\sqrt{\frac{\rho_\delta}{r}}\mathbf{X}_\delta\right] \tag{2.114}$$

where $\mathbf{X}_\tau\in\mathbb{C}^{r\times r}$ such that $\mathrm{rank}\,(\mathbf{X}_\tau)=r$ and $\mathrm{tr}\,(\mathbf{X}_\tau^{\mathsf{H}}\mathbf{X}_\tau)=r^2$ is the pilot matrix, and $\mathbf{X}_\delta\in\mathbb{C}^{r\times(T-r)}$ is the data matrix containing $\mathcal{CN}(0,1)$ entries. The average pilot and data powers satisfy $\rho_\tau r+\rho_\delta(T-r)\le\rho T$.

The received signal during the training phase is then $\mathbf{Y}_\tau\triangleq\mathbf{Y}_{[1:r]}=\sqrt{\frac{\rho_\tau}{r}}\mathbf{G}\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{\Phi}\mathbf{X}_\tau+\mathbf{W}_{[1:r]}$. The equivalent channel $\mathbf{\Omega}\triangleq\mathbf{G}\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{\Phi}$ has correlation matrix $\bar{\mathbf{R}}=\mathbf{\Phi}^{\mathsf{H}}\mathbf{\Sigma}\mathbf{\Phi}=\mathbf{V}^{\mathsf{H}}\mathbf{R}\mathbf{V}$. According to Lemma 3, the MMSE channel estimate for the equivalent channel $\mathbf{\Omega}$ is given by

$$\hat{\mathbf{\Omega}}=\sqrt{\frac{\rho_\tau}{r}}\mathbf{Y}_\tau\left(\frac{\rho_\tau}{r}\mathbf{X}_\tau^{\mathsf{H}}\bar{\mathbf{R}}\mathbf{X}_\tau+\mathbf{I}_r\right)^{-1}\mathbf{X}_\tau^{\mathsf{H}}\bar{\mathbf{R}}. \tag{2.115}$$

The estimate $\hat{\mathbf{\Omega}}$ and the estimation error $\tilde{\mathbf{\Omega}} = \mathbf{G}\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{\Phi} - \hat{\mathbf{\Omega}}$ have zero mean and row covariance

$$\frac{1}{N}\mathbb{E}[\hat{\mathbf{\Omega}}^{\mathsf{H}}\hat{\mathbf{\Omega}}] = \frac{\rho_\tau}{r}\bar{\mathbf{R}}\mathbf{X}_\tau\Big(\frac{\rho_\tau}{r}\mathbf{X}_\tau^{\mathsf{H}}\bar{\mathbf{R}}\mathbf{X}_\tau + \mathbf{I}_r\Big)^{-1}\mathbf{X}_\tau^{\mathsf{H}}\bar{\mathbf{R}}, \tag{2.116}$$

$$\frac{1}{N}\mathbb{E}[\tilde{\mathbf{\Omega}}^{\mathsf{H}}\tilde{\mathbf{\Omega}}] = \bar{\mathbf{R}} - \frac{\rho_\tau}{r}\bar{\mathbf{R}}\mathbf{X}_\tau\Big(\frac{\rho_\tau}{r}\mathbf{X}_\tau^{\mathsf{H}}\bar{\mathbf{R}}\mathbf{X}_\tau + \mathbf{I}_r\Big)^{-1}\mathbf{X}_\tau^{\mathsf{H}}\bar{\mathbf{R}} = \big(\bar{\mathbf{R}}^{-1} + \frac{\rho_\tau}{r}\mathbf{X}_\tau\mathbf{X}_\tau^{\mathsf{H}}\big)^{-1}. \tag{2.117}$$

In the data transmission phase, the received signal is

$$\mathbf{Y}_\delta \triangleq \mathbf{Y}_{[r+1:T]} = \sqrt{\frac{\rho_\delta}{r}}\mathbf{G}\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{\Phi}\mathbf{X}_\delta + \mathbf{W}_{[r+1:T]} = \sqrt{\frac{\rho_\delta}{r}}\hat{\mathbf{\Omega}}\mathbf{X}_\delta + \mathbf{W}_\delta, \tag{2.118}$$

where $\mathbf{W}_\delta \triangleq \sqrt{\frac{\rho_\delta}{r}}\tilde{\mathbf{\Omega}}\mathbf{X}_\delta + \mathbf{W}_{[r+1:T]}$. From Lemma 2, a lower bound on the achievable rate is obtained by replacing $\mathbf{W}_\delta$ with i.i.d. Gaussian noise with the same variance

$$\sigma^2_{\mathbf{W}_\delta} = \frac{1}{N(T-r)}\mathrm{tr}\big(\mathbb{E}[\mathbf{W}_\delta^{\mathsf{H}}\mathbf{W}_\delta]\big) = \frac{\rho_\delta}{r}\mathrm{tr}\Big(\big(\bar{\mathbf{R}}^{-1} + \frac{\rho_\tau}{r}\mathbf{X}_\tau\mathbf{X}_\tau^{\mathsf{H}}\big)^{-1}\Big) + 1. \tag{2.119}$$

The corresponding achievable rate lower bound is

$$R = \frac{T-r}{T}\mathbb{E}\Big[\log\det\Big(\mathbf{I}_N + \frac{\rho_\delta}{r\sigma^2_{\mathbf{W}_\delta}}\hat{\mathbf{\Omega}}\hat{\mathbf{\Omega}}^{\mathsf{H}}\Big)\Big] \tag{2.120}$$

where the rows of $\hat{\mathbf{\Omega}}$ obey $\mathcal{CN}\big(0^t, \bar{\mathbf{R}} - \mathbf{B}\big)$ with $\mathbf{B} \triangleq \big(\bar{\mathbf{R}}^{-1} + \frac{\rho_\tau}{r}\mathbf{X}_\tau\mathbf{X}_\tau^{\mathsf{H}}\big)^{-1}$ and are independent with each other.

Taking $\mathbf{X}_\tau$ such that $\mathbf{X}_\tau\mathbf{X}_\tau^{\mathsf{H}} = r\mathbf{I}_r$ (i.e., orthogonal pilots), we have $\mathbf{B} = \big(\bar{\mathbf{R}}^{-1} + \rho_\tau\mathbf{I}_r\big)^{-1}$, and the achievable rate R is given in (2.60).

We can also optimize the pilot $\mathbf{X}_\tau$ so as to maximize R. The pilot matrix $\mathbf{X}_\tau$ affects the achievable rate bound primarily through the effective SNR

$$\rho_{\text{eff}} = \frac{\rho_\delta}{r\sigma^2_{\mathbf{W}_\delta}}\frac{1}{N}\mathbb{E}\big(\mathrm{tr}\big[\hat{\mathbf{\Omega}}^{\mathsf{H}}\hat{\mathbf{\Omega}}\big]\big) = \frac{\rho_\delta\mathrm{tr}(\bar{\mathbf{R}} - \mathbf{B})}{\rho_\delta\mathrm{tr}(\mathbf{B}) + r} \tag{2.121}$$

which decreases with $\mathrm{tr}\,(\mathbf{B})$. Therefore, to maximize R, we would like to minimize $\mathrm{tr}\,(\mathbf{B})$. That is

$$\min_{\mathrm{tr}(\mathbf{X}_\tau^{\mathsf{H}}\mathbf{X}_\tau) = r^2} \mathrm{tr}\Big(\big(\bar{\mathbf{R}}^{-1} + \frac{\rho_\tau}{r}\mathbf{X}_\tau\mathbf{X}_\tau^{\mathsf{H}}\big)^{-1}\Big). \tag{2.122}$$

Using Lagrange multiplier λ, we minimize

$$L(\mathbf{X}_\tau, \lambda) = \operatorname{tr}\Big(\big(\bar{\mathbf{R}}^{-1} + \frac{\rho_\tau}{r}\mathbf{X}_\tau\mathbf{X}_\tau^{\mathsf{H}}\big)^{-1}\Big) + \lambda\big(\operatorname{tr}(\mathbf{X}_\tau\mathbf{X}_\tau^{\mathsf{H}}) - r^2\big). \tag{2.123}$$

Solving $\frac{\partial L(\mathbf{X}_\tau,\lambda)}{\partial \mathbf{X}_\tau\mathbf{X}_\tau^{\mathsf{H}}} = 0$, we obtain the minimizer $\mathbf{X}_\tau\mathbf{X}_\tau^{\mathsf{H}} = \sqrt{\frac{r}{\rho_\tau\lambda}}\mathbf{I}_r - \frac{r}{\rho_\tau}\bar{\mathbf{R}}^{-1}$. Using the constrain $\operatorname{tr}(\mathbf{X}_\tau^{\mathsf{H}}\mathbf{X}_\tau) = r^2$, we find that $\frac{\rho_\tau}{r}\mathbf{X}_\tau\mathbf{X}_\tau^{\mathsf{H}} = \big(\rho_\tau + \frac{1}{r}\operatorname{tr}(\bar{\mathbf{R}}^{-1})\big)\mathbf{I}_r - \bar{\mathbf{R}}^{-1}$. With this, $\mathbf{B} = \big(\rho_\tau + \frac{1}{r}\operatorname{tr}(\bar{\mathbf{R}}^{-1})\big)^{-1}\mathbf{I}_r$, and the rate R is given in (2.61). The effective SNR is now written as

$$\rho_{\text{eff}} = \frac{\rho_\delta}{\rho_\delta r\big(\rho_\tau + \frac{1}{r}\operatorname{tr}\left(\bar{\mathbf{R}}^{-1}\right)\big)^{-1} + r}\Big[\operatorname{tr}\left(\bar{\mathbf{R}}\right) - r\Big(\rho_\tau + \frac{1}{r}\operatorname{tr}\left(\bar{\mathbf{R}}^{-1}\right)\Big)^{-1}\Big]. \tag{2.124}$$

Let $\rho_\tau r = (1-\alpha)\rho T$ and $\rho_\delta(T-r) = \alpha\rho T$ for $\alpha \in (0,1)$, we can derive that

$$\rho_{\text{eff}} = \frac{\rho T \operatorname{tr}\left(\bar{\mathbf{R}}\right)}{r(T-2r)}\frac{-\alpha^2 + a\alpha}{-\alpha + b} \tag{2.125}$$

where $a \triangleq 1 + \frac{\operatorname{tr}(\bar{\mathbf{R}}^{-1})}{\rho T} - \frac{r^2}{\rho T \operatorname{tr}(\bar{\mathbf{R}})}$ and $b \triangleq \frac{T-r}{T-2r}\Big(1 + \frac{\operatorname{tr}(\bar{\mathbf{R}}^{-1})}{\rho T}\Big)$. Noting that $T - 2r \geq 0$, we obtain the optimal value of α that maximizes ρ_{eff} as given in (2.62). This completes the proof.

2.8.2 Proof of Theorem 5

This achievable rate region is fully characterized by the mutual information $I(\mathbf{Y}_k; \mathbf{X}_k, \mathbf{X}_0)$, $I(\mathbf{Y}_k; \mathbf{X}_k \,|\, \mathbf{X}_0)$, and $I(\mathbf{Y}_k; \mathbf{X}_0 \,|\, \mathbf{X}_k)$, $k \in \{1, 2\}$. We cacluate the achievable rates for the following input distribution:

$$\mathbf{X}_0 = \left[\sqrt{\rho_{0\tau}}\mathbf{I}_{s_0} \;\; \mathbf{0}_{s_0\times s_1} \;\; \sqrt{\frac{\rho_{0\delta}}{s_0}}\mathbf{S}_0\right], \tag{2.126}$$

$$\mathbf{X}_1 = \left[\mathbf{0}_{s_1\times s_0} \;\; \sqrt{\rho_{1\tau}}\mathbf{I}_{s_1} \;\; \sqrt{\frac{\rho_{1\delta}}{s_1}}\mathbf{S}_1\right], \tag{2.127}$$

$$\mathbf{X}_2 = \left[\mathbf{0}_{s_2\times s_0} \;\; \sqrt{\rho_{2\tau}}\mathbf{I}_{s_2} \;\; \sqrt{\frac{\rho_{2\delta}}{s_2}}\mathbf{S}_2\right], \tag{2.128}$$

where $\mathbf{S}_0 \in \mathbb{C}^{s_0\times(T-s_1-s_0)}$, $\mathbf{S}_1 \in \mathbb{C}^{s_1\times(T-s_1-s_0)}$, and $\mathbf{S}_2 \in \mathbb{C}^{s_2\times(T-s_2-s_0)}$ are data matrices containing independent $\mathcal{CN}(0,1)$ symbols, for powers $\rho_{i\tau}, \rho_\delta$, $i \in \{0,1,2\}$, such that

$$\rho_{0\tau}s_0 + \rho_{0\delta}(T - s_1 - s_0) + \sum_{i=1}^{2}\left[\rho_{i\tau}s_i + \rho_{i\delta}(T - s_i - s_0)\right] = \rho T. \tag{2.129}$$

The received signal at User 1 is

$$\mathbf{Y}_1 = \mathbf{G}_1\mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{\Phi}_1 \begin{bmatrix} \sqrt{\rho_{0\tau}}\mathbf{I}_{s_0} & \mathbf{0} & \sqrt{\frac{\rho_{0\delta}}{s_0}}\mathbf{S}_0 \\ \mathbf{0} & \sqrt{\rho_{1\tau}}\mathbf{I}_{s_1} & \sqrt{\frac{\rho_{1\delta}}{s_1}}\mathbf{S}_1 \end{bmatrix} + \mathbf{W}_1 \tag{2.130}$$

$$= \Bigg[\underbrace{\mathbf{G}_1\mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{\Phi}_1\mathbf{P}_{1\tau}^{\frac{1}{2}} + \mathbf{W}_{1[1:s_1+s_0]}}_{\mathbf{Y}_{1\tau}} \quad \underbrace{\mathbf{G}_1\mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{\Phi}_1\mathbf{P}_{1\delta}^{\frac{1}{2}}\begin{bmatrix}\mathbf{S}_0\\ \mathbf{S}_1\end{bmatrix} + \mathbf{W}_{1[s_1+s_0+1;T]}}_{\mathbf{Y}_{1\delta}} \Bigg] \tag{2.131}$$

where $\mathbf{P}_{1\tau} \triangleq \begin{bmatrix} \rho_{0\tau}\mathbf{I}_{s_0} & \mathbf{0} \\ \mathbf{0} & \rho_{1\tau}\mathbf{I}_{s_1} \end{bmatrix}$ and $\mathbf{P}_{1\delta} \triangleq \begin{bmatrix} \frac{\rho_{0\delta}}{s_0}\mathbf{I}_{s_0} & \mathbf{0} \\ \mathbf{0} & \frac{\rho_{1\delta}}{s_1}\mathbf{I}_{s_1} \end{bmatrix}$ are the power matrices for the pilot and data, respectively.

The equivalent channel $\mathbf{\Omega}_1 \triangleq \mathbf{G}_1\mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{\Phi}_1$ has correlation matrix $\bar{\mathbf{R}}_1$. Following Lemma 3, User 1 performs a MMSE channel estimation based on $\mathbf{Y}_{1\tau}$ as

$$\hat{\mathbf{\Omega}}_1 = \mathbf{Y}_{1\tau}\big(\mathbf{P}_{1\tau}^{\frac{1}{2}}\bar{\mathbf{R}}_1\mathbf{P}_{1\tau}^{\frac{1}{2}} + \mathbf{I}_{s_1+s_0}\big)^{-1}\mathbf{P}_{1\tau}^{\frac{1}{2}}\bar{\mathbf{R}}_1. \tag{2.132}$$

The estimate $\hat{\mathbf{\Omega}}_1$ and the estimation error $\tilde{\mathbf{\Omega}}_1 = \mathbf{G}_1\mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{\Phi}_1 - \hat{\mathbf{\Omega}}_1$ have zero mean and row covariance

$$\frac{1}{N_1}\mathbb{E}[\hat{\mathbf{\Omega}}_1^{\mathsf{H}}\hat{\mathbf{\Omega}}_1] = \bar{\mathbf{R}}_1\mathbf{P}_{1\tau}^{\frac{1}{2}}\Big(\mathbf{P}_{1\tau}^{\frac{1}{2}}\bar{\mathbf{R}}_1\mathbf{P}_{1\tau}^{\frac{1}{2}} + \mathbf{I}_{s_1+s_0}\Big)^{-1}\mathbf{P}_{1\tau}^{\frac{1}{2}}\bar{\mathbf{R}}_1, \tag{2.133}$$

$$\frac{1}{N_1}\mathbb{E}[\tilde{\mathbf{\Omega}}_1^{\mathsf{H}}\tilde{\mathbf{\Omega}}_1] = \bar{\mathbf{R}}_1 - \bar{\mathbf{R}}_1\mathbf{P}_{1\tau}^{\frac{1}{2}}\Big(\mathbf{P}_{1\tau}^{\frac{1}{2}}\bar{\mathbf{R}}_1\mathbf{P}_{1\tau}^{\frac{1}{2}} + \mathbf{I}_{s_1+s_0}\Big)^{-1}\mathbf{P}_{1\tau}^{\frac{1}{2}}\bar{\mathbf{R}}_1 = \big(\bar{\mathbf{R}}_1^{-1} + \mathbf{P}_{1\tau}\big)^{-1}. \tag{2.134}$$

<u>*Lower bounding $I(\mathbf{Y}_1; \mathbf{X}_1, \mathbf{X}_0)$:*</u> The received signal during the data transmission phase can be written as

$$\mathbf{Y}_{1\delta} = \hat{\mathbf{G}}_1\mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{\Phi}_1\mathbf{P}_{1\delta}^{\frac{1}{2}}\begin{bmatrix}\mathbf{S}_0\\ \mathbf{S}_1\end{bmatrix} + \mathbf{W}_{1\delta}, \tag{2.135}$$

where $\mathbf{W}_{1\delta} \triangleq \tilde{\mathbf{\Omega}}_1\mathbf{P}_{1\delta}^{\frac{1}{2}}\begin{bmatrix}\mathbf{S}_0\\ \mathbf{S}_1\end{bmatrix} + \mathbf{W}_{1[s_1+s_0+1:T]}$ is the combined noise and residual interference due to channel estimation error. Define $\bar{\mathbf{\Omega}}_1 \in \mathbb{C}^{N_1\times(s_1+s_0)}$ with independent rows obeying

$\mathcal{CN}\big(\mathbf{0}^t, \mathbf{P}_{1\tau}^{\frac{1}{2}}(\mathbf{P}_{1\tau}^{\frac{1}{2}}\bar{\mathbf{R}}_1\mathbf{P}_{1\tau}^{\frac{1}{2}} + \mathbf{I}_{s_1+s_0})^{-1}\mathbf{P}_{1\tau}^{\frac{1}{2}}\big)$. By a similar analysis using Lemma 2 as for (2.60) in Theorem 4, we have

$$
\begin{aligned}
&I(\mathbf{Y}_1; \mathbf{X}_1, \mathbf{X}_0) \\
&= I(\mathbf{Y}_{1\delta}; \mathbf{S}_1, \mathbf{S}_0 \,|\, \mathbf{Y}_{1\tau}) + \underbrace{I(\mathbf{Y}_{1\tau}; \mathbf{S}_1, \mathbf{S}_0)}_{=0} && (2.136)\\
&= I(\mathbf{Y}_{1\delta}; \mathbf{S}_1, \mathbf{S}_0 \,|\, \hat{\mathbf{\Omega}}_1) && (2.137)\\
&\geq \Big(T - s_1 - s_0\Big)\mathbb{E}\Bigg[\log\det\Bigg(\mathbf{I}_{N_1} + \frac{1}{\mathrm{tr}\Big(\Big(\bar{\mathbf{R}}_1^{-1} + \mathbf{P}_{1\tau}\Big)^{-1}\mathbf{P}_{1\delta}\Big) + 1}\hat{\mathbf{\Omega}}_1\mathbf{P}_{1\delta}\hat{\mathbf{\Omega}}_1^{\mathsf{H}}\Bigg)\Bigg] && (2.138)\\
&\geq \Big(T - s_1 - s_0\Big)\mathbb{E}\Bigg[\log\det\Bigg(\mathbf{I}_{N_1} + \frac{1}{\mathrm{tr}\Big(\Big(\bar{\mathbf{R}}_1^{-1} + \mathbf{P}_{1\tau}\Big)^{-1}\mathbf{P}_{1\delta}\Big) + 1}\bar{\mathbf{\Omega}}_1\bar{\mathbf{R}}_1\mathbf{P}_{1\delta}\bar{\mathbf{R}}_1^{\mathsf{H}}\bar{\mathbf{\Omega}}_1^{\mathsf{H}}\Bigg)\Bigg]. && (2.139)
\end{aligned}
$$

<u>*Lower bounding $I(\mathbf{Y}_1; \mathbf{X}_1 \,|\, \mathbf{X}_0)$:*</u> We rewrite $\mathbf{Y}_{1\delta}$ as

$$
\mathbf{Y}_{1\delta} = \sqrt{\frac{\rho_{1\delta}}{s_1}}\mathbf{G}_1\mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{\Phi}_{11}\mathbf{S}_1 + \sqrt{\frac{\rho_{0\delta}}{s_0}}\mathbf{G}_1\mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{\Phi}_{10}\mathbf{S}_0 + \mathbf{W}_{1\delta}. \tag{2.140}
$$

While decoding $\mathbf{S}_1$, the term $\sqrt{\frac{\rho_{0\delta}}{s_0}}\mathbf{G}_1\mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{\Phi}_{10}\mathbf{S}_0$ is an interference. Given the knowledge of $\mathbf{S}_0$ and the channel estimate $\hat{\mathbf{\Omega}}_1 = \big[\hat{\mathbf{\Omega}}_{10}\ \hat{\mathbf{\Omega}}_{11}\big]$, where $\hat{\mathbf{\Omega}}_{10}$ and $\hat{\mathbf{\Omega}}_{11}$ are respectively the estimates of $\mathbf{G}_1\mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{\Phi}_{10}$ and $\mathbf{G}_1\mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{\Phi}_{11}$, the receiver can remove partly the interference to obtain

$$
\mathbf{Y}_{1\delta} - \sqrt{\frac{\rho_{0\delta}}{s_0}}\hat{\mathbf{\Omega}}_{10}\mathbf{S}_0 = \sqrt{\frac{\rho_{1\delta}}{s_1}}\mathbf{G}_1\mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{\Phi}_{11}\mathbf{S}_1 + \sqrt{\frac{\rho_{0\delta}}{s_0}}\Big[\mathbf{G}_1\mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{\Phi}_{10} - \hat{\mathbf{\Omega}}_{10}\Big]\mathbf{S}_0 + \mathbf{W}_{1[s_1+s_0+1:T]} \tag{2.141}
$$

$$
= \sqrt{\frac{\rho_{1\delta}}{s_1}}\hat{\mathbf{\Omega}}_{11}\mathbf{S}_1 + \mathbf{W}_{1\delta}. \tag{2.142}
$$

With a similar analysis, using Lemma 2 as for (2.60) in Theorem 4,

$$
\begin{aligned}
&I(\mathbf{Y}_1; \mathbf{X}_1 \,|\, \mathbf{X}_0) \\
&= I\big(\mathbf{Y}_{1\delta}; \mathbf{S}_1 \big| \mathbf{S}_0, \mathbf{Y}_{1\tau}\big) && (2.143) \\
&= I\big(\mathbf{Y}_{1\delta}; \mathbf{S}_1 \big| \mathbf{S}_0, \hat{\mathbf{\Omega}}_1\big) && (2.144) \\
&= I\big(\mathbf{Y}_{1\delta} - \sqrt{\frac{\rho_{0\delta}}{s_0}} \hat{\mathbf{\Omega}}_{10}\mathbf{S}_0; \mathbf{S}_1 \;\big|\; \mathbf{S}_0, \hat{\mathbf{\Omega}}_1\big) && (2.145) \\
&= I\big(\sqrt{\frac{\rho_{1\delta}}{s_1}} \hat{\mathbf{\Omega}}_{11}\mathbf{S}_1 + \mathbf{W}_{1\delta}; \mathbf{S}_1 \;\big|\; \hat{\mathbf{\Omega}}_{11}\big) && (2.146) \\
&\geq \Big(T - s_1 - s_0\Big) \mathbb{E}\Bigg[\log\det\Bigg(\mathbf{I}_{N_1} + \frac{\rho_{1\delta}}{s_1\big[\mathrm{tr}\big((\bar{\mathbf{R}}_1^{-1} + \mathbf{P}_{1\tau})^{-1}\mathbf{P}_{1\delta}\big) + 1\big]} \bar{\mathbf{\Omega}}_1 \bar{\mathbf{R}}_{11} \bar{\mathbf{R}}_{11}^{\mathsf{H}} \bar{\mathbf{\Omega}}_1^{\mathsf{H}}\Bigg)\Bigg] && (2.147)
\end{aligned}
$$

<u>*Lower bounding $I(\mathbf{Y}_1; \mathbf{X}_0 \,|\, \mathbf{X}_1)$:*</u> Given $\mathbf{S}_1$ and the channel estimate $\hat{\mathbf{\Omega}}_1 = \big[\hat{\mathbf{\Omega}}_{10} \quad \hat{\mathbf{\Omega}}_{11}\big]$, the receiver can remove partly the interference in (2.140) to obtain

$$
\mathbf{Y}_{1\delta} - \sqrt{\frac{\rho_{1\delta}}{s_1}} \hat{\mathbf{\Omega}}_{11}\mathbf{S}_1 = \sqrt{\frac{\rho_{0\delta}}{s_0}} \mathbf{G}_1 \mathbf{\Sigma}_1^{\frac{1}{2}} \mathbf{\Phi}_{10} \mathbf{S}_0 + \sqrt{\frac{\rho_{1\delta}}{s_1}} \Big[\mathbf{G}_1 \mathbf{\Sigma}_1^{\frac{1}{2}} \mathbf{\Phi}_{11} - \hat{\mathbf{\Omega}}_{11}\mathbf{S}_1\Big] + \mathbf{W}_{1[s_1+s_0+1:T]} \tag{2.148}
$$

$$
= \sqrt{\frac{\rho_{0\delta}}{s_0}} \hat{\mathbf{\Omega}}_{10}\mathbf{S}_0 + \mathbf{W}_{1\delta}. \tag{2.149}
$$

Using reasoning similar to (2.60) in Theorem 4,

$$
\begin{aligned}
&I(\mathbf{Y}_1; \mathbf{X}_0 \,|\, \mathbf{X}_1) \\
&= I\big(\mathbf{Y}_{1\delta}; \mathbf{S}_0 \big| \mathbf{S}_1, \mathbf{Y}_{1\tau}\big) && (2.150) \\
&= I\big(\mathbf{Y}_{1\delta}; \mathbf{S}_0 \big| \mathbf{S}_1, \hat{\mathbf{\Omega}}_1\big) && (2.151) \\
&= I\big(\mathbf{Y}_{1\delta} - \sqrt{\frac{\rho_{1\delta}}{s_1}} \hat{\mathbf{\Omega}}_{11}\mathbf{S}_1; \mathbf{S}_0 \;\big|\; \mathbf{S}_1, \hat{\mathbf{\Omega}}_1\big) && (2.152) \\
&= I\big(\sqrt{\frac{\rho_{0\delta}}{s_0}} \hat{\mathbf{\Omega}}_{10}\mathbf{S}_0 + \mathbf{W}_{1\delta}; \mathbf{S}_0 \;\big|\; \hat{\mathbf{\Omega}}_{10}\big) && (2.153) \\
&\geq \Big(T - s_1 - s_0\Big) \mathbb{E}\Bigg[\log\det\Bigg(\mathbf{I}_{N_1} + \frac{\rho_{0\delta}}{s_0\big[\mathrm{tr}\big((\bar{\mathbf{R}}_1^{-1} + \mathbf{P}_{1\tau}\big)^{-1}\mathbf{P}_{1\delta}\big) + 1\big]} \bar{\mathbf{\Omega}}_1 \bar{\mathbf{R}}_{10} \bar{\mathbf{R}}_{10}^{\mathsf{H}} \bar{\mathbf{\Omega}}_1^{\mathsf{H}}\Bigg)\Bigg] && (2.154)
\end{aligned}
$$

The received signal at User 2 is

$$\mathbf{Y}_2 = \mathbf{G}_2\mathbf{\Sigma}_2^{\frac{1}{2}}\mathbf{\Phi}_2 \begin{bmatrix} \sqrt{\rho_{0\tau}}\mathbf{I}_{s_0} & \mathbf{0} & \mathbf{0}_{s_0\times(s_1-s_2)} & \sqrt{\frac{\rho_{0\delta}}{s_0}}\mathbf{S}_0 \\ \mathbf{0} & \sqrt{\rho_{2\tau}}\mathbf{I}_{s_2} & \sqrt{\frac{\rho_{2\delta}}{s_2}}\mathbf{S}_{2a} & \sqrt{\frac{\rho_{2\delta}}{s_2}}\mathbf{S}_{2b} \end{bmatrix} + \mathbf{W}_2 \tag{2.155}$$

$$= \Bigg[\underbrace{\mathbf{G}_2\mathbf{\Sigma}_2^{\frac{1}{2}}\mathbf{\Phi}_2\mathbf{P}_{2\tau}^{\frac{1}{2}} + \mathbf{W}_{2[1:s_2+s_0]}}_{\mathbf{Y}_{2\tau}},$$

$$\underbrace{\sqrt{\frac{\rho_{2\delta}}{s_2}}\mathbf{G}_2\mathbf{\Sigma}_2^{\frac{1}{2}}\mathbf{\Phi}_{22}\mathbf{S}_{2a} + \mathbf{W}_{2[s_2+s_0+1:s_1+s_0]}}_{\mathbf{Y}_{2\delta a}},\ \underbrace{\mathbf{G}_2\mathbf{\Sigma}_2^{\frac{1}{2}}\mathbf{\Phi}_2\mathbf{P}_{2\delta}^{\frac{1}{2}}\begin{bmatrix}\mathbf{S}_0\\ \mathbf{S}_{2b}\end{bmatrix} + \mathbf{W}_{2[s_1+s_0+1:T]}}_{\mathbf{Y}_{2\delta b}}\Bigg], \tag{2.156}$$

where $\mathbf{S}_{2a}$ and $\mathbf{S}_{2b}$ are respectively the first $s_1 - s_2$ columns and the remaining $T - s_1 - s_0$ columns of $\mathbf{S}_2$; $\mathbf{P}_{2\tau} \triangleq \begin{bmatrix} \rho_{0\tau}\mathbf{I}_{s_0} & \mathbf{0} \\ \mathbf{0} & \rho_{2\tau}\mathbf{I}_{s_2} \end{bmatrix}$ and $\mathbf{P}_{2\delta} \triangleq \begin{bmatrix} \frac{\rho_{0\delta}}{s_0}\mathbf{I}_{s_0} & \mathbf{0} \\ \mathbf{0} & \frac{\rho_{2\delta}}{s_2}\mathbf{I}_{s_2} \end{bmatrix}$ are the power matrices for the pilot and data, respectively. Following Lemma 3, user 2 performs a MMSE channel estimation of $\mathbf{\Omega}_2 \triangleq \mathbf{G}_2\mathbf{\Sigma}_2^{\frac{1}{2}}\mathbf{\Phi}_2 = [\mathbf{\Omega}_{20}\ \mathbf{\Omega}_{22}] = \left[\mathbf{G}_2\mathbf{\Sigma}_2^{\frac{1}{2}}\mathbf{\Phi}_{20}\ \mathbf{G}_2\mathbf{\Sigma}_2^{\frac{1}{2}}\mathbf{\Phi}_{22}\right]$ based on $\mathbf{Y}_{2\tau}$ as

$$\hat{\mathbf{\Omega}}_2 = \mathbf{Y}_{2\tau}\big(\mathbf{P}_{2\tau}^{\frac{1}{2}}\bar{\mathbf{R}}_2\mathbf{P}_{2\tau}^{\frac{1}{2}} + \mathbf{I}_{s_2+s_0}\big)^{-1}\mathbf{P}_{2\tau}^{\frac{1}{2}}\bar{\mathbf{R}}_2. \tag{2.157}$$

The estimate $\hat{\mathbf{\Omega}}_2 = \left[\hat{\mathbf{\Omega}}_{20}\ \hat{\mathbf{\Omega}}_{22}\right]$ and the estimation error $\tilde{\mathbf{\Omega}}_2 = \mathbf{G}_2\mathbf{\Sigma}_2^{\frac{1}{2}}\mathbf{\Phi}_2 - \hat{\mathbf{\Omega}}_2$ have zero mean and row covariance

$$\frac{1}{N_2}\mathbb{E}[\hat{\mathbf{\Omega}}_2^{\mathsf{H}}\hat{\mathbf{\Omega}}_2] = \bar{\mathbf{R}}_2\mathbf{P}_{2\tau}^{\frac{1}{2}}\Big(\mathbf{P}_{2\tau}^{\frac{1}{2}}\bar{\mathbf{R}}_2\mathbf{P}_{2\tau}^{\frac{1}{2}} + \mathbf{I}_{s_2+s_0}\Big)^{-1}\mathbf{P}_{2\tau}^{\frac{1}{2}}\bar{\mathbf{R}}_2, \tag{2.158}$$

$$\frac{1}{N_2}\mathbb{E}[\tilde{\mathbf{\Omega}}_2^{\mathsf{H}}\tilde{\mathbf{\Omega}}_2] = \bar{\mathbf{R}}_2 - \bar{\mathbf{R}}_2\mathbf{P}_{2\tau}^{\frac{1}{2}}\Big(\mathbf{P}_{2\tau}^{\frac{1}{2}}\bar{\mathbf{R}}_2\mathbf{P}_{2\tau}^{\frac{1}{2}} + \mathbf{I}_{s_2+s_0}\Big)^{-1}\mathbf{P}_{2\tau}^{\frac{1}{2}}\bar{\mathbf{R}}_2 = \big(\bar{\mathbf{R}}_2^{-1} + \mathbf{P}_{2\tau}\big)^{-1}. \tag{2.159}$$

<u>*Lower bounding $I(\mathbf{Y}_2; \mathbf{X}_2, \mathbf{X}_0)$:*</u> Using the chain rule,

$$\begin{aligned}
I(\mathbf{Y}_2; \mathbf{X}_2, \mathbf{X}_0) &= I(\mathbf{Y}_{2\tau}, \mathbf{Y}_{2\delta a}, \mathbf{Y}_{2\delta b}; \mathbf{S}_0, \mathbf{S}_{2a}, \mathbf{S}_{2b}) && (2.160)\\
&= I(\mathbf{Y}_{2\delta a}, \mathbf{Y}_{2\delta b}; \mathbf{S}_0, \mathbf{S}_{2a}, \mathbf{S}_{2b} \,|\, \mathbf{Y}_{2\tau}) + \underbrace{I(\mathbf{Y}_{2\tau}; \mathbf{S}_0, \mathbf{S}_{2a}, \mathbf{S}_{2b})}_{=0} && (2.161)\\
&= I(\mathbf{Y}_{2\delta a}, \mathbf{Y}_{2\delta b}; \mathbf{S}_0, \mathbf{S}_{2a}, \mathbf{S}_{2b} \,|\, \hat{\mathbf{\Omega}}_2) && (2.162)\\
&= I(\mathbf{Y}_{2\delta a}; \mathbf{S}_{2a} \,|\, \hat{\mathbf{\Omega}}_2) + \underbrace{I(\mathbf{Y}_{2\delta a}; \mathbf{S}_0, \mathbf{S}_{2b} \,|\, \mathbf{S}_{2a}, \hat{\mathbf{\Omega}}_2}_{=0} \\
&\quad + \underbrace{I(\mathbf{Y}_{2\delta b}; \mathbf{S}_0, \mathbf{S}_{2b} \,|\, \mathbf{Y}_{2\delta a}, \hat{\mathbf{\Omega}}_2)}_{\geq I(\mathbf{Y}_{2\delta b}; \mathbf{S}_0, \mathbf{S}_{2b} \,|\, \hat{\mathbf{\Omega}}_2)} + \underbrace{I(\mathbf{Y}_{2\delta b}; \mathbf{S}_{2a} \,|\, \mathbf{S}_0, \mathbf{S}_{2b}, \mathbf{Y}_{2\delta a}, \hat{\mathbf{\Omega}}_2)}_{=0} && (2.163)\\
&\geq I(\mathbf{Y}_{2\delta a}; \mathbf{S}_{2a} \,|\, \hat{\mathbf{\Omega}}_{22}) + I(\mathbf{Y}_{2\delta b}; \mathbf{S}_0, \mathbf{S}_{2b} \,|\, \hat{\mathbf{\Omega}}_2). && (2.164)
\end{aligned}$$

Define $\bar{\mathbf{\Omega}}_2 \in \mathbb{C}^{N_2 \times (s_2+s_0)}$ with independent rows obeying $\mathcal{CN}\big(\mathbf{0}^t, \mathbf{P}_{2\tau}^{\frac{1}{2}}(\mathbf{P}_{2\tau}^{\frac{1}{2}}\bar{\mathbf{R}}_2\mathbf{P}_{2\tau}^{\frac{1}{2}}+\mathbf{I}_{s_2+s_0})^{-1}\mathbf{P}_{2\tau}^{\frac{1}{2}}\big)$. Following analysis similar to (2.60) in Theorem 4,

$$\begin{aligned}
&I(\mathbf{Y}_{2\delta a}; \mathbf{S}_{2a} \,|\, \hat{\mathbf{\Omega}}_{22}) \\
&\geq \Big(s_1 - s_2\Big)\mathbb{E}\bigg[\log\det\bigg(\mathbf{I}_{N_2} + \frac{\rho_{2\delta}}{\rho_{2\delta}\mathrm{tr}\big(\bar{\mathbf{R}}_{22}^{\mathsf{H}}(\bar{\mathbf{R}}_2 + \bar{\mathbf{R}}_2\mathbf{P}_{2\tau}\bar{\mathbf{R}}_2)^{-1}\bar{\mathbf{R}}_{22}\big) + s_2}\bar{\mathbf{\Omega}}_2\bar{\mathbf{R}}_{22}\bar{\mathbf{R}}_{22}^{\mathsf{H}}\bar{\mathbf{\Omega}}_2^{\mathsf{H}}\bigg)\bigg]
\end{aligned} \tag{2.165}$$

and

$$\begin{aligned}
&I(\mathbf{Y}_{2\delta b}; \mathbf{S}_0, \mathbf{S}_{2b} \,|\, \hat{\mathbf{\Omega}}_2) \\
&\geq \Big(T - s_1 - s_0\Big)\mathbb{E}\bigg[\log\det\bigg(\mathbf{I}_{N_2} + \frac{1}{\mathrm{tr}\big((\bar{\mathbf{R}}_2^{-1} + \mathbf{P}_{2\tau})^{-1}\mathbf{P}_{2\delta}\big) + 1}\bar{\mathbf{\Omega}}_2\bar{\mathbf{R}}_2\mathbf{P}_{2\delta}\bar{\mathbf{R}}_2^{\mathsf{H}}\bar{\mathbf{\Omega}}_2^{\mathsf{H}}\bigg)\bigg].
\end{aligned} \tag{2.166}$$

<u>*Lower bounding $I(\mathbf{Y}_2; \mathbf{X}_2 \,|\, \mathbf{X}_0)$:*</u> We write $\mathbf{Y}_{2\delta} \triangleq [\mathbf{Y}_{2\delta a} \ \mathbf{Y}_{2\delta b}]$ as

$$\mathbf{Y}_{2\delta} = \sqrt{\frac{\rho_{2\delta}}{s_2}}\mathbf{G}_2\mathbf{\Sigma}_2^{\frac{1}{2}}\mathbf{\Phi}_{22}\mathbf{S}_2 + \sqrt{\frac{\rho_{0\delta}}{s_0}}\mathbf{G}_2\mathbf{\Sigma}_2^{\frac{1}{2}}\mathbf{\Phi}_{20}[\mathbf{0} \ \mathbf{S}_0] + \mathbf{W}_{2[s_2+s_0+1:T]}. \tag{2.167}$$

Similar to $I(\mathbf{Y}_1; \mathbf{X}_1 \,|\, \mathbf{X}_0)$, using interference cancellation and wort-case additive noise,

$$
\begin{aligned}
&I(\mathbf{Y}_2; \mathbf{X}_2 \,|\, \mathbf{X}_0) \\
&= I\big(\mathbf{Y}_{2\delta}; \mathbf{S}_2 \big| \mathbf{S}_0, \hat{\mathbf{\Omega}}_2\big) && (2.168)\\
&= I\Big(\mathbf{Y}_{2\delta} - \sqrt{\frac{\rho_{0\delta}}{s_0}} \hat{\mathbf{\Omega}}_{20} [\mathbf{0} \ \mathbf{S}_0]; \mathbf{S}_2 \big| \mathbf{S}_0, \hat{\mathbf{\Omega}}_2\Big) && (2.169)\\
&\geq \Big(s_1 - s_2\Big) \mathbb{E}\left[\log\det\left(\mathbf{I}_{N_2} + \frac{\rho_{2\delta}}{\rho_{2\delta} \mathrm{tr}\big(\bar{\mathbf{R}}_{22}^{\mathsf{H}} (\bar{\mathbf{R}}_2 + \bar{\mathbf{R}}_2 \mathbf{P}_{2\tau} \bar{\mathbf{R}}_2)^{-1} \bar{\mathbf{R}}_{22}\big) + s_2} \bar{\mathbf{\Omega}}_2 \bar{\mathbf{R}}_{22} \bar{\mathbf{R}}_2 \bar{\mathbf{R}}_2^{\mathsf{H}} \bar{\mathbf{R}}_{22}^{\mathsf{H}} \bar{\mathbf{\Omega}}_2^{\mathsf{H}} \right) \right] \\
&+ \Big(T - s_1 - s_0\Big) \mathbb{E}\left[\log\det\left(\mathbf{I}_{N_2} + \frac{\rho_{2\delta}}{s_2 \big[\mathrm{tr}\big((\bar{\mathbf{R}}_2^{-1} + \mathbf{P}_{2\tau})^{-1} \mathbf{P}_{2\delta}\big) + 1\big]} \bar{\mathbf{\Omega}}_2 \bar{\mathbf{R}}_{22} \bar{\mathbf{R}}_2 \bar{\mathbf{R}}_2^{\mathsf{H}} \bar{\mathbf{R}}_{22}^{\mathsf{H}} \bar{\mathbf{\Omega}}_2^{\mathsf{H}} \right) \right].
\end{aligned}
$$
$$(2.170)$$

<u>*Lower bounding $I(\mathbf{Y}_2; \mathbf{X}_0 \,|\, \mathbf{X}_2)$:*</u> Again, using interference cancellation and a similar analysis as for (2.60) in Theorem 4,

$$
\begin{aligned}
&I(\mathbf{Y}_2; \mathbf{X}_0 \,|\, \mathbf{X}_2) \\
&\geq I\big(\mathbf{Y}_{2\delta b}; \mathbf{S}_0 \big| \mathbf{S}_{2b}, \hat{\mathbf{\Omega}}_2\big) && (2.171)\\
&= I\Big(\mathbf{Y}_{2\delta b} - \sqrt{\frac{\rho_{2\delta}}{s_2}} \hat{\mathbf{\Omega}}_{22} \mathbf{S}_{2b}; \mathbf{S}_0 \ \big| \ \mathbf{S}_{2b}, \hat{\mathbf{\Omega}}_2\Big) && (2.172)\\
&\geq \Big(T - s_1 - s_0\Big) \mathbb{E}\left[\log\det\left(\mathbf{I}_{N_2} + \frac{\rho_{0\delta}}{s_0 \big[\mathrm{tr}\big((\bar{\mathbf{R}}_2^{-1} + \mathbf{P}_{2\tau})^{-1} \mathbf{P}_{2\delta}\big) + 1\big]} \bar{\mathbf{\Omega}}_2 \bar{\mathbf{R}}_{20} \bar{\mathbf{R}}_2 \bar{\mathbf{R}}_2^{\mathsf{H}} \bar{\mathbf{R}}_{20}^{\mathsf{H}} \bar{\mathbf{\Omega}}_2^{\mathsf{H}} \right) \right].
\end{aligned}
$$
$$(2.173)$$

Substituting (2.165) and (2.166) into (2.164), then substituting (2.139), (2.147), (2.154), (2.164), (2.170), and (2.173) into (2.76)-(2.78), and taking the convex hull over all possible power allocation satisfying (2.129) and all feasible values of s_0, s_1, s_2, an achievable rate region is found with rate splitting for the broadcast channel. This concludes the proof of Theorem 5.

2.8.3 Proof of Theorem 6

Under product superposition, the input to the channel is constructed as follows:

$$\mathbf{X} = [\mathbf{V}_0 \ \mathbf{V}_2] \mathbf{X}_1 \mathbf{X}_2, \tag{2.174}$$

with

$$\mathbf{X}_1 = \begin{bmatrix} \sqrt{\nu_{1\tau}}\mathbf{I}_{s_0} & \sqrt{\frac{\nu_{1\delta}}{s_0}}\mathbf{S}_1 \\ \mathbf{0} & \sqrt{\nu_{1a}}\mathbf{I}_{s_2} \end{bmatrix}, \tag{2.175}$$

$$\mathbf{X}_2 = \begin{bmatrix} \sqrt{\rho_{2\tau}}\mathbf{I}_{s_2+s_0} & \sqrt{\frac{\rho_{2\delta}}{s_2+s_0}}\mathbf{S}_2 \end{bmatrix}, \tag{2.176}$$

where $\mathbf{S}_1 \in \mathbb{C}^{s_0 \times s_2}$ and $\mathbf{S}_2 \in \mathbb{C}^{(s_2+s_0)\times(T-s_2-s_0)}$ are the data matrices of User 1 and User 2 respectively, both contain i.i.d. $\mathcal{CN}(0,1)$ symbols. As in earlier developments, integers s_0, s_1, s_2 are designed to allocate transmit dimensions to the components of product superposition, and take values in the range $s_0 \le r_0$ and $s_2 \le r_2 - r_0$.

The power constraint $\mathbb{E}\left[\mathrm{tr}\left(\mathbf{X}^{\mathsf{H}}\mathbf{X}\right)\right] \le \rho T$ translates to

$$(s_0\nu_{1\tau} + s_2(\nu_{1\delta}+\nu_{1a}))\Big(\rho_{2\tau} + \frac{T-s_2-s_0}{s_2+s_0}\rho_{2\delta}\Big) \le \rho T. \tag{2.177}$$

In the first s_2+s_0 channel uses, User 1 receives

$$\mathbf{Y}_{1[1:s_2+s_0]} = \sqrt{\rho_{2\tau}}\mathbf{G}_1\mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{\Phi}_{10}\begin{bmatrix} \sqrt{\nu_{1\tau}}\mathbf{I}_{s_0} & \sqrt{\frac{\nu_{1\delta}}{s_0}}\mathbf{S}_1 \end{bmatrix} + \mathbf{W}_{1[1:s_2+s_0]} \tag{2.178}$$

$$= \Bigg[\underbrace{\sqrt{\nu_{1\tau}\rho_{2\tau}}\mathbf{G}_1\mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{\Phi}_{10} + \mathbf{W}_{1[1:s_0]}}_{\mathbf{Y}_{1\tau}} \quad \underbrace{\sqrt{\frac{\nu_{1\delta}\rho_{2\tau}}{s_0}}\mathbf{G}_1\mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{\Phi}_{10}\mathbf{S}_1 + \mathbf{W}_{1[s_0+1:s_2+s_0]}}_{\mathbf{Y}_{1\delta}} \Bigg]. \tag{2.179}$$

Following Lemma 3, User 1 estimates the equivalent channel $\mathbf{G}_1\mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{\Phi}_{10}$ using a MMSE estimator based on $\mathbf{Y}_{1\tau}$ as

$$\hat{\mathbf{\Omega}}_{10} = \sqrt{\nu_{1\tau}\rho_{2\tau}}\mathbf{Y}_{1\tau}\Big(\nu_{1\tau}\rho_{2\tau}\breve{\mathbf{R}}_{10} + \mathbf{I}_{s_0}\Big)^{-1}\breve{\mathbf{R}}_{10}. \tag{2.180}$$

The estimate $\hat{\mathbf{\Omega}}_{10}$ and the estimation error $\tilde{\mathbf{\Omega}}_{10} = \mathbf{G}_1\mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{\Phi}_{10} - \hat{\mathbf{\Omega}}_{10}$ have zero mean and row covariance

$$\frac{1}{N_1}\mathbb{E}[\hat{\mathbf{\Omega}}_{10}^{\mathsf{H}}\hat{\mathbf{\Omega}}_{10}] = \nu_{1\tau}\rho_{2\tau}\breve{\mathbf{R}}_{10}\Big(\nu_{1\tau}\rho_{2\tau}\breve{\mathbf{R}}_{10} + \mathbf{I}_{s_0}\Big)^{-1}\breve{\mathbf{R}}_{10}, \tag{2.181}$$

$$\frac{1}{N_1}\mathbb{E}[\tilde{\mathbf{\Omega}}_{10}^{\mathsf{H}}\tilde{\mathbf{\Omega}}_{10}] = \breve{\mathbf{R}}_{10} - \nu_{1\tau}\rho_{2\tau}\breve{\mathbf{R}}_{10}\Big(\nu_{1\tau}\rho_{2\tau}\breve{\mathbf{R}}_{10} + \mathbf{I}_{s_0}\Big)^{-1}\breve{\mathbf{R}}_{10}\Big(\breve{\mathbf{R}}_{10}^{-1} + \nu_{1\tau}\rho_{2\tau}\mathbf{I}_{s_0}\Big)^{-1}. \tag{2.182}$$

Using data processing inequality,

$$I(\mathbf{Y}_1;\mathbf{X}_1) \geq I(\mathbf{Y}_{1[1:s_2+s_0]};\mathbf{X}_1) = I(\mathbf{Y}_{1\delta};\mathbf{S}_1 \,|\, \mathbf{Y}_{1\tau}) = I(\mathbf{Y}_{1\delta};\mathbf{S}_1 \,|\, \hat{\mathbf{\Omega}}_{10}). \tag{2.183}$$

Then, using the worst-case noise argument and Lemma 2, the following lower bound on $I(\mathbf{Y}_{1\delta};\mathbf{S}_1 \,|\, \hat{\mathbf{\Omega}}_{10})$, is established, giving an achievable rate for User 1:

$$R_1 = \frac{s_2}{T}\mathbb{E}\left[\log\det\left(\mathbf{I}_{N_1} + \frac{\nu_{1\delta}\rho_{2\tau}}{s_0 + \nu_{1\delta}\rho_{2\tau}\mathrm{tr}\left((\breve{\mathbf{R}}_{10}^{-1} + \nu_{1\tau}\rho_{2\tau}\mathbf{I}_{s_0})^{-1}\right)}\hat{\mathbf{\Omega}}_{10}\hat{\mathbf{\Omega}}_{10}^{\mathsf{H}}\right)\right] \tag{2.184}$$

The received signal at User 2 is

$$\mathbf{Y}_2 = \mathbf{G}_2\mathbf{\Sigma}_2^{\frac{1}{2}}\mathbf{\Phi}_2\mathbf{X}_1\left[\sqrt{\rho_{2\tau}}\mathbf{I}_{s_2+s_0} \quad \sqrt{\frac{\rho_{2\delta}}{s_2+s_0}}\mathbf{S}_2\right] + \mathbf{W}_2 \tag{2.185}$$

$$= \Bigg[\underbrace{\sqrt{\rho_{2\tau}}\mathbf{G}_{2e} + \mathbf{W}_{2[1:s_2+s_0]}}_{\mathbf{Y}_{2\tau}} \quad \underbrace{\sqrt{\frac{\rho_{2\delta}}{s_2+s_0}}\mathbf{G}_{2e}\mathbf{S}_2 + \mathbf{W}_{2[s_2+s_0+1:T]}}_{\mathbf{Y}_{2\delta}}\Bigg], \tag{2.186}$$

where $\mathbf{G}_{2e} \triangleq \mathbf{G}_2\mathbf{\Sigma}_2^{\frac{1}{2}}\mathbf{\Phi}_2\mathbf{X}_1$ is the equivalent channel with the correlation matrix

$$\mathbf{R}_{2e} \triangleq \frac{1}{N_2}\mathbb{E}\left[\mathbf{G}_{2e}^{\mathsf{H}}\mathbf{G}_{2e}\right] = \begin{bmatrix} \nu_{1\tau}\breve{\mathbf{R}}_{20} & \sqrt{\nu_{1\tau}\nu_{1a}}\mathbf{\Phi}_{20}^{\mathsf{H}}\mathbf{\Sigma}_2\mathbf{\Phi}_{22} \\ \sqrt{\nu_{1\tau}\nu_{1a}}\mathbf{\Phi}_{22}^{\mathsf{H}}\mathbf{\Sigma}_2\mathbf{\Phi}_{20} & \frac{\nu_{1\delta}}{s_0}\mathrm{tr}\left(\breve{\mathbf{R}}_{20}\right)\mathbf{I}_{s_2} + \nu_{1a}\breve{\mathbf{R}}_{22} \end{bmatrix}. \tag{2.187}$$

Following Lemma 3, User 2 estimates the equivalent channel $\mathbf{G}_{2e}$ using a MMSE estimator based on $\mathbf{Y}_{2\tau}$ as

$$\hat{\mathbf{G}}_{2e} = \sqrt{\rho_{2\tau}}\mathbf{Y}_{2\tau}\left(\rho_{2\tau}\mathbf{R}_{2e} + \mathbf{I}_{s_2+s_0}\right)^{-1}\mathbf{R}_{2e}. \tag{2.188}$$

The estimate $\hat{\mathbf{G}}_{2e}$ and the estimation error $\tilde{\mathbf{G}}_{2e} = \mathbf{G}_{2e} - \hat{\mathbf{G}}_{2e}$ have zero mean and row covariance

$$\frac{1}{N_2}\mathbb{E}[\hat{\mathbf{G}}_{2e}^{\mathsf{H}}\hat{\mathbf{G}}_{2e}] = \rho_{2\tau}\mathbf{R}_{2e}\left(\rho_{2\tau}\mathbf{R}_{2e} + \mathbf{I}_{s_2+s_0}\right)^{-1}\mathbf{R}_{2e}, \tag{2.189}$$

$$\frac{1}{N_2}\mathbb{E}[\tilde{\mathbf{G}}_{2e}^{\mathsf{H}}\tilde{\mathbf{G}}_{2e}] = \mathbf{R}_{2e} - \rho_{2\tau}\mathbf{R}_{2e}\left(\rho_{2\tau}\mathbf{R}_{2e} + \mathbf{I}_{s_2+s_0}\right)^{-1}\mathbf{R}_{2e} = \left(\mathbf{R}_{2e}^{-1} + \rho_{2\tau}\mathbf{I}_{s_2+s_0}\right)^{-1}. \tag{2.190}$$

Using the worst-case noise argument and Lemma 2, the following achievable rate for User 2 is established:

$$R_2 = \Big(1 - \frac{s_2 + s_0}{T}\Big)\mathbb{E}\left[\log\det\left(\mathbf{I}_{N_2} + \frac{\rho_{2\delta}}{s_2 + s_0 + \rho_{2\delta}\mathrm{tr}\big((\mathbf{R}_{2e}^{-1} + \rho_{2\tau}\mathbf{I}_{s_2+s_0})^{-1}\big)}\hat{\mathbf{G}}_{2e}\hat{\mathbf{G}}_{2e}^{\mathsf{H}}\right)\right] \tag{2.191}$$

where the distribution of $\hat{\mathbf{G}}_{2e}$ is imposed by (2.188).

From (2.184) and (2.191), the rate pair (R_1, R_2) is achievable. By swapping the users' role, another achievable rate pair is obtained. The overall achievable rate region is the convex hull of these pairs over all possible power allocations satisfying (2.177) and all feasible values of s_0, s_1, s_2. This concludes the proof of Theorem 6.

2.8.4 Proof of Theorem 7

The transmitted signal is

$$\mathbf{X} = [\mathbf{V}_0\ \mathbf{V}_1]\ \mathbf{X}_2'\mathbf{X}_1 + \mathbf{V}_2\mathbf{X}_2 \tag{2.192}$$

with

$$\mathbf{X}_1 = \left[\sqrt{\rho_{1\tau}}\mathbf{I}_{s_1+s_0}\ \sqrt{\frac{\rho_{1\delta}}{s_1+s_0}}\mathbf{S}_1\right] \in \mathbb{C}^{(s_1+s_0)\times T}, \tag{2.193}$$

$$\mathbf{X}_2 = \left[\mathbf{0}_{s_2\times s_0}\ \sqrt{\rho_{2\tau}}\mathbf{I}_{s_2}\ \sqrt{\frac{\rho_{2\delta}}{s_2}}\mathbf{S}_2\right] \in \mathbb{C}^{s_2\times T}, \tag{2.194}$$

$$\mathbf{X}_2' = \begin{bmatrix} \sqrt{\nu_{2\tau}}\mathbf{I}_{s_0} & \left[\mathbf{0}_{s_0\times s_2}\ \sqrt{\frac{\nu_{2\delta}}{s_0}}\mathbf{S}_2'\right] \\ \mathbf{0} & \sqrt{\nu_{2a}}\mathbf{I}_{s_1} \end{bmatrix} \in \mathbb{C}^{(s_1+s_0)\times(s_1+s_0)}, \tag{2.195}$$

where $\mathbf{S}_1 \in \mathbb{C}^{(s_1+s_0)\times(T-s_1-s_0)}$, $\mathbf{S}_2 \in \mathbb{C}^{s_2\times(T-s_2-s_0)}$, and $\mathbf{S}_2' \in \mathbb{C}^{s_0\times(s_1-s_2)}$ are data matrices containing $\mathcal{CN}(0,1)$ entries. The power constraint $\mathbb{E}\left[\mathrm{tr}\left(\mathbf{X}^{\mathsf{H}}\mathbf{X}\right)\right] \le \rho T$ translates to

$$\big(s_0\nu_{2\tau} + s_1\nu_{2a} + (s_1 - s_2)\nu_{2\delta}\big)\Big(\rho_{1\tau} + \frac{T - s_1 - s_0}{s_1 + s_0}\rho_{1\delta}\Big) + s_2\rho_{2\tau} + (T - s_2 - s_0)\rho_{2\delta} \le \rho T. \tag{2.196}$$

We begin by analyzing the rate of User 1. The received signal at User 1 is

$$\mathbf{Y}_1 = \mathbf{G}_1\mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{\Phi}_1\mathbf{X}_2'\left[\sqrt{\rho_{1\tau}}\mathbf{I}_{s_1+s_0}\ \sqrt{\frac{\rho_{1\delta}}{s_1+s_0}}\mathbf{S}_1\right] + \mathbf{W}_1 \tag{2.197}$$

$$= \left[\underbrace{\sqrt{\rho_{1\tau}}\mathbf{G}_{1e} + \mathbf{W}_{1[1:s_1+s_0]}}_{\mathbf{Y}_{1\tau}}\quad \underbrace{\sqrt{\frac{\rho_{1\delta}}{s_1+s_0}}\mathbf{G}_{1e}\mathbf{S}_1 + \mathbf{W}_{1[s_1+s_0+1:T]}}_{\mathbf{Y}_{1\delta}}\right], \tag{2.198}$$

where $\mathbf{G}_{1e} \triangleq \mathbf{G}_1\mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{\Phi}_1\mathbf{X}_2'$ is the equivalent channel with correlation matrix

$$\mathbf{R}_{1e} \triangleq \frac{1}{N_1}\mathbb{E}\left[\mathbf{G}_{1e}^{\mathsf{H}}\mathbf{G}_{1e}\right] = \begin{bmatrix} \nu_{2\tau}\breve{\mathbf{R}}_{10} & \sqrt{\nu_{2\tau}\nu_{2a}}\mathbf{\Phi}_{10}^{\mathsf{H}}\mathbf{\Sigma}_1\mathbf{\Phi}_{11} \\ \sqrt{\nu_{2\tau}\nu_{2a}}\mathbf{\Phi}_{11}^{\mathsf{H}}\mathbf{\Sigma}_1\mathbf{\Phi}_{10} & \begin{bmatrix}\mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{\nu_{2\delta}}{s_0}\mathrm{tr}\left(\breve{\mathbf{R}}_{10}\right)\mathbf{I}_{s_1-s_2}\end{bmatrix} + \nu_{2a}\breve{\mathbf{R}}_{22}\end{bmatrix}. \tag{2.199}$$

Following Lemma 3, User 1 estimates the equivalent channel $\mathbf{G}_{1e}$ using a MMSE estimator based on $\mathbf{Y}_{1\tau}$ as

$$\hat{\mathbf{G}}_{1e} = \sqrt{\rho_{1\tau}}\mathbf{Y}_{1\tau}\left(\rho_{1\tau}\mathbf{R}_{1e} + \mathbf{I}_{s_1+s_0}\right)^{-1}\mathbf{R}_{1e}. \tag{2.200}$$

The estimate $\hat{\mathbf{G}}_{1e}$ and the estimation error $\tilde{\mathbf{G}}_{1e} = \mathbf{G}_{1e} - \hat{\mathbf{G}}_{1e}$ have zero mean and row covariance

$$\frac{1}{N_1}\mathbb{E}[\hat{\mathbf{G}}_{1e}^{\mathsf{H}}\hat{\mathbf{G}}_{1e}] = \rho_{1\tau}\mathbf{R}_{1e}\left(\rho_{1\tau}\mathbf{R}_{1e} + \mathbf{I}_{s_1+s_0}\right)^{-1}\mathbf{R}_{1e}, \tag{2.201}$$

$$\frac{1}{N_1}\mathbb{E}[\tilde{\mathbf{G}}_{1e}^{\mathsf{H}}\tilde{\mathbf{G}}_{1e}] = \mathbf{R}_{1e} - \rho_{1\tau}\mathbf{R}_{1e}\left(\rho_{1\tau}\mathbf{R}_{1e} + \mathbf{I}_{s_1+s_0}\right)^{-1}\mathbf{R}_{1e} = \left(\mathbf{R}_{1e}^{-1} + \rho_{1\tau}\mathbf{I}_{s_1+s_0}\right)^{-1}. \tag{2.202}$$

Using the worst-case noise argument and Lemma 2 as before, the following achievable rate for User 1 is obtained:

$$R_1 = \left(1 - \frac{s_1+s_0}{T}\right)\mathbb{E}\left[\log\det\left(\mathbf{I}_{N_1} + \frac{\rho_{1\delta}}{s_1+s_0+\rho_{1\delta}\mathrm{tr}\left(\left(\mathbf{R}_{1e}^{-1}+\rho_{1\tau}\mathbf{I}_{s_1+s_0}\right)^{-1}\right)}\hat{\mathbf{G}}_{1e}\hat{\mathbf{G}}_{1e}^{\mathsf{H}}\right)\right], \tag{2.203}$$

where the distribution of $\hat{\mathbf{G}}_{1e}$ is imposed by (2.200).

Now, we turn to analyzing the achievable rate for User 2. The received signal at User 2 can be written as

$$\mathbf{Y}_2 = \mathbf{G}_2\mathbf{\Sigma}_2^{\frac{1}{2}}\mathbf{\Phi}_2 \begin{bmatrix} \sqrt{\nu_{2\tau}\rho_{1\tau}}\mathbf{I}_{s_0} & \mathbf{0} & \left[\sqrt{\frac{\nu_{2\delta}\rho_{1\tau}}{s_0}}\mathbf{S}_2' \ \mathbf{A}\right] \\ \mathbf{0} & \sqrt{\rho_{2\tau}}\mathbf{I}_{s_2} & \sqrt{\frac{\rho_{2\delta}}{s_2}}\mathbf{S}_2 \end{bmatrix} + \mathbf{W}_2 \tag{2.204}$$

$$= \Big[\mathbf{Y}_{2\tau} \ \underbrace{\mathbf{Y}_{2\delta a} \ \mathbf{Y}_{2\delta b}}_{\mathbf{Y}_{2\delta}}\Big], \tag{2.205}$$

where $\mathbf{A} \triangleq \left[\sqrt{\nu_{2\tau}}\mathbf{I}_{s_0} \ \mathbf{0} \ \sqrt{\frac{\nu_{2\delta}}{s_0}}\mathbf{S}_2'\right]\sqrt{\frac{\rho_{1\delta}}{s_1+s_0}}\mathbf{S}_1$ and

$$\mathbf{Y}_{2\tau} \triangleq \mathbf{G}_2\mathbf{\Sigma}_2^{\frac{1}{2}}\mathbf{\Phi}_2\mathbf{P}_{2\tau}^{\frac{1}{2}} + \mathbf{W}_{2[1:s_2+s_0]}, \tag{2.206}$$

$$\mathbf{Y}_{2\delta a} \triangleq \mathbf{G}_2\mathbf{\Sigma}_2^{\frac{1}{2}}\mathbf{\Phi}_2 \begin{bmatrix} \sqrt{\frac{\nu_{2\delta}\rho_{1\tau}}{s_0}}\mathbf{S}_2' \\ \sqrt{\frac{\rho_{2\delta}}{s_2}}\mathbf{S}_{2[1:s_1-s_2]} \end{bmatrix} + \mathbf{W}_{2[s_2+s_0+1:s_1+s_0]}, \tag{2.207}$$

$$\mathbf{Y}_{2\delta b} \triangleq \mathbf{G}_2\mathbf{\Sigma}_2^{\frac{1}{2}}\mathbf{\Phi}_2 \begin{bmatrix} \mathbf{A} \\ \sqrt{\frac{\rho_{2\delta}}{s_2}}\mathbf{S}_{2[s_1-s_2+1:T-s_2]} \end{bmatrix} + \mathbf{W}_{2[s_1+s_0+1:T]}, \tag{2.208}$$

where $\mathbf{P}_{2\tau} \triangleq \begin{bmatrix} \nu_{2\tau}\rho_{1\tau}\mathbf{I}_{s_0} & \mathbf{0} \\ \mathbf{0} & \rho_{2\tau}\mathbf{I}_{s_2} \end{bmatrix}$. The rate that User 2 can achieve is $\frac{1}{T}I(\mathbf{Y}_2;\mathbf{S}_2',\mathbf{S}_2)$ bits/channel use with

$$I(\mathbf{Y}_2;\mathbf{S}_2',\mathbf{S}_2) = I(\mathbf{Y}_{2\tau},\mathbf{Y}_{2\delta};\mathbf{S}_2',\mathbf{S}_2) \tag{2.209}$$

$$= \underbrace{I(\mathbf{Y}_{2\tau};\mathbf{S}_2',\mathbf{S}_2)}_{=0} + I(\mathbf{Y}_{2\delta};\mathbf{S}_2',\mathbf{S}_2 \,|\, \mathbf{Y}_{2\tau}) \tag{2.210}$$

$$= I(\mathbf{Y}_{2\delta};\mathbf{S}_2',\mathbf{S}_2,\mathbf{A} \,|\, \mathbf{Y}_{2\tau}) - I(\mathbf{Y}_{2\delta};\mathbf{A} \,|\, \mathbf{Y}_{2\tau},\mathbf{S}_2',\mathbf{S}_2) \tag{2.211}$$

where the second and third equalities follow from the chain rule.

Define $\bar{\mathbf{\Omega}}_2 \in \mathbb{C}^{N_2\times(s_2+s_0)}$ with independent rows obeying $\mathcal{CN}\big(\mathbf{0}^t, \bar{\mathbf{R}}_2^{\mathsf{H}}\big(\bar{\mathbf{R}}_2+\mathbf{P}_{2\tau}^{-1}\big)^{-1}\bar{\mathbf{R}}_2\big)$ and $\bar{\mathbf{\Omega}}_{20} \in \mathbb{C}^{N_2\times s_0}$ with independent rows obeying $\mathcal{CN}\big(\mathbf{0}^t, \breve{\mathbf{R}}_{20}\big)$. For $I(\mathbf{Y}_{2\delta};\mathbf{S}_2',\mathbf{S},\mathbf{A} \,|\, \mathbf{Y}_{2\tau})$, using the worst-case noise argument and Lemma 2 as before, we have the bound

$$
\begin{aligned}
& I(\mathbf{Y}_{2\delta}; \mathbf{S}_2', \mathbf{S}_{2[1:s_1-s_2]}, \mathbf{A} \big| \mathbf{Y}_{2\tau}) \\
& \geq \big(s_1 - s_2\big)\mathbb{E}\bigg[\log\det\bigg(\mathbf{I}_{N_2} + \frac{1}{\mathrm{tr}\big((\bar{\mathbf{R}}_2^{-1} + \mathbf{P}_{2\tau})^{-1}\mathbf{P}_{2\delta a}\big) + 1}\bar{\boldsymbol{\Omega}}_2\mathbf{P}_{2\delta a}\bar{\boldsymbol{\Omega}}_2^{\mathsf{H}}\bigg)\bigg] \qquad (2.212) \\
& + \big(T - s_1 - s_0\big)\mathbb{E}\bigg[\log\det\bigg(\mathbf{I}_{N_2} + \frac{1}{\mathrm{tr}\big((\bar{\mathbf{R}}_2^{-1} + \mathbf{P}_{2\tau})^{-1}\mathbf{P}_{2\delta b}\big) + 1}\bar{\boldsymbol{\Omega}}_2\mathbf{P}_{2\delta b}\bar{\boldsymbol{\Omega}}_2^{\mathsf{H}}\bigg)\bigg]
\end{aligned}
$$

$$(2.213)$$

where $\mathbf{P}_{2\delta a} \triangleq \begin{bmatrix} \frac{\nu_{2\delta}\rho_{1\tau}}{s_0}\mathbf{I}_{s_0} & \mathbf{0} \\ \mathbf{0} & \frac{\rho_{2\delta}}{s_2}\mathbf{I}_{s_2} \end{bmatrix}$ and $\mathbf{P}_{2\delta b} \triangleq \begin{bmatrix} \frac{\rho_{1\delta}}{T-s_1-s_0}\big(\nu_{1\tau} + \nu_{2\delta}\frac{s_1-s_2}{s_0}\big)\mathbf{I}_{s_0} & \mathbf{0} \\ \mathbf{0} & \frac{\rho_{2\delta}}{s_2}\mathbf{I}_{s_2} \end{bmatrix}$.

The term $I(\mathbf{Y}_{2\delta}; \mathbf{A} \big| \mathbf{Y}_{2\tau}, \mathbf{S}_2', \mathbf{S}_2)$ can be upper bounded as follows:

$$
\begin{aligned}
& I(\mathbf{Y}_{2\delta}; \mathbf{A} \,|\, \mathbf{Y}_{2\tau}, \mathbf{S}_2', \mathbf{S}_2) \\
& = I(\mathbf{Y}_{2\delta b}; \mathbf{A} \,|\, \mathbf{S}_2', \mathbf{S}_2, \mathbf{Y}_{2\tau}) && (2.214) \\
& = I(\mathbf{Y}_{2\delta b}; \mathbf{A} \,|\, \mathbf{S}_2, \mathbf{Y}_{2\tau}) - I(\mathbf{Y}_{2\delta b}; \mathbf{S}_2' \,|\, \mathbf{S}_2, \mathbf{Y}_{2\tau}) && (2.215) \\
& \leq I(\mathbf{Y}_{2\delta b}; \mathbf{A} \,|\, \mathbf{S}_{2[s_1-s_2+1:T-s_2-s_0]}, \mathbf{Y}_{2\tau}) && (2.216) \\
& = h(\mathbf{A}|\mathbf{S}_{2[s_1-s_2+1:T-s_2-s_0]}, \mathbf{Y}_{2\tau}) - h(\mathbf{A}|\mathbf{S}_{2[s_1-s_2+1:T-s_2-s_0]}, \mathbf{Y}_{2\tau}, \mathbf{Y}_{2\delta b}) && (2.217) \\
& \leq h(\mathbf{A}|\mathbf{S}_{2[s_1-s_2+1:T-s_2-s_0]}, \mathbf{Y}_{2\tau}) - h(\mathbf{A}|\mathbf{S}_{2[s_1-s_2+1:T-s_2-s_0]}, \mathbf{Y}_{2\tau}, \mathbf{Y}_{2\delta b}, \mathbf{G}_2\boldsymbol{\Sigma}_2^{\frac{1}{2}}\boldsymbol{\Phi}_2) && (2.218) \\
& = h(\mathbf{A}|\mathbf{S}_{2[s_1-s_2+1:T-s_2-s_0]}, \mathbf{G}_2\boldsymbol{\Sigma}_2^{\frac{1}{2}}\boldsymbol{\Phi}_2) - h(\mathbf{A}|\mathbf{S}_{2[s_1-s_2+1:T-s_2-s_0]}, \mathbf{Y}_{2\delta b}, \mathbf{G}_2\boldsymbol{\Sigma}_2^{\frac{1}{2}}\boldsymbol{\Phi}_2) && (2.219) \\
& = I(\mathbf{Y}_{2\delta b}; \mathbf{A} \,|\, \mathbf{S}_{2[s_1-s_2+1:T-s_2-s_0]}, \mathbf{G}_2\boldsymbol{\Sigma}_2^{\frac{1}{2}}\boldsymbol{\Phi}_2) && (2.220) \\
& = I\bigg(\mathbf{Y}_{2\delta b} - \sqrt{\frac{\rho_{2\delta}}{s_2}}\mathbf{G}_2\boldsymbol{\Sigma}_2^{\frac{1}{2}}\boldsymbol{\Phi}_{22}\mathbf{S}_{2[s_1-s_2+1:T-s_2-s_0]}; \mathbf{A} \,\bigg|\, \mathbf{S}_{2[s_1-s_2+1:T-s_2-s_0]}, \mathbf{G}_2\boldsymbol{\Sigma}_2^{\frac{1}{2}}\boldsymbol{\Phi}_{20}, \mathbf{G}_2\boldsymbol{\Sigma}_2^{\frac{1}{2}}\boldsymbol{\Phi}_{22}\bigg) \\
& && (2.221) \\
& = I\bigg(\mathbf{G}_2\boldsymbol{\Sigma}_2^{\frac{1}{2}}\boldsymbol{\Phi}_{20}\mathbf{A} + \mathbf{W}_{2[s_1+s_0+1:T]}; \mathbf{A} \,\bigg|\, \mathbf{G}_2\boldsymbol{\Sigma}_2^{\frac{1}{2}}\boldsymbol{\Phi}_{20}\bigg) && (2.222) \\
& = \big(T - s_1 - s_0\big)\mathbb{E}\bigg[\log\det\bigg(\mathbf{I}_{N_2} + \rho_{1\delta}\Big(\nu_{2\tau} + \nu_{2\delta}\frac{s_1 - s_2}{s_0}\Big)\bar{\boldsymbol{\Omega}}_{20}\bar{\boldsymbol{\Omega}}_{20}^{\mathsf{H}}\bigg)\bigg] && (2.223)
\end{aligned}
$$

where (2.214) and (2.215) follow from the Markov chains $\mathbf{Y}_{2\delta a} \leftrightarrow \mathbf{S}_2' \leftrightarrow \mathbf{A}$ and $\mathbf{Y}_{2\delta b} \leftrightarrow \mathbf{A} \leftrightarrow \mathbf{S}_2'$, respectively; (2.216) holds because mutual information is non-negative and both $\mathbf{Y}_{2\delta b}$

and $\mathbf{A}$ are independent of $\mathbf{S}_{2[1:s_1-s_2]}$; (2.218) holds because conditioning reduces entropy; (2.219) holds because $\mathbf{A}$ is independent of both $\mathbf{Y}_{2\tau}$ and $\mathbf{G}_2\mathbf{\Sigma}_2^{\frac{1}{2}}\mathbf{\Phi}_2$, while given $\mathbf{Y}_{2\delta b}$, $\mathbf{A}$ depends on $\mathbf{Y}_{2\tau}$ only through $\mathbf{G}_2\mathbf{\Sigma}_2^{\frac{1}{2}}\mathbf{\Phi}_2$; and in the last equality, we used that $\mathbb{E}[\mathbf{A}\mathbf{A}^{\mathsf{H}}] = \rho_{1\delta}\left(\nu_{2\tau} + \nu_{2\delta}\frac{s_1-s_2}{s_0}\right)\mathbf{I}_{s_0}$.

Substituting (2.213) and (2.223) into (2.211), an achievable rate for User 2 is obtained. This rate and (2.203) give an achievable rate pair. Taking the convex hull of this pair over all possible power allocations satisfying (2.196) and all feasible values of s_0, s_1, s_2 provides an overall achievable rate region. This concludes the proof of Theorem 7.

CHAPTER 3

SPATIAL CORRELATION IN MASSIVE MIMO

In a massive MIMO system [63], the base station needs the CSI to beamform. However, due to the large number of antennas, the overhead for channel estimation is large. On the other hand, due to the limited space between the transmit antennas, the channel responses are normally spatially correlated. In this section, we exploit the spatial correlation to reduce the training overhead and compare the scheme with the conventional training method.

We consider a multiuser massive MIMO system with a base station equipped with M antennas communicating with K single-antenna users with different spatial correlations. The channel vector corresponding to user $k \in [K]$ is $\mathbf{h}_k \in \mathbb{C}^M$. The received signal of User k at time t is $y(t) = \mathbf{h}_k^\intercal \mathbf{x}(t) + w(t)$, and during a coherence block is

$$\mathbf{y}_k^\intercal = [y(1)\ y(2)\ \ldots\ y(T)] = \mathbf{h}_k^\intercal \mathbf{X} + \mathbf{w}_k^\intercal, \tag{3.1}$$

where $\mathbf{X} = [\mathbf{x}(1)\ \mathbf{x}(2)\ \ldots\ \mathbf{x}(T)]$ and $\mathbf{w}_k = [w(1)\ w(2)\ \ldots\ w(T)]^\intercal \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_T)$. We assume that the system operates in FDD mode and focus on the downlink transmission. The transmission has two phases: the pilot phase and the data phase. During the pilot phase, pilot signal is sent so that the users can estimate the channel and then feedback the channel estimates to the base station. For simplicity and to focus on the gain of exploiting spatial correlation, we assume that feedback is perfect and instantaneous. After that, the base station sends data via beamforming.

3.1 The Two-User Case

We first consider the two-user scenario and assume that User 1 has uncorrelated channel and User 2 has spatially correlated channel of rank r_2. To extract an uncorrelated equivalent representation of $\mathbf{h}_2$, we define $\mathbf{g}_2 \in \mathbb{C}^{r_2}$ via

$$\mathbf{h}_2 = \mathbf{U}\mathbf{g}_2 \tag{3.2}$$

where $\mathbf{U} \triangleq [\mathbf{u}_1 \ \ldots \ \mathbf{u}_M]^\intercal \in \mathbb{C}^{M\times r_2}$ is a truncated unitary matrix.

Consider one coherence block. During the pilot phase, the transmitted signal is

$$\mathbf{X}_{[1:M]} = \sqrt{\rho}\ \mathrm{diag}\left(x_1, x_2, \ldots, x_M\right) \tag{3.3}$$

where $x_t = 1$ for $t \in \{1, 2, ..., r_2\}$, and x_t is a Gaussian random variable following $\mathcal{CN}(0, 1)$ for $t \in \{r_2 + 1, r_2 + 2, ..., M\}$. In time slots $t = 1, 2, \ldots, r_2$, the received signal at User 2 is $y_2(t) = \sqrt{\rho}\mathbf{g}_2^\intercal \mathbf{u}_t + w_2(t)$. User 2 estimates $\mathbf{g}_2$ with a MMSE estimator

$$\hat{\mathbf{g}}_2 = \sqrt{\rho}[\mathbf{u}_1 \ \ldots \ \mathbf{u}_{r_2}]^\mathsf{H}(\mathbf{I}_{r_2} + \rho[\mathbf{u}_1 \ \ldots \ \mathbf{u}_{r_2}][\mathbf{u}_1 \ \ldots \ \mathbf{u}_{r_2}]^\mathsf{H})^{-1}[y_2(1), \ldots, y_2(r_1)]^\intercal. \tag{3.4}$$

The estimation error is $\tilde{\mathbf{g}}_2 = \mathbf{g}_2 - \hat{\mathbf{g}}_2$. In time slots $t = r_2 + 1, \ldots, M$, User 2 receives the signal $y_2(t) = \sqrt{\rho}\mathbf{g}_2^\intercal \mathbf{u}_t x_t + w_2(t)$. User 2 uses the estimated channel to decode $[x_{r_1+1}, \ldots, x_M]$, achieving the rate

$$\Delta R_2 = \frac{M - r_2}{T}\mathbb{E}\left[\log\left(1 + \frac{\rho}{\rho\mathbb{E}\left[\|\tilde{\mathbf{g}}_2^\intercal \mathbf{u}_t\|^2\right] + 1}\|\hat{\mathbf{g}}_2^\intercal \mathbf{u}_t\|^2\right)\right]. \tag{3.5}$$

The received signal at User 1 in the pilot phase is

$$(\mathbf{y}_1^\intercal)_{[1:M]} = [y_1(1) \ \ldots \ y_1(M)] = \mathbf{h}_1^\intercal \mathbf{X} + (\mathbf{w}_1^\intercal)_{[1:M]}. \tag{3.6}$$

User 2 estimates $\mathbf{h}_1^\intercal \mathbf{X}$ by $\frac{\rho}{\rho+1}(\mathbf{y}_1^\intercal)_{[1:M]}$ and feeds back to the base station. Because the base station knows $\mathbf{X}$, it can obtain the estimation of $\mathbf{h}_1$ as $\hat{\mathbf{h}}_1 = \frac{\rho}{\rho+1}\mathbf{X}^{-\intercal}(\mathbf{y}_1)_{[1:M]}$. The estimation error is $\tilde{\mathbf{h}}_1 = \mathbf{h}_1 - \hat{\mathbf{h}}_1$.

Let $\hat{\mathbf{h}}_2 = \mathbf{U}\hat{\mathbf{g}}_2$ and $\tilde{\mathbf{h}}_2 = \mathbf{U}\tilde{\mathbf{g}}_1$. During the data phase, i.e. time slots $t = M+1, \ldots, T$, the transmitted signal via conjugate beamforming is $\mathbf{x}(t) = \sqrt{\frac{\rho}{2}}\frac{\hat{\mathbf{h}}_1^*}{\|\hat{\mathbf{h}}_1\|}s_1(t) + \sqrt{\frac{\rho}{2}}\frac{\hat{\mathbf{h}}_2^*}{\|\hat{\mathbf{h}}_2\|}s_2(t)$, where $s_k(t)$ is the data symbol for user $k \in \{1, 2\}$ following the $\mathcal{CN}(0, 1)$ distribution. The received signals at the two users are

$$y_1(t) = \sqrt{\frac{\rho}{2}}\frac{\mathbf{h}_1^\intercal \hat{\mathbf{h}}_1^*}{\|\hat{\mathbf{h}}_1\|}s_1(t) + \sqrt{\frac{\rho}{2}}\frac{\mathbf{h}_1^\intercal \hat{\mathbf{h}}_2^*}{\|\hat{\mathbf{h}}_2\|}s_2(t) + w_1(t), \tag{3.7}$$

$$y_2(t) = \sqrt{\frac{\rho}{2}}\frac{\mathbf{h}_2^\intercal \hat{\mathbf{h}}_2^*}{\|\hat{\mathbf{h}}_2\|}s_2(t) + \sqrt{\frac{\rho}{2}}\frac{\mathbf{h}_2^\intercal \hat{\mathbf{h}}_1^*}{\|\hat{\mathbf{h}}_1\|}s_1(t) + w_2(t). \tag{3.8}$$

The achievable rate for User k is:

$$R_k = \left(1 - \frac{M}{T}\right)\mathbb{E}\left[\log\left(1 + \rho_i \|\hat{\mathbf{h}}_i\|^2\right)\right], \quad k = 1, 2, \tag{3.9}$$

where the equivalent SNRs are defined as

$$\rho_1 \triangleq \left(\mathbb{E}\left[\frac{|\tilde{\mathbf{h}}_1^{\intercal}\hat{\mathbf{h}}_1^*|^2}{\|\hat{\mathbf{h}}_1\|^2} + \frac{|\mathbf{h}_1^{\intercal}\hat{\mathbf{h}}_2^*|^2}{\|\hat{\mathbf{h}}_2\|^2}\right] + \frac{2}{\rho}\right)^{-1} \tag{3.10}$$

and

$$\rho_2 \triangleq \left(\mathbb{E}\left[\frac{|\tilde{\mathbf{h}}_2^{\intercal}\hat{\mathbf{h}}_2^*|^2}{\|\hat{\mathbf{h}}_2\|^2} + \frac{|\mathbf{h}_2^{\intercal}\hat{\mathbf{h}}_1^*|^2}{\|\hat{\mathbf{h}}_1\|^2}\right] + \frac{2}{\rho}\right)^{-1}. \tag{3.11}$$

The achievable sum rate is

$$R = R_1 + R_2 + \Delta R_2. \tag{3.12}$$

For conventional transmission, the transmitter ignores the condition that two users need different number of pilots and sends M pilots over M time slots, the users estimate the channel and feedback to the transmitter. Then the transmitter communicates with the users via conjugate beamforming [63]. Figure 3.1 shows the performance of the proposed scheme in comparison with the conventional one under Rayleigh fading, $M = 32$, $T = 64$, User 1 has fully correlated channel and User 2 has uncorrelated channel.

We now generalize to the case where both users experience spatially correlated links and have partially overlapping eigenspaces. Recall that the eigendirections for the two users are $\mathbf{U}_k$, where $\mathbf{U}_k \in \mathcal{C}^{M\times r_k}$, for $k = 1, 2$. We assume without loss of generality that $r_1 \geq r_2$. We find transmit eigendirections with orthonormal columns $\mathbf{V}_0$ that are aligned with the common part of the two channel eigenspaces and $\mathbf{V}_1, \mathbf{V}_2$ that are aligned with the non-common parts, i.e., $\mathbf{V}_0 \in \mathbb{C}^{M\times r_0}$, $\mathbf{V}_1 \in \mathbb{C}^{M\times(r_1-r_0)}$, $\mathbf{V}_2 \in \mathbb{C}^{M\times(r_2-r_0)}$ such that

$$\mathrm{Span}(\mathbf{V}_0) = \mathrm{Span}(\mathbf{U}_1) \cap \mathrm{Span}(\mathbf{U}_2), \tag{3.13}$$

$$\mathrm{Span}(\mathbf{V}_1) = \mathrm{Span}(\mathbf{U}_1) \cap \mathrm{Span}(\mathbf{U}_2)^{\perp}, \tag{3.14}$$

$$\mathrm{Span}(\mathbf{V}_2) = \mathrm{Span}(\mathbf{U}_2) \cap \mathrm{Span}(\mathbf{U}_1)^{\perp}. \tag{3.15}$$

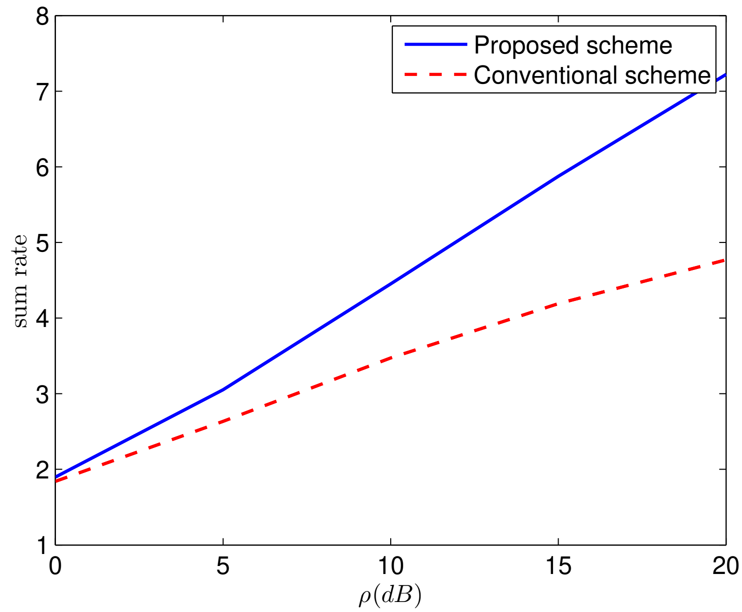

Figure 3.1. Sum rate of an FDD massive MIMO system where $K = 2, M = 32$, $T = 64$, User 1 has fully correlated channel, and User 2 has uncorrelated channel.

Therefore, we can write $\mathbf{h}_k = [\mathbf{V}_0\ \mathbf{V}_k]\mathbf{g}_k$ where $\mathbf{g}_k \in \mathbb{C}^{r_k}$, $k = 1, 2$.

The proposed scheme has two phases. The pilot phase has r_1 time slots, and the data phase has $T - r_1$ time slots. In the pilot phase, the base station sends pilots in the subspace of $\mathbf{V}_0$ in time slots 1 to r_0, $\mathbf{X}_{[1:r_0]} = \sqrt{\rho}\mathbf{V}_0^*$. The received signal at User k is

$$(\mathbf{y}_k^\intercal)_{[1:r_0]} = \sqrt{\rho}\mathbf{h}_k^\intercal \mathbf{V}_0^* + (\mathbf{w}_k^\intercal)_{[1:r_0]} = \sqrt{\rho}\mathbf{g}_k^\intercal \begin{bmatrix} \mathbf{I}_{r_0} \\ \mathbf{0}_{(r_k - r_0)\times r_0} \end{bmatrix} + (\mathbf{w}_k^\intercal)_{[1:r_0]}. \tag{3.16}$$

In the next $r_2 - r_0$ time slots, the base station sends pilots to two users simultaneously in subspaces $\mathbf{V}_1$ and $\mathbf{V}_2$, the transmitted signal is

$$\mathbf{X}_{[r_0+1:r_2]} = \sqrt{\frac{\rho}{2}}\left(\mathbf{V}_1^* \begin{bmatrix} \mathbf{I}_{r_2 - r_0} \\ \mathbf{0}_{(r_1 - r_2)\times(r_2 - r_0)} \end{bmatrix} + \mathbf{V}_2^*\right). \tag{3.17}$$

The received signals at two users are:

$$(\mathbf{y}_1^\intercal)_{[r_0+1:r_2]} = \mathbf{h}_1^\intercal \mathbf{X}_{[r_0+1:r_2]} + (\mathbf{w}_1^\intercal)_{[r_0+1:r_2]} = \sqrt{\frac{\rho}{2}}\mathbf{g}_1^\intercal \begin{bmatrix} \mathbf{0}_{r_0\times(r_2-r_0)} \\ \mathbf{I}_{r_2-r_0} \\ \mathbf{0}_{(r_1-r_2)\times(r_2-r_0)} \end{bmatrix} + (\mathbf{w}_1^\intercal)_{[r_0+1:r_2]}, \quad (3.18)$$

$$(\mathbf{y}_2^\intercal)_{[r_0+1:r_2]} = \mathbf{h}_2^\intercal \mathbf{X}_{[r_0+1:r_2]} + (\mathbf{w}_2^\intercal)_{[r_0+1:r_2]} = \sqrt{\frac{\rho}{2}}\mathbf{g}_2^\intercal \begin{bmatrix} \mathbf{0}_{r_0\times(r_2-r_0)} \\ \mathbf{I}_{r_2-r_0} \end{bmatrix} + (\mathbf{w}_2^\intercal)_{[r_0+1:r_2]}. \quad (3.19)$$

Based on $(\mathbf{y}_2^\intercal)_{[1:r_2]}$, User 2 obtains a MMSE estimates $\hat{\mathbf{g}}_2 = \frac{\sqrt{\rho/2}}{\rho/2+1}(\mathbf{y}_2)_{[1:r_2]}$ of $\mathbf{g}_2$ and feeds back to the base station. The estimation error is $\tilde{\mathbf{g}}_2 = \mathbf{g}_2 - \hat{\mathbf{g}}_2$. In time slots r_2+1 to r_1, the base station sends pilots for User 1 in the remaining eigenspaces and sends data to User 2 via beamforming as

$$\mathbf{X}_{[r_2+1:r_1]} = \sqrt{\frac{\rho}{2}}\left(\mathbf{V}_1^* \begin{bmatrix} \mathbf{0}_{r_2\times(r_1-r_2)} \\ \mathbf{I}_{r_1-r_2} \end{bmatrix} + [\mathbf{0}_{r_2\times r_0}\ \ \mathbf{V}_2^*] \frac{\hat{\mathbf{g}}_2^*}{\|(\hat{\mathbf{g}}_2^\intercal)_{[r_0+1:r_2]}\|}\mathbf{s}_{22}^\intercal\right), \quad (3.20)$$

where $\mathbf{s}_{22} \in \mathbb{C}^{r_1-r_2}$ contains i.i.d. $\mathcal{CN}(0,1)$ data symbols. The received signal at User 1 is:

$$(\mathbf{y}_1^\intercal)_{[r_2+1:r_1]} = \mathbf{h}_1^\intercal \mathbf{X}_{[r_2+1:r_1]} + (\mathbf{w}_1^\intercal)_{[r_2+1:r_1]} = \sqrt{\frac{\rho}{2}}\mathbf{g}_1^\intercal \begin{bmatrix} \mathbf{0}_{r_2\times(r_1-r_2)} \\ \mathbf{I}_{r_1-r_2} \end{bmatrix} + (\mathbf{w}_1^\intercal)_{[r_2+1:r_1]} \quad (3.21)$$

Based on $(\mathbf{y}_1^\intercal)_{[1:r_1]}$, User 1 obtains a MMSE estimates $\hat{\mathbf{g}}_1 = \frac{\sqrt{\rho/2}}{\rho/2+1}(\mathbf{y}_1)_{[1:r_1]}$ of $\mathbf{g}_1$ and feeds back to the base station. The estimation error is $\tilde{\mathbf{g}}_1 = \mathbf{g}_1 - \hat{\mathbf{g}}_1$. The received signal at User 2 is

$$(\mathbf{y}_2^\intercal)_{[r_2+1:r_1]} = \mathbf{h}_2^\intercal \mathbf{X}_{[r_2+1:r_1]} + (\mathbf{w}_2^\intercal)_{[r_2+1:r_1]} \quad (3.22)$$

$$= \sqrt{\frac{\rho}{2}}\mathbf{g}_2^\intercal \begin{bmatrix} \mathbf{0}_{r_0\times r_0} & \mathbf{0}_{r_0\times(r_2-r_0)} \\ \mathbf{0}_{(r_2-r_0)\times r_0} & \mathbf{I}_{r_2-r_0} \end{bmatrix} \frac{\hat{\mathbf{g}}_2^*}{\|(\hat{\mathbf{g}}_2^\intercal)_{[r_0+1:r_2]}\|}\mathbf{s}_{22}^\intercal + (\mathbf{w}_2^\intercal)_{[r_2+1:r_1]} \quad (3.23)$$

$$= \sqrt{\frac{\rho}{2}}\|(\hat{\mathbf{g}}_2^\intercal)_{[r_0+1:r_2]}\|\mathbf{s}_{22}^\intercal + \sqrt{\frac{\rho}{2}}\frac{(\tilde{\mathbf{g}}_2^\intercal)_{[r_0+1:r_2]}(\hat{\mathbf{g}}_2^*)_{[r_0+1:r_2]}}{\|(\hat{\mathbf{g}}_2^\intercal)_{[r_0+1:r_2]}\|}\mathbf{s}_{22}^\intercal + (\mathbf{w}_2^\intercal)_{[r_2+1:r_1]}. \quad (3.24)$$

User 2 decodes $\mathbf{s}_{22}$ and achieves the rate

$$\Delta R_2 = \frac{r_1 - r_2}{T}\mathbb{E}\left[\log\left(1 + \frac{\frac{\rho}{2}\|(\hat{\mathbf{g}}_2^{\mathsf{T}})_{[r_0+1:r_2]}\|^2}{\frac{\rho}{2}\mathbb{E}[\frac{|(\tilde{\mathbf{g}}_2^{\mathsf{T}})_{[r_0+1:r_2]}(\hat{\mathbf{g}}_2^{*})_{[r_0+1:r_2]}|^2}{\|(\hat{\mathbf{g}}_2^{\mathsf{T}})_{[r_0+1:r_2]}\|^2}] + 1}\right)\right]. \tag{3.25}$$

With the help of the feedback, the base station generates estimation for the two channels via $\hat{\mathbf{h}}_1 = [\mathbf{V}_0\ \mathbf{V}_1]\hat{\mathbf{g}}_1$, and $\hat{\mathbf{h}}_2 = [\mathbf{V}_0\ \mathbf{V}_2]\hat{\mathbf{g}}_2$. The estimation errors are $\tilde{\mathbf{h}}_1 = \mathbf{h}_1 - \hat{\mathbf{h}}_1$ and $\tilde{\mathbf{h}}_2 = \mathbf{h}_2 - \hat{\mathbf{h}}_2$. During the data phase, the transmitted signal via conjugate beamforming is

$$\mathbf{X}_{[r_1+1:T]} = \sqrt{\frac{\rho}{2}}\frac{\hat{\mathbf{h}}_1^*}{\|\hat{\mathbf{h}}_1\|}\mathbf{s}_1^{\mathsf{T}} + \sqrt{\frac{\rho}{2}}\frac{\hat{\mathbf{h}}_2^*}{\|\hat{\mathbf{h}}_2\|}\mathbf{s}_2^{\mathsf{T}}. \tag{3.26}$$

where $\mathbf{s}_k \in \mathbb{C}^{T-r_1}$, $k = 1, 2$, contains i.i.d. $\mathcal{CN}(0, 1)$ data symbols for User k. The received signals at the two users are

$$(\mathbf{y}_1^{\mathsf{T}})_{[r_1+1:T]} = \sqrt{\frac{\rho}{2}}\frac{\mathbf{h}_1^{\mathsf{T}}\hat{\mathbf{h}}_1^*}{\|\hat{\mathbf{h}}_1\|}\mathbf{s}_1^{\mathsf{T}} + \sqrt{\frac{\rho}{2}}\frac{\mathbf{h}_1^{\mathsf{T}}\hat{\mathbf{h}}_2^*}{\|\hat{\mathbf{h}}_2\|}\mathbf{s}_2^{\mathsf{T}} + (\mathbf{w}_1^{\mathsf{T}})_{[r_1+1:T]}, \tag{3.27}$$

$$(\mathbf{y}_2^{\mathsf{T}})_{[r_1+1:T]} = \sqrt{\frac{\rho}{2}}\frac{\mathbf{h}_2^{\mathsf{T}}\hat{\mathbf{h}}_2^*}{\|\hat{\mathbf{h}}_2\|}\mathbf{s}_2^{\mathsf{T}} + \sqrt{\frac{\rho}{2}}\frac{\mathbf{h}_2^{\mathsf{T}}\hat{\mathbf{h}}_1^*}{\|\hat{\mathbf{h}}_1\|}\mathbf{s}_1^{\mathsf{T}} + (\mathbf{w}_2^{\mathsf{T}})_{[r_1+1:T]}. \tag{3.28}$$

User k decodes $\mathbf{s}_k$ and achieves the rate

$$R_k = \left(1 - \frac{r_1}{T}\right)\mathbb{E}\left[\log\left(1 + \rho_k\|\hat{\mathbf{h}}_k\|^2\right)\right], \quad k = 1, 2, \tag{3.29}$$

with the equivalent SNRs $\rho_1 \triangleq \left(\mathbb{E}\left[\frac{|\tilde{\mathbf{h}}_1^{\mathsf{T}}\hat{\mathbf{h}}_1^*|^2}{\|\hat{\mathbf{h}}_1\|^2} + \frac{|\mathbf{h}_1^{\mathsf{T}}\hat{\mathbf{h}}_2^*|^2}{\|\hat{\mathbf{h}}_2\|^2}\right] + \frac{2}{\rho}\right)^{-1}$ and $\rho_2 \triangleq \left(\mathbb{E}\left[\frac{|\tilde{\mathbf{h}}_2^{\mathsf{T}}\hat{\mathbf{h}}_2^*|^2}{\|\hat{\mathbf{h}}_2\|^2} + \frac{|\mathbf{h}_2^{\mathsf{T}}\hat{\mathbf{h}}_1^*|^2}{\|\hat{\mathbf{h}}_1\|^2}\right] + \frac{2}{\rho}\right)^{-1}$.

The achievable sum rate is:

$$R = R_1 + R_2 + \Delta R_2. \tag{3.30}$$

3.2 The K-User Case with Symmetric Eigenspace

In the following subsections, we consider the K-user case. In this case, for a general (irregular) correlation structure, the signal design matching the correlations is complicated. Therefore, in order to emphasize the gain of correlation-based rate splitting and product superposition, we focus on some special configurations of the eigenspaces. The first considered special eigenspace configuration for the K-user case is the symmetric correlation structure. We first present the case when $K = 3$. Under the symmetry assumption, we have $r_{\{1\}} = r_{\{2\}} = r_{\{3\}} \triangleq p_1$, $r_{\{1,2\}} = r_{\{1,3\}} = r_{\{2,3\}} \triangleq p_2$, and $r_{\{1,2,3\}} \triangleq p_3$.

Define the matrix $\mathbf{V}$ as the collection of all the eigendirection vectors, which means

$$\mathbf{V} = \left[\mathbf{V}_{\{1\}}\ \mathbf{V}_{\{2\}}\ \mathbf{V}_{\{3\}}\ \mathbf{V}_{\{1,2\}}\ \mathbf{V}_{\{1,3\}}\ \mathbf{V}_{\{2,3\}}\ \mathbf{V}_{\{1,2,3\}}\right] \tag{3.31}$$

where $\mathbf{V}_{\mathcal{J}} \in \mathbb{C}^{M \times r_{\mathcal{J}}}$ contains the eigenvectors spanning the subspaces of all users in $\mathcal{J}$. Now we decompose the channel as $\mathbf{h}_k = [\mathbf{V}_{\mathcal{J}}]_{k \in \mathcal{J}} \mathbf{g}_k$ where $\mathbf{g}_k \in \mathbb{C}^{r_k}$. For example, $\mathbf{h}_1 = \left[\mathbf{V}_{\{1\}}\ \mathbf{V}_{\{1,2\}}\ \mathbf{V}_{\{1,3\}}\ \mathbf{V}_{\{1,2,3\}}\right]\mathbf{g}_1$.

In the first p_1 time slots, the base station sends pilots to three users simultaneously in subspaces $\mathbf{V}_{\{1\}}$,$\mathbf{V}_{\{2\}}$ and $\mathbf{V}_{\{3\}}$. The transmitted signal is

$$\mathbf{X}_{[1:p_1]} = \sqrt{\frac{\rho}{3}}(\mathbf{V}^*_{\{1\}} + \mathbf{V}^*_{\{2\}} + \mathbf{V}^*_{\{3\}}). \tag{3.32}$$

The received signal at User k is

$$(\mathbf{y}_k^\intercal)_{[1:p_1]} = \sqrt{\frac{\rho}{3}}\mathbf{h}_k^\intercal \mathbf{V}^*_{\{k\}} + (\mathbf{w}_k^\intercal)_{[1:p_1]}. \tag{3.33}$$

User k estimates $\mathbf{h}_k^\intercal \mathbf{V}^*_{\{k\}}$ to obtain $\hat{\mathbf{h}}_k^\intercal \mathbf{V}^*_{\{k\}}$ and feeds back to the base station. The estimation error is $\tilde{\mathbf{h}}_k^\intercal \mathbf{V}^*_{\{k\}} = \mathbf{h}_k^\intercal \mathbf{V}^*_{\{k\}} - \hat{\mathbf{h}}_k^\intercal \mathbf{V}^*_{\{k\}}$. In the next $3p_2$ time slots, the base station sends pilots to users i and j in the subspace of $\mathbf{V}_{\{i,j\}}(i \neq j)$ and data to the remaining user via conjugate beamforming. For example, in the first p_2 time slots, it sends

$$\mathbf{X}_{[p_1+1:p_1+p_2]} = \sqrt{\frac{\rho}{2}}\mathbf{V}^*_{\{2,3\}} + \sqrt{\frac{\rho}{2}}\mathbf{V}^*_{\{1\}} \frac{\mathbf{V}^\intercal_{\{1\}} \hat{\mathbf{h}}_1^*}{\|\mathbf{V}^\intercal_{\{1\}} \hat{\mathbf{h}}_1^*\|}\mathbf{s}_{11}^\intercal, \tag{3.34}$$

where $\mathbf{s}_{11} \in \mathbb{C}^{p_2}$ contains i.i.d. $\mathcal{CN}(0,1)$ data symbols. The received signal at User 2 or User 3 is

$$(\mathbf{y}_k^\intercal)_{[p_1+1:p_1+p_2]} = \sqrt{\frac{\rho}{2}}\mathbf{h}_k^\intercal \mathbf{V}^*_{\{2,3\}} + (\mathbf{w}_k^\intercal)_{[p_1+1:p_1+p_2]}, \quad k = 2, 3. \tag{3.35}$$

User k ($k = 2, 3$) estimates $\mathbf{h}_k^\intercal \mathbf{V}_{\{2,3\}}^*$ to obtain $\hat{\mathbf{h}}_k^\intercal \mathbf{V}_{\{2,3\}}^*$ and feeds back to the base station. The received signal at User 1 is

$$(\mathbf{y}_1^\intercal)_{[p_1+1:p_1+p_2]} = \sqrt{\frac{\rho}{2}} \mathbf{h}_1^\intercal \mathbf{V}_{\{1\}}^* \frac{\mathbf{V}_{\{1\}}^\intercal \hat{\mathbf{h}}_1^*}{\|\mathbf{V}_{\{1\}}^\intercal \hat{\mathbf{h}}_1^*\|} \mathbf{s}_{11} + (\mathbf{w}_1^\intercal)_{[p_1+1:p_1+p_2]} \tag{3.36}$$

$$= \sqrt{\frac{\rho}{2}} \|\mathbf{V}_{\{1\}}^\intercal \hat{\mathbf{h}}_1^*\| \mathbf{s}_{11} + \sqrt{\frac{\rho}{2}} \tilde{\mathbf{h}}_1^\intercal \mathbf{V}_{\{1\}}^* \frac{\mathbf{V}_{\{1\}}^\intercal \hat{\mathbf{h}}_1^*}{\|\mathbf{V}_{\{1\}}^\intercal \hat{\mathbf{h}}_1^*\|} \mathbf{s}_{11} + (\mathbf{w}_1^\intercal)_{[p_1+1:p_1+p_2]}. \tag{3.37}$$

User 1 decodes $\mathbf{s}_{11}$ and achieves the rate

$$\Delta R_1 = \frac{p_2}{T} \mathbb{E}\left[\log\left(1 + \frac{\frac{\rho}{2} \|\mathbf{V}_{\{1\}}^\intercal \hat{\mathbf{h}}_1^*\|^2}{\frac{\rho}{2} [\frac{|\tilde{\mathbf{h}}_1^\intercal \mathbf{V}_{\{1\}}^* \mathbf{V}_{\{1\}}^\intercal \hat{\mathbf{h}}_1^*|^2}{\|\mathbf{V}_{\{1\}}^\intercal \hat{\mathbf{h}}_1^*\|^2}] + 1} \right) \right], \tag{3.38}$$

In the subsequent p_2 time slots, the channel coefficients in $\mathcal{V}_{\{1,3\}}$, $\mathcal{V}_{\{1,2\}}$ are estimated and and fed back, and the achievable rate for User 2 and User 3 can be calculated similarly.

In the following p_3 time slots, the base station transmits pilots in $\mathcal{V}_{\{1,2,3\}}$ as $\mathbf{X}_{[3p_2+1:3p_2+p_3]} = \sqrt{\rho} \mathbf{V}_{\{1,2,3\}}^*$. User k receives $(\mathbf{y}_k^T)_{[3p_2+1:3p_2+p_3]} = \sqrt{\rho} \mathbf{h}_k^\intercal \mathbf{V}_{\{1,2,3\}}^* + (\mathbf{y}_k^T)_{[3p_2+1:3p_2+p_3]}$, estimates $\mathbf{h}_k^\intercal \mathbf{V}_{\{1,2,3\}}^*$ to obtain $\hat{\mathbf{h}}_k^\intercal \mathbf{V}_{\{1,2,3\}}^*$ and feeds back to the base station. From the feedbacks in the first $T_\tau = p_1 + 3p_2 + p3$ time slots, the base station obtains estimates $\hat{\mathbf{h}}_k$ of $\mathbf{h}_k$, $k = 1, 2, 3$. The estimation error is $\tilde{\mathbf{h}}_k = \mathbf{h}_k - \hat{\mathbf{h}}_k$.

During the data phase, the transmitted signal via conjugate beamforming is

$$\mathbf{X}_{[T_\tau+1:T]} = \sqrt{\frac{\rho}{3}} \frac{\hat{\mathbf{h}}_1^*}{\|\hat{\mathbf{h}}_1\|} \mathbf{s}_1^\intercal + \sqrt{\frac{\rho}{3}} \frac{\hat{\mathbf{h}}_2^*}{\|\hat{\mathbf{h}}_2\|} \mathbf{s}_2^\intercal + \sqrt{\frac{\rho}{3}} \frac{\hat{\mathbf{h}}_3^*}{\|\hat{\mathbf{h}}_3\|} \mathbf{s}_3^\intercal, \tag{3.39}$$

where $\mathbf{s}_k \in \mathbb{C}^{T-T_\tau}$ contains i.i.d. $\mathcal{CN}(0,1)$ data symbols for User k. The received signals at User 1 is

$$(\mathbf{y}_1^\intercal)_{[T_\tau+1:T]} = \sqrt{\frac{\rho}{3}} \frac{\mathbf{h}_1^\intercal \hat{\mathbf{h}}_1^*}{\|\hat{\mathbf{h}}_1\|} \mathbf{s}_1^\intercal + \sqrt{\frac{\rho}{3}} \frac{\mathbf{h}_1^\intercal \hat{\mathbf{h}}_2^*}{\|\hat{\mathbf{h}}_2\|} \mathbf{s}_2^\intercal + \sqrt{\frac{\rho}{3}} \frac{\mathbf{h}_1^\intercal \hat{\mathbf{h}}_3^*}{\|\hat{\mathbf{h}}_3\|} \mathbf{s}_3^\intercal + (\mathbf{w}_1^\intercal)_{[T_\tau+1:T]}. \tag{3.40}$$

User 1 decodes $\mathbf{s}_1$ and achieves the rate

$$R_1 = \left(1 - \frac{p_1 + 3p_2 + p_3}{T}\right) \mathbb{E}\left[\log\left(1 + \frac{\frac{\rho}{3} \|\hat{\mathbf{h}}_1\|^2}{\frac{\rho}{3} \mathbb{E}\left[\frac{|\tilde{\mathbf{h}}_1^\intercal \hat{\mathbf{h}}_1^*|^2}{\|\hat{\mathbf{h}}_1\|^2} + \frac{|\mathbf{h}_1^\intercal \hat{\mathbf{h}}_2^*|^2}{\|\hat{\mathbf{h}}_2\|^2} + \frac{|\mathbf{h}_1^\intercal \hat{\mathbf{h}}_3^*|^2}{\|\hat{\mathbf{h}}_3\|^2} \right] + 1} \right) \right]. \tag{3.41}$$

The achievable rate of User 2 and User 3 can be calculated in the same way.

The achievable sum rate is

$$R = \sum_{k=1}^{3} (R_k + \Delta R_k). \tag{3.42}$$

Now we extend this scheme to the K-user scenario. Following the signaling structure developed in the 3-user case, the transmit scheme has three phases. In the first phase, some pilot signals are transmitted. In the second phase, the remaining pilots are transmitted while at the same time, some users also receive data. In the third phase, the channel state is known (due to pilots transmitted in the earlier two phases) and the base station beamforms to all users.

The first phase has $\sum_{l=1}^{\lfloor K/2 \rfloor} \chi(G(K,l))p_l$ time slots, in the first $\chi(G(K,1))p_1 = p_1$ time slots, the base station sends $\frac{\rho}{K} \sum_{i=1}^{K} \mathbf{V}_{\{i\}}$. In the same way, during the following time slots, the base station sends pilots which will not interfere with each other. The users estimate the channel coefficients in these subspaces and feed back to the base station.

The second phase has $\sum_{l=\lfloor K/2 \rfloor+1}^{K} \chi(G(K,l))p_l$ time slots, where $\chi(G(K,l)) = \binom{K}{l}$. In this phase, the base station sends pilot in some eigendirections and simultaneously beamforms to the users which are not interfered by the pilots. For example, when sending the pilots in $\mathcal{V}_{\{1,2,\ldots,K-2\}}$, the transmitted signal is

$$\mathbf{X} = \sqrt{\frac{\rho}{3}} \mathbf{V}^*_{\{1,2,\ldots,K-2\}} + \sqrt{\frac{\rho}{3}} \mathbf{V}^*_{\{K-1\}} \frac{\mathbf{V}^\intercal_{\{K-1\}} \hat{\mathbf{h}}^*_{K-1}}{\|\mathbf{V}^\intercal_{\{K-1\}} \hat{\mathbf{h}}^*_{K-1}\|} \mathbf{s}^\intercal_{K-1} + \sqrt{\frac{\rho}{3}} \mathbf{V}^*_{\{K\}} \frac{\mathbf{V}^\intercal_{\{K\}} \hat{\mathbf{h}}^*_{K}}{\|\mathbf{V}^\intercal_{\{K\}} \hat{\mathbf{h}}^*_{K}\|} \mathbf{s}^\intercal_{K}, \tag{3.43}$$

where the equivalent channels $\mathbf{V}^\intercal_{\{K-1\}} \hat{\mathbf{h}}^*_{K-1}$ and $\mathbf{V}^\intercal_{\{K\}} \hat{\mathbf{h}}^*_{K}$ have been estimated and fed back in the first phase. During these time slots, User 1 to User $K-2$ can estimate their channel coefficients in the direction of $\mathbf{V}_{\{1,2,\ldots,K-2\}}$, while user $K-1$ can decode $\mathbf{s}_{K-1}$ and User k can decode $\mathbf{s}_K$.

In the third phase, which has $T - T_\tau(K,0)$ time slots, the base station beamforms to all users with the estimated channel by sending

$$\mathbf{x}(t) = \sqrt{\frac{\rho}{K}} \sum_{k=1}^{K} \frac{\hat{\mathbf{h}}_k}{\|\hat{\mathbf{h}}_k\|} s_k(t) \tag{3.44}$$

at time slot $t = T - T_\tau(K, 0) + 1, \ldots, T$.

Finally, the total rate that can be achieved is the sum of the rates achieved during phases two and three.

3.3 The K-User Case with On-Off Correlation

The second special correlation configuration is motivated as follows. Experience shows that small values of correlation are often inconsequential to the rate and thus can be treated as uncorrelation in signal design. Furthermore, interference-free pilot reuse is only made possible under rank deficient correlation matrices, i.e., some transmit antenna gains are fully deterministic conditioned on the others. Therefore, we consider a K-user channel where the pairs of transmit antennas are either uncorrelated or fully correlated for each user, and refer to it as *on-off correlation*. Specifically, consider the channel vector $\mathbf{h}_k = [h_{k,1}\ h_{k,2}\ \ldots\ h_{k,M}]^\intercal$ of any User k, for any $i, j \in [M]$, we assume that either $h_{k,i} = h_{k,j}$ (fully correlated) or $\mathbb{E}\left[h^*_{k,i} h_{k,j}\right] = 0$ (uncorrelated).

Consider the case where the channel coefficients of User 1 are fully correlated, the channel coefficients of User k are uncorrelated, while the remaining $K-2$ users have fully correlated channel coefficients with respect to some antennas. Let us group the antennas into $L+1$ groups: the first group has the first antenna, the l-th group has $\frac{M-1}{L}$ antennas from $\frac{(M-1)(l-1)}{L} + 2$ to $\frac{(M-1)l}{L} + 1$. We assign the users to each group as follows: User k is assigned to group l if the channel coefficients of User k corresponding to the antennas in group l are fully correlated, i.e $h_{k, \frac{(M-1)(l-1)}{L}+1+i} = h_{k, \frac{(M-1)(l-1)}{L}+1+j}$, for $1 \le i, j \le \frac{M-1}{L}$. Because User 1 has fully correlated channel coefficients, it is assigned to every group.

The base station transmits the following signal in the pilot phase:

$$\mathbf{X}_{[1:M]} = \sqrt{\rho}\ \mathrm{diag}\left(v_0, v_1, v_1\mathbf{u}_1^\intercal, v_2, v_2\mathbf{u}_2^\intercal, \ldots, v_L, v_L\mathbf{u}_L^\intercal\right), \tag{3.45}$$

where $v_0 = 1$ and $v_l \in \mathbb{C}$, $\mathbf{u}_l \in \mathbb{C}^{\frac{M-1}{L}-1}$, $l = 1, 2, ..., L$ are mutually independent random variables following the distribution $\mathcal{CN}(0, 1)$. Here $\{v_l\}_{l=1}^L$ are the symbols for User 1 and $\mathbf{u}_l$

is for one of the users in group l. The received signal at User k is:

$$(\mathbf{y}_k^\mathsf{T})_{[1:M]} = \sqrt{\rho}\mathbf{h}_k^\mathsf{T}\mathbf{X} + (\mathbf{w}_k^\mathsf{T})_{[1:M]}. \tag{3.46}$$

User k estimates $\mathbf{X}^\mathsf{T}\mathbf{h}_k$ via MMSE and feeds back the estimated version $\frac{\sqrt{\rho}}{\rho+1}(\mathbf{y}_k)_{[1:M]}$ to the base station. Because the base station knows $\mathbf{X}$, it obtains an estimated version of the channel of User k as $\hat{\mathbf{h}}_k = \frac{\sqrt{\rho}}{\rho+1}\mathbf{X}^{-\mathsf{T}}(\mathbf{y}_k)_{[1:M]}$. The estimation error is $\tilde{\mathbf{h}}_k = \mathbf{h}_k - \hat{\mathbf{h}}_k$.

Denote the fully correlated channel coefficient of User 1 as $\bar{h}_1 \triangleq h_{1,1} = h_{1,2} = \cdots = h_{1,M}$. In the first time slot, User 1 receives $y_1(1) = \sqrt{\rho}\bar{h}_1 + w_1(1)$. It estimates $\bar{h}_1$ by $\hat{\bar{h}}_1 = \frac{\sqrt{\rho}}{\rho+1}y_1(1)$ and the estimation error is $\tilde{\bar{h}}_1 = \bar{h}_1 - \hat{\bar{h}}_1$. We have that $\hat{\bar{h}}_1 \sim \mathcal{CN}\big(0, \frac{\rho}{\rho+1}\big)$ and $\tilde{\bar{h}}_1 \sim \mathcal{CN}\big(0, \frac{1}{\rho+1}\big)$. In the time slots $\frac{(M-1)(l-1)}{L}+2$, $l = 1, \dots, L$, User 1 receive $y_1(\frac{(M-1)(l-1)}{L} + 2) = \sqrt{\rho}\bar{h}_1 v_l + w_1(\frac{(M-1)(l-1)}{L} + 2)$. User 1 can decode $\{v_l\}$ and achieves the rate

$$\begin{aligned}\Delta R_1 &= \frac{L}{T}\mathbb{E}\left[\log\left(1 + \frac{\rho|\hat{\bar{h}}_1|^2}{\rho\mathbb{E}\left[|\tilde{\bar{h}}_1|^2\right] + 1}\right)\right] && (3.47)\\ &= \frac{L}{T}\mathbb{E}\left[\log\left(1 + \frac{\rho(\rho+1)}{2\rho+1}|\hat{\bar{h}}_1|^2\right)\right] && (3.48)\\ &= \frac{L}{T}\log(e)\exp\Big(\frac{2\rho+1}{\rho^2}\Big)E_1\Big(\frac{2\rho+1}{\rho^2}\Big), && (3.49)\end{aligned}$$

where $E_1(x) \triangleq \int_x^\infty \frac{e^{-t}}{t}\mathrm{d}t$ is the exponential integral function.

In addition, if User k is assigned to group $l+1$, $l = 1, \dots, L$, denote $h_{k,\frac{(M-1)l}{L}+2}v_{l+1} = h_{k,\frac{(M-1)l}{L}+3}v_{l+1} = \cdots = h_{k,\frac{(M-1)(l+1)}{L}+1}v_{l+1} \triangleq \bar{h}_{k,l+1}$. In time slot $\frac{(M-1)l}{L}+2$, the received signal of User k is

$$y_k\big(\tfrac{(M-1)l}{L} + 2\big) = \sqrt{\rho}\bar{h}_{k,l+1} + w_k\big(\tfrac{(M-1)l}{L} + 2\big). \tag{3.50}$$

User k can estimate the equivalent channel $\bar{h}_{k,l+1}$ by $\hat{\bar{h}}_{k,l+1} = \frac{\sqrt{\rho}}{\rho+1}y_k\big(\frac{(M-1)l}{L}+2\big)$ and the estimation error is $\tilde{\bar{h}}_{k,l+1} = \bar{h}_{k,l+1} - \hat{\bar{h}}_{k,l+1}$. We have that $\mathbb{E}\left[|\hat{\bar{h}}_{k,l+1}|^2\right] = \frac{\rho}{\rho+1}$ and $\mathbb{E}\left[|\hat{\bar{h}}_{k,l+1}|^2\right] = \frac{1}{\rho+1}$. In the next $\frac{M-1}{L} - 1$ time slots, User k receives

$$y_k\big(\tfrac{(M-1)l}{L} + 2 + t\big) = \sqrt{\rho}\bar{h}_{k,l+1}u_{l+1,t} + w_k\big(\tfrac{(M-1)l}{L} + 2 + t\big) \tag{3.51}$$

for $t = 1, 2, \ldots, \frac{M-1}{L} - 1$. Therefore, User k can decode $\mathbf{u}_{l+1}$ and achieve the rate

$$\Delta R_{l+1} = \frac{(M-1)/L - 1}{T} \mathbb{E}\left[\log\left(1 + \frac{\rho\left|\hat{\bar{h}}_{k,l+1}\right|^2}{\rho\mathbb{E}\left[\left|\tilde{\bar{h}}_{k,l+1}\right|^2\right] + 1}\right)\right] \tag{3.52}$$

$$= \frac{(M-1)/L - 1}{T} \mathbb{E}\left[\log\left(1 + \frac{\rho(\rho+1)}{2\rho+1}\left|\hat{\bar{h}}_{k,l+1}\right|^2\right)\right]. \tag{3.53}$$

In the beamforming phase, the base station beamforms to the users according to the estimated channel with equal power. The transmitted signal is

$$\mathbf{X}_{[M+1:T]} = \sqrt{\frac{\rho}{K}} \sum_{k=1}^{K} \frac{\hat{\mathbf{h}}_k^*}{\|\hat{\mathbf{h}}_k\|} \mathbf{s}_k^{\mathsf{T}}, \tag{3.54}$$

where $\mathbf{s}_k \in \mathbb{C}^{T-M}$ contains i.i.d. $\mathcal{CN}(0,1)$ data symbols for User k. The received signal at User k is:

$$(\mathbf{y}_k^{\mathsf{T}})_{[M+1:T]} = \mathbf{h}_k^{\mathsf{T}} \mathbf{X}_{[M+1:T]} + (\mathbf{w}_k^{\mathsf{T}})_{[M+1:T]} \tag{3.55}$$

$$= \sqrt{\frac{\rho}{K}} \sum_{l=1}^{K} \frac{\mathbf{h}_k^{\mathsf{T}} \hat{\mathbf{h}}_l}{\|\hat{\mathbf{h}}_l\|} \mathbf{s}_l^{\mathsf{T}} + (\mathbf{w}_k^{\mathsf{T}})_{[M+1:T]}. \tag{3.56}$$

User k decodes $\mathbf{s}_k$ and achieves the rate

$$R_i = \left(1 - \frac{M}{T}\right) \mathbb{E}\left[\log\left(1 + \frac{\frac{\rho}{K}\|\hat{\mathbf{h}}_k\|^2}{\frac{\rho}{K}\mathbb{E}\left[\frac{|\tilde{\mathbf{h}}_k^{\mathsf{T}} \hat{\mathbf{h}}_k^*|^2}{\|\hat{\mathbf{h}}_k\|^2} + \sum_{l \neq k} \frac{|\mathbf{h}_k^{\mathsf{T}} \hat{\mathbf{h}}_l^*|^2}{\|\hat{\mathbf{h}}_l\|^2}\right] + 1}\right)\right]. \tag{3.57}$$

Finally, the achievable sum rate is:

$$R = \sum_{k=1}^{K} R_i + \sum_{l=1}^{L+1} \Delta R_l. \tag{3.58}$$

Figure 3.2 shows the performance gain of the proposed scheme with respect to the conventional one under the following configuration: $K = 10$, $L = 9$, $M = 64$, $T = 128$.

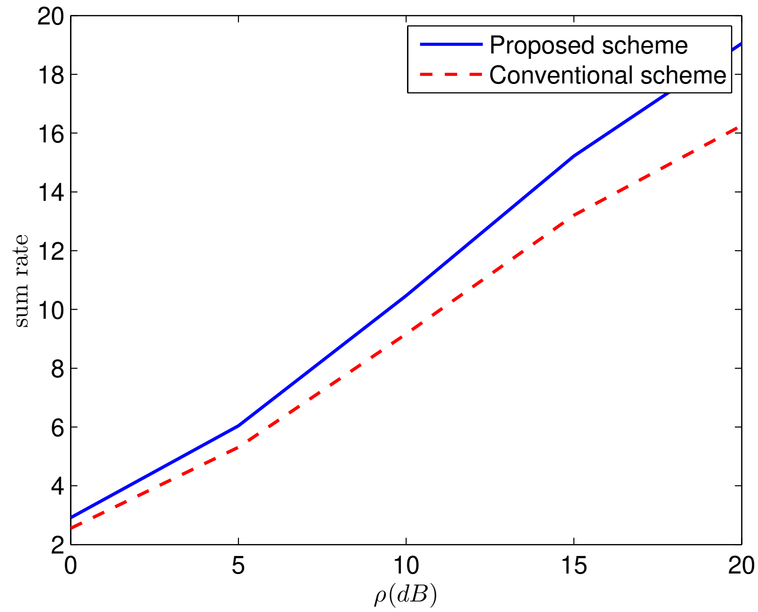

Figure 3.2. Sum rate of an FDD massive MIMO system in on-off correlated fading where $K = 10, M = 64$, $T = 128$.

CHAPTER 4

COHERENCE DIVERSITY IN THE MIMO RELAY CHANNEL[1] [2]

In this chapter, we begin by proving the theorem: under identical coherence intervals for the source-relay, relay-destination, and source-destination links, the relay cannot provide any DoF gains compared with the direct link alone. This is a simple but important negative result that is independent of antenna configurations at the three nodes and is used as a reference. When the coherence intervals are unequal, we start with a representative example, design signaling appropriately for the unequal coherence intervals, and show the resulting DoF gains. Then we broaden the result by removing the constraints from the length and alignment of the coherence blocks, showing that the DoF gains persist in the more general case. Furthermore, a new scheme combining product superposition and relay scheduling is proposed, motivated by the following observation: Whenever a pilot-based relay is activated, the relay pilots impose a cost (in degrees of freedom) due to their interference with source-destination transmission. In the new scheme, this cost is compared to the relay gains, and the relay is activated accordingly. We show the extent to which this new scheme improves the degrees of freedom of the relay channel. This paper also studies multiple parallel relays under non-identical coherence intervals, wherein transmission strategies are studied and achievable degrees of freedom are calculated.

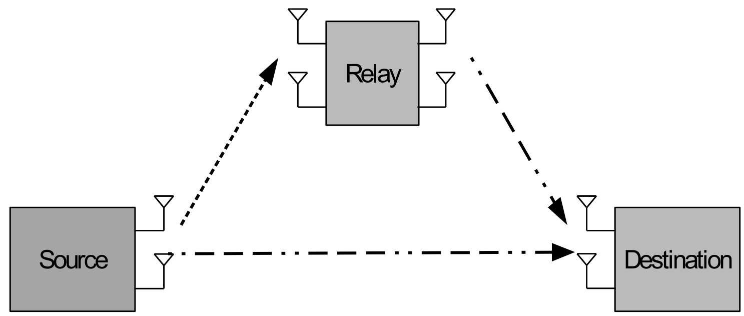

Figure 4.1. Relay channel with coherence diversity

4.1 System Model

Consider a MIMO relay in full-duplex mode as in Figure 4.1. The source and destination are equipped with N_S and N_D antennas, respectively. The relay has N_R receive antennas and n_R transmit antennas. The number of active (powered) relay transmit antennas in a transmission scheme is represented with n_r, which is optimized in each scenario. Obviously $n_r \leq n_R$. The received signals at the relay and destination are:

$$\mathbf{y}_R = \mathbf{H}_{SR}\mathbf{x}_S + \mathbf{w}_R \tag{4.1}$$

$$\mathbf{y}_D = \mathbf{H}_{SD}\mathbf{x}_S + \mathbf{H}_{RD}\mathbf{x}_R + \mathbf{w}_D, \tag{4.2}$$

where $\mathbf{x}_S$ and $\mathbf{x}_R$ are signals transmitted from the source and relay. $\mathbf{w}_R$ and $\mathbf{w}_D$ are i.i.d. white Gaussian noise and $\mathbf{H}_{SR}$, $\mathbf{H}_{RD}$ and $\mathbf{H}_{SD}$ are channel gain matrices whose entries are i.i.d. Gaussian. Channel gain entries and noise components are zero-mean and have unit variance. Channel gains experience block fading, remaining constant during the coherence intervals which are, respectively, of length T_{SR}, T_{RD} and T_{SD}, satisfying $T_{SR} \geq 2\max(N_S, N_R)$, $T_{RD} \geq 2\max(N_R, N_D)$ and $T_{SD} \geq 2\max(N_S, N_D)$. Channel gains are independent across blocks [58]. The source and relay obey the power constraints $\mathbb{E}[\mathrm{tr}(\mathbf{x}_S\mathbf{x}_S')] \leq \rho$ and $\mathbb{E}[\mathrm{tr}(\mathbf{x}_R\mathbf{x}_R')] \leq \rho$. We assume there is no CSIT at the source or relay.

The source sends messages to the destination with rate $R(\rho)$ at signal-to-noise ratio ρ. The degrees of freedom at the destination achieving rate $R(\rho)$ are defined as

$$d = \lim_{\rho \to \infty} \frac{R(\rho)}{\log(\rho)}. \tag{4.3}$$

4.2 Aligned Coherence Blocks

We first show that the relay cannot provide any gains in degrees of freedom under *identical* coherence intervals. Then we analyze the scenarios where the coherence times are unequal.

4.2.1 Identical Coherence Times

Proposition 3. *When relay link coherence times are identical ($T_{SD} = T_{SR} = T_{RD} = T$), the relay does not improve the degrees of freedom of the source-destination link, namely:*

$$d = \min(N_S, N_D)(1 - \frac{\min(N_S, N_D)}{T}). \tag{4.4}$$

Proof. From the cut-set bound,

$$R \leq \min\{\mathrm{I}(\mathbf{x}_S; \mathbf{y}_R, \mathbf{y}_S | \mathbf{x}_R), \mathrm{I}(\mathbf{x}_S, \mathbf{x}_R; \mathbf{y}_D)\}\ . \tag{4.5}$$

If $N_S \leq N_D$, consider the broadcast component of the cutset bound: $R \leq \mathrm{I}(\mathbf{x}_S; \mathbf{y}_R, \mathbf{y}_S | \mathbf{x}_R)$. Because $T_{SD} = T_{SR} = T$ and there is no CSIT, the right hand side in the inequality is upper bounded by the capacity of a point-to-point channel having N_S transmit antennas and $(N_D + N_R)$ receive antennas with coherence time T, which is $N_S(1 - \frac{N_S}{T}) \log \rho + o(\log \rho)$. Then we have

$$d \leq N_S(1 - \frac{N_S}{T}), \tag{4.6}$$

which can be achieved by the direct link alone.

If $N_S \geq N_D$, we focus on the MAC component of the cutset bound: $R \leq \mathrm{I}(\mathbf{x}_S, \mathbf{x}_R; \mathbf{y}_D)$. Since $T_{SD} = T_{SR} = T$, the right hand side is upper bounded by the capacity of a point-to-point channel having $(N_D + N_R)$ transmit antennas and N_D receive antennas with coherence time T, whose capacity is $N_D(1 - \frac{N_D}{T}) \log \rho + o(\log \rho)$. Then we have

$$d \leq N_D(1 - \frac{N_D}{T}), \tag{4.7}$$

and this degrees of freedom can also be achieved by the direct link alone. This completes the proof. □

4.2.2 A Representative Example for Unequal Coherence Times

To pave the way for the analysis to come, and to motivate the direction taken by this paper, we provide an example whose purpose is to illuminate the main features of the problem in a simple setting. In this example, the source and relay are equipped with two antennas and the destination is equipped with three antennas. The coherence times of the three links are as follows: $T_{SD} = T_{RD} = 8$ and $T_{SR} = \infty$, i.e., the source-relay channel is static, therefore the cost of training over this link is amortized over a large number of samples, so we can assume the relay knows $\mathbf{H}_{SR}$.

The source uses product superposition, sending

$$\mathbf{X}_S = \mathbf{U}[\mathbf{I}_2, \mathbf{0}_{2\times 1}, \mathbf{V}], \tag{4.8}$$

where $\mathbf{U} \in \mathbb{C}^{2\times 2}$ and $\mathbf{V} \in \mathbb{C}^{2\times 5}$.

At the relay, the received signal is

$$\mathbf{Y}_R = \mathbf{H}_{SR}\mathbf{X}_S + \mathbf{W}_R = \mathbf{H}_{SR}\mathbf{U}[\mathbf{I}_2, \mathbf{0}_{2\times 1}, \mathbf{V}] + \mathbf{W}_R. \tag{4.9}$$

The received signal during the first two time slots is

$$\mathbf{Y}'_R = \mathbf{H}_{SR}\mathbf{U} + \mathbf{W}'_R. \tag{4.10}$$

The relay knows $\mathbf{H}_{SR}$ and decodes $\mathbf{U}$. The signal decoded by the relay in the previous block is $\mathbf{U}'$ and the two rows of $\mathbf{U}'$ are $\mathbf{u}'_1, \mathbf{u}'_2 \in \mathbb{C}^{1\times 2}$.

The relay powers only one antenna for transmission and sends

$$\mathbf{X}_R = [\mathbf{0}_{1\times 2}, 1, \mathbf{u}'_1, \mathbf{u}'_2, 0] \in \mathbb{C}^{1\times 8}. \tag{4.11}$$

The received signal at the destination is:

$$\begin{aligned}
\mathbf{Y}_D &= \mathbf{H}_{SD}\mathbf{X}_S + \mathbf{H}_{RD}\mathbf{X}_R + \mathbf{W}_D \\
&= [\mathbf{H}_{SD}, \mathbf{H}_{RD}] \begin{bmatrix} \mathbf{U}[\mathbf{I}_2, \mathbf{0}_{2\times 1}, \mathbf{V}] \\ \mathbf{0}_{1\times 2}, 1, \mathbf{u}', \mathbf{u}', 0 \end{bmatrix} + \mathbf{W}_D \\
&= [\mathbf{H}_{SD}\mathbf{U}, \mathbf{H}_{RD}] \Big[\mathbf{I}_3, \begin{bmatrix} \mathbf{V} \\ \mathbf{u}'_1, \mathbf{u}'_2, 0 \end{bmatrix} \Big] + \mathbf{W}_D,
\end{aligned} \tag{4.12}$$

The destination estimates the equivalent channel $[\mathbf{H}_{SD}\mathbf{U}, \mathbf{H}_{RD}]$ in the first three time slots and decodes $\mathbf{V}, \mathbf{u}'_1$ and $\mathbf{u}'_2$. In this proposed scheme, the destination can achieve the degrees of freedom $(2 \times 5 + 2 \times 1 \times 2)/8 = 1.75$. In comparison, a traditional relaying scheme assigns pilots and training according to the smallest coherence time and achieves the degrees of freedom $2 \times (8 - 2)/8 = 1.5$.

4.2.3 Coherence Conditions $T_{SR} = \infty$

When $T_{SR} = \infty$, the training resources required for the source-relay link can be amortized over a long period and are therefore negligible. This scenario occurs when source and relay are either stationary, have a dominant line-of-sight component, or both.

Theorem 9. *In a relay channel with $T_{SR} = \infty$ and $N_S < N_D$, the following degrees of freedom are achievable, where $N_S^* \triangleq \min\{N_S, N_R\}$:*
If $T_{SD} = T_{RD}$,

$$d = (1 - \frac{N_S + n_r}{T_{SD}}) \max_{n_r} \min\{N_S + n_r, N_S + \frac{N_S^* N_S}{T_{SD} - n_r - N_S}\}. \tag{4.13}$$

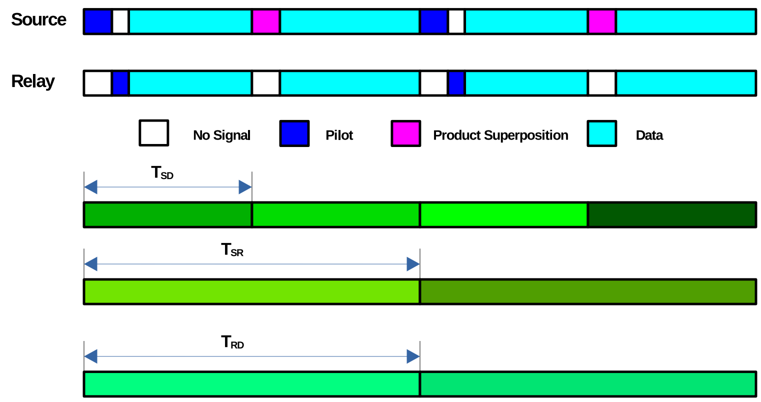

Figure 4.2. Signaling structure of product superposition

If $T_{RD} = KT_{SD}$

$$d = (1 - \frac{N_S + n_r}{T_{SD}}) \max_{n_r} \min \left\{ (N_S + n_r)(1 + \frac{(K-1)n_r}{K(T_{SD} - n_r - N_S)}), \right. \\ \left. N_S(1 + \frac{N_S^* + (K-1)n_r}{T_{SD} - n_r - N_S}) \right\}. \quad (4.14)$$

If $T_{SD} = KT_{RD}$

$$d = (1 - \frac{n_r}{T_{RD}}) \max_{n_r} \min \left\{ (N_S + n_r)(1 + \frac{N_S}{K(T_{RD} - n_r)}), N_S(1 + \frac{KN_S^* - N_S}{K(T_{RD} - n_r)}) \right\}. \quad (4.15)$$

Proof. When $T_{SD} = T_{RD}$, the proposed signaling structure is presented in Figure 4.2. In this case, the source sends the product superposition signal:

$$\mathbf{X}_S = \mathbf{U}[\mathbf{I}_{N_S}, \mathbf{0}_{N_S \times n_r}, \mathbf{V}_S], \quad (4.16)$$

where $n_r \leq \min\{N_S, N_D - N_S\}$,$\mathbf{U} \in \mathbb{C}^{N_S \times N_S}$ and $\mathbf{V}_S \in \mathbb{C}^{N_S \times (T_{SD} - n_r - N_S)}$.

At the relay, the received signal is

$$\mathbf{Y}_R = \mathbf{H}_{SR}\mathbf{X}_S + \mathbf{W}_R = \mathbf{H}_{SR}\mathbf{U}[\mathbf{I}_{N_S}, \mathbf{0}_{N_S \times n_r}, \mathbf{V}_S] + \mathbf{W}_R. \quad (4.17)$$

The received signal during the first N_S time slots is

$$\mathbf{Y}'_R = \mathbf{H}_{SR}\mathbf{U} + \mathbf{W}'_R. \quad (4.18)$$

The relay knows $\mathbf{H}_{SR}$ and decodes $\mathbf{U}$. Assume the message decoded by the relay in the previous block is $\mathbf{U}'$. The relay uses n_r transmit antennas, sending

$$\mathbf{X}_R = [\mathbf{0}_{n_r \times N_S}, \mathbf{I}_{n_r}, \mathbf{V}_S] \in \mathbb{C}^{n_r \times T_{SD}}, \tag{4.19}$$

where $\mathbf{V}_S \in \mathbb{C}^{n_r \times (T_{SD} - n_r - N_S)}$.

The received signal at the destination is

$$\begin{aligned}
\mathbf{Y}_D &= \mathbf{H}_{SD}\mathbf{X}_S + \mathbf{H}_{RD}\mathbf{X}_R + \mathbf{W}_D \\
&= [\mathbf{H}_{SD}, \mathbf{H}_{RD}] \begin{bmatrix} \mathbf{U}[\mathbf{I}_{N_S}, \mathbf{0}_{N_S \times n_r}, \mathbf{V}_S] \\ \mathbf{0}_{n_r \times N_S}, \mathbf{I}_{n_r}, \mathbf{V}_R \end{bmatrix} + \mathbf{W}_D \\
&= [\mathbf{H}_{SD}\mathbf{U}, \mathbf{H}_{RD}][\mathbf{I}_{(N_S + n_r)}, \begin{bmatrix} \mathbf{V}_S \\ \mathbf{V}_R \end{bmatrix} + \mathbf{W}_D.
\end{aligned} \tag{4.20}$$

The destination estimates the equivalent channel $[\mathbf{H}_{SD}\mathbf{U}.\mathbf{H}_{RD}]$ during the first $(N_S + n_r)$ time slots and then decodes $\mathbf{V}_S$ and $\mathbf{V}_R$. At the destination, the decoded messages have two parts: $\mathbf{V}_S$ from the source and $\mathbf{V}_R$ from the relay, which provide degrees of freedom $N_S(T_{SD} - n_r - N_S)$ and $n_r(T_{SD} - n_r - N_S)$. The message in $\mathbf{V}_R$ is from $\mathbf{U}'$. The degrees of freedom the relay can decode from $\mathbf{U}'$ are $N_S^* N_S$. The rate of the message emitted by the relay is bounded by the rate it decodes from the source. Adding up the degrees of freedom provided by the source and the relay, and optimizing the number of transmit antennas at the relay, the end-to-end degrees of freedom are (4.13).

When $T_{RD} = KT_{SD}$, our scheme has a transmission block from the source that has length KT_{SD}, which we divide into sub-blocks of length T_{SD}. During the first sub-block, the source sends the signal

$$\mathbf{X}_S^1 = \mathbf{U}^1[\mathbf{I}_{N_S}, \mathbf{0}_{N_S \times n_r}, \mathbf{V}_S^1], \tag{4.21}$$

where $n_r \leq \min\{N_S, N_D - N_S\}$,$\mathbf{U}^1 \in \mathbb{C}^{N_S \times N_S}$ and $\mathbf{V}_S^1 \in \mathbb{C}^{N_S \times (T_{SD} - n_r - N_S)}$.

The relay decodes $\mathbf{U}^1$ and uses n_r transmit antennas and sends

$$\mathbf{X}_R^1 = [\mathbf{0}_{n_r\times N_S}\ \mathbf{I}_{n_r}\ \mathbf{V}_R^1] \in \mathbb{C}^{n_r\times T_{SD}}, \tag{4.22}$$

where $\mathbf{V}_R^1 \in \mathbb{C}^{n_r\times(T_{SD}-n_r-N_S)}$. The received signal at the destination is

$$\mathbf{Y}_D^1 = \mathbf{H}_{SD}^1\mathbf{X}_S^1 + \mathbf{H}_{RD}\mathbf{X}_R^1 + \mathbf{W}_D^1 = [\mathbf{H}_{SD}^1\mathbf{U}^1, \mathbf{H}_{RD}]\Big[\mathbf{I}_{(N_S+n_r)}, \begin{bmatrix}\mathbf{V}_S^1\\ \mathbf{V}_R^1\end{bmatrix}\Big] + \mathbf{W}_D^1. \tag{4.23}$$

In the first sub-block, the three signal components $\mathbf{V}_S^1$, $\mathbf{V}_R$ and $\mathbf{U}^1$ respectively provide for the degrees of freedom $N_S(T-n_r-N_S)$, $n_r(T-n_r-N_S)$ and $N_S^*N_S$.

During the following $(K-1)$ sub-blocks, the source sends the signal

$$\mathbf{X}_S^k = \mathbf{U}^k[\mathbf{I}_{N_S}, \mathbf{V}_S^k],\ 2\le k\le K, \tag{4.24}$$

where $\mathbf{V}_S^k \in \mathbb{C}^{N_S\times(T_{SD}-n_r-N_S)}$.

The relay uses n_r transmit antennas and sends:

$$\mathbf{X}_R^k = [\mathbf{0}_{n_r\times N_S}, \mathbf{V}_R^k] \in \mathbb{C}^{n_r\times T_{SD}}, \tag{4.25}$$

where $\mathbf{V}_R^k \in \mathbb{C}^{n_r\times(T_{SD}-N_S)}$. The received signal at the destination is

$$\mathbf{Y}_D^k = \mathbf{H}_{SD}^k\mathbf{X}_S^k + \mathbf{H}_{RD}\mathbf{X}_R^k + \mathbf{W}_D^k = [\mathbf{H}_{SD}^k\mathbf{U}^k, \mathbf{H}_{RD}]\Big[\mathbf{I}_{N_S}, \begin{bmatrix}\mathbf{V}_S^k\\ \mathbf{V}_R^k\end{bmatrix}\Big] + \mathbf{W}_D^k. \tag{4.26}$$

During Sub-block k, the destination can decode $\mathbf{V}_S^k$, $\mathbf{V}_R^k$ and $\mathbf{U}^k$, which respectively provide degrees of freedom $N_S(T-N_S)$,$n_r(T-N_S)$ and $N_S^*N_S$. Therefore, the end-to-end degrees of freedom are (4.14).

When $T_{SD} = KT_{RD}$, our source transmission block has length KT_{RD}, with sub-blocks of length T_{RD}. For the first sub-block, the source uses product superposition, sending

$$\mathbf{X}_S^1 = \mathbf{U}[\mathbf{I}_{N_S}, \mathbf{0}_{N_S\times n_r}, \mathbf{V}_S^1] \in \mathbb{C}^{N_S\times T_{RD}}. \tag{4.27}$$

In the remaining $K-1$ sub-blocks with length T_{RD}, the source sends

$$\mathbf{X}_S^k = \mathbf{U}[\mathbf{0}_{N_S \times n_r}, \mathbf{V}_S^k] \in \mathbb{C}^{N_S \times T_{RD}}, \tag{4.28}$$

where $n_r \leq \min\{N_S, N_D - N_S\}$, $\mathbf{U} \in \mathbb{C}^{N_S \times N_S}$, $\mathbf{V}_S^1 \in \mathbb{C}^{N_S \times (T_{RD}-n_r-N_s)}$ and $\mathbf{V}_S^k \in \mathbb{C}^{N_S \times (T_{RD}-n_r)}$, $k = 2, 3, \ldots, K$.

The received signal during the first N_S time slots is

$$\mathbf{Y}'_R = \mathbf{H}_{SR}\mathbf{U} + \mathbf{W}'_R. \tag{4.29}$$

The relay knows $\mathbf{H}_{SR}$ and decodes $\mathbf{U}$. Then it uses n_r transmit antennas and sends

$$\mathbf{X}_R^1 = [\mathbf{0}_{n_r \times N_S}, \mathbf{I}_{n_r}, \mathbf{V}_R^1] \in \mathbb{C}^{n_r \times T_{RD}}, \tag{4.30}$$

during the first sub-block with length T_{RD}. In the remaining $K-1$ sub-block the relay sends

$$\mathbf{X}_R^k = [\mathbf{I}_{n_r}, \mathbf{V}_R^k] \in \mathbb{C}^{n_r \times T_{RD}}. \tag{4.31}$$

During the first sub-block, the received signal at the destination is

$$\begin{aligned}
\mathbf{Y}_D^1 &= \mathbf{H}_{SD}\mathbf{X}_S^1 + \mathbf{H}_{RD}^1\mathbf{X}_R^1 + \mathbf{W}_D^1 \\
&= [\mathbf{H}_{SD}, \mathbf{H}_{RD}^1] \begin{bmatrix} \mathbf{U}[\mathbf{I}_{N_S}, \mathbf{0}_{N_S \times n_r}, \mathbf{V}_S^1] \\ \mathbf{0}_{n_r \times N_S}, \mathbf{I}_{n_r}, \mathbf{V}_R^1 \end{bmatrix} + \mathbf{W}_D^1 \\
&= [\mathbf{H}_{SD}\mathbf{U}, \mathbf{H}_{RD}^1] \left[\mathbf{I}_{(N_S+n_r)}, \begin{bmatrix} \mathbf{V}_S^1 \\ \mathbf{X}_D^1 \end{bmatrix} \right] + \mathbf{W}_D^1.
\end{aligned} \tag{4.32}$$

The destination estimates the equivalent channel $[\mathbf{H}_{SD}\mathbf{U}, \mathbf{H}_{RD}^1]$ during the first (N_S+n_r) time slots and decodes $\mathbf{V}_S^1$.

During Sub-block k, the received signal at the destination is

$$\begin{aligned}\mathbf{Y}_D^k &= \mathbf{H}_{SD}\mathbf{X}_S^k + \mathbf{H}_{RD}^k\mathbf{X}_R^k + \mathbf{W}_D^k \\ &= [\mathbf{H}_{SD}, \mathbf{H}_{RD}^k]\begin{bmatrix}\mathbf{U}[\mathbf{0}_{N_S\times n_r}, \mathbf{V}_S^k] \\ \mathbf{I}_{n_r}, \mathbf{V}_R^k\end{bmatrix} + \mathbf{W}_D^k \\ &= \left[\mathbf{H}_{RD}^k, [\mathbf{H}_{SD}\mathbf{U}, \mathbf{H}_{RD}^k]\begin{bmatrix}\mathbf{V}_S^k \\ \mathbf{V}_R^k\end{bmatrix}\right] + \mathbf{W}_D^k.\end{aligned} \tag{4.33}$$

The destination estimates $\mathbf{H}_{RD}^k$ during the first n_r time slots. Because the destination is already estimated $\mathbf{H}_{SD}\mathbf{U}$, it knows the equivalent channel $[\mathbf{H}_{SD}\mathbf{U},\ \mathbf{H}_{RD}^k]$. During the remaining time slots, the destination decodes $\mathbf{V}_S^k$, $\mathbf{V}_R^k$, which respectively provide degrees of freedom $N_S(T - n_r - N_s)$ and $N_S(T - n_r)$. $\mathbf{V}_R^1$ provides degrees of freedom $n_r(T - n_r - N_s)$ and $\mathbf{V}_R^k(2 \le k \le K)$ provides degrees of freedom $n_r(T - n_r)$. Adding up the degrees of freedom and optimizing the number of transmit antennas at the relay produces (4.15). This completes the proof. □

Corollary 3. *The degrees of freedom in Theorem 9 are optimal under channel conditions* $T_{SD} = T_{SR}$ *and antenna configuration:*

$$(N_D^* - N_S)(T_{SD} - N_D^*) \le N_S^* N_S \tag{4.34}$$

where $N_D^* \triangleq \min\{N_S + n_R, N_D\}$. *In this case, the DoF is:*

$$d_{opt} = N_D^*(1 - \frac{N_D^*}{T_{SD}}). \tag{4.35}$$

Proof. For achievability, the relay activates $n_r = N_D^* - N_S$ antennas for transmission. Because the condition (4.34) holds (equivalent to $d_2 \le d_3$), according to Theorem 9, the degrees of freedom

$$\frac{1}{T_{SD}}\{N_S(T_{SD} - n_r - N_S) + n_r(T_{SD} - n_r - N_S)\} = N_D^*(1 - \frac{N_D^*}{T_{SD}})$$

are achievable. For the converse, from the cut-set bound, the capacity of the relay is upper bounded by $I(\mathbf{Y}_D; \mathbf{X}_R, \mathbf{X}_S)$. Because the coherence times of the source-destination and relay-destination links are identical and the coherence blocks are aligned, this mutual information is equivalent to the capacity of a point-to-point channel with $N_S + n_R$ transmit antennas and N_D receive antennas with coherence time T_{SD}. The degrees of freedom upper bound for this point-to-point channel is $N_D^*(1 - \frac{N_D^*}{T_{SD}})$. This completes the proof. □

Corollary 4. *When $T_{SD} = T_{RD}$ and $T_{SR} = \infty$, the relay degrees of freedom are strictly greater than the degrees of freedom of source-destination link alone.*

Proof. From Theorem 9, the direct link alone can achieve the following degrees of freedom: $d' = \frac{N_S}{T_{SD}} \times (T_{SD} - N_S)$. Choose $n_r = 1$. If $d_3 \leq d_2$, the degrees of freedom achieved by the proposed scheme are

$$d \geq \frac{1}{T_{SD}}(N_S(T_{SD} - 1 - N_S) + N_S^2) = \frac{N_S}{T_{SD}}(T_{SD} - N_S + N_S^* - 1). \tag{4.36}$$

Obviously, $d \geq d'$; if $d_2 \leq d_3$, the degrees of freedom achieved are

$$d \geq \frac{1}{T_{SD}}(N_S(T_{SD} - 1 - N_S) + (T_{SD} - 1 - N_S)) = \frac{N_S + 1}{T_{SD}}(T_{SD} - 1 - N_S). \tag{4.37}$$

Because $T_{SD} \geq 2N_D \geq 2N_S + 2$,

$$d \geq = \frac{N_S + 1}{T_{SD}}(T_{SD} - 1 - N_S) > \frac{N_S}{T_{SD}}(T_{SD} - N_S) = d'. \tag{4.38}$$

This completes the proof. □

Remark 3. *Theorem 9 highlights strictly positive gains, e.g., $N_S = N_R = 3, N_D = 5, T = 6, K = 2$, $d = \frac{7}{3}, d' = \frac{9}{4}$. But under some conditions, e.g. when the relay-destination coherence time is too short, the relay pilot requirements will eat into the gains. For example, when $N_S = N_R = 3, N_D = 5, T = 4, K = 3$, $d = 2, d' = \frac{9}{4}$, the relay does not provide any DoF gains.*

4.2.4 Coherence Conditions $T_{SR} < \infty$

When T_{SR} is bounded, one can no longer assume that the relay knows $\mathbf{H}_{SR}$ with negligible training cost. The following theorem states the achievable degrees of freedom.

Theorem 10. *In a relay channel with link coherence times $T_{SR} = KT_{SD}$, and antenna configuration $N_S < N_D$, the following degrees of freedom are achievable:*
If $T_{SD} = T_{RD}$,

$$d = (1 - \frac{N_S + n_r}{T_{SD}}) \max_{n_r} \min\{N_S + n_r, N_S + \frac{(K-1)N_S^* N_S}{K(T_{SD} - n_r - N_S)}\}. \tag{4.39}$$

If $T_{RD} = K'T_{SD}$ and all coherence length pairs have integer ratios, equivalently $\frac{\max(K,K')}{\min(K,K')} \in \mathbb{N}$,

$$\begin{aligned} d = (1 - \frac{N_S + n_r}{T_{SD}}) \max_{n_r} \min\Big\{&(N_S + n_r)(1 + \frac{(K'-1)n_r}{K'(T_{SD} - n_r - N_S)}), \\ &N_S(1 + \frac{K'(K-1)N_S^* + K(K'-1)n_r}{KK'(T_{SD} - n_r - N_S)})\Big\}. \end{aligned} \tag{4.40}$$

If $T_{SD} = K'T_{RD}$

$$d = (1 - \frac{n_r}{T_{RD}}) \max_{n_r} \min\Big\{(N_S + n_r)(1 - \frac{N_S}{K'(T_{RD} - n_r)}), N_S(1 + \frac{(K-1)N_S^* - KN_S}{KK'(T_{RD} - n_r)})\Big\}. \tag{4.41}$$

Proof. When $T_{SD} = T_{RD}$, our transmission block has length KT_{SD}. This transmit block has K sub-blocks with length T_{SD}. During the first sub-block, the source sends the signal

$$\mathbf{X}_S^1 = [\mathbf{I}_{N_S}, \mathbf{0}_{N_S \times n_r}, \mathbf{V}_S^1], \tag{4.42}$$

where $\mathbf{V}^1 \in \mathbb{C}^{N_S \times (T_{SD} - n_r - N_S)}$. The destination estimates the channel $\mathbf{H}_{SD}$. The relay estimates $\mathbf{H}_{SR}$ during the first N_S time slots.

In the next $(K-1)$ sub-blocks, the source sends

$$\mathbf{X}_S^k = \mathbf{U}^k[\mathbf{I}_{N_S}, \mathbf{0}_{N_S \times n_r}, \mathbf{V}_S^k],\ k = 2, \ldots, K, \tag{4.43}$$

where $n_r \le \min\{N_S, N_D - N_S\}$, $\mathbf{U}^k \in \mathbb{C}^{N_S \times N_S}$ and $\mathbf{V}_S^k \in \mathbb{C}^{N_S \times (T_{SD} - n_r - N_S)}$.

The received signal at the relay is

$$\mathbf{Y}_R^k = \mathbf{H}_{SR}\mathbf{U}^k[\mathbf{I}_{N_S}, \mathbf{0}_{N_S\times n_r}, \mathbf{V}_S^k]. \tag{4.44}$$

The relay knows $\mathbf{H}_{SR}$ and decodes $\mathbf{U}^k$ and uses n_r transmit antennas, sending

$$\mathbf{X}_R^k = [\mathbf{0}_{n_r\times N_S}, \mathbf{I}_{n_r}, \mathbf{V}_R^k] \in \mathbb{C}^{n_r\times T_{SD}}, \tag{4.45}$$

where $\mathbf{V}_R^k \in \mathbb{C}^{n_r\times(T_{SD}-n_r-N_S)}$. The received signal at the destination is:

$$\mathbf{Y}_D^k = \mathbf{H}_{SD}^k\mathbf{X}_S^k + \mathbf{H}_{RD}^k\mathbf{X}_R^k + \mathbf{W}_D^k = [\mathbf{H}_{SD}^k\mathbf{U}^k, \mathbf{H}_{RD}^k]\Big[\mathbf{I}_{(N_S+n_r)}, \begin{bmatrix}\mathbf{V}_S^k \\ \mathbf{V}_R^k\end{bmatrix}\Big] + \mathbf{W}_D^k. \tag{4.46}$$

The destination estimates the equivalent channel $[\mathbf{H}_{SD}^k\mathbf{U}^k, \mathbf{H}_{RD}^k]$ during the first $(N_S + n_r)$ time slots , and decodes $\mathbf{V}_S^k$ and $\mathbf{V}_R^k$, respectively provide degrees of freedom $N_S(T-n_r-N_S)$ and $n_r(T-n_r-N_S)$ per transmit block of length T_{SD}. For all $k \in \{2, \ldots, K\}$, the degrees of freedom provided via $\mathbf{U}^k$ are $N_S^*N_S$, hence the total degrees of freedom the relay can decode are $(K-1)N_S^*N_S$ per transmit block of length KT_{SD}. Therefore, adding up the degrees of freedom during the super block of length T_{SD} and optimizing the number of relay transmit antennas, the end-to-end degrees of freedom are given by (4.39). This completes the first part of the theorem.

We now consider $T_{RD} = K'T_{SD}$. Recall that in this section we are focusing on fading blocks that are aligned, thus the ratio of any pair of coherence times is an integer. Therefore we have the following two cases for the coherence time configurations.

In the first case, $T_{RD} = (K'/K)T_{SR} = K'T_{SD}$, where (K'/K) is an integer. Our transmission block from the source has length T_{RD} and is divided into sub-blocks with length T_{SR}. During the first sub-block, from time slot 1 to T_{SD}, the source sends the signal

$$\mathbf{X}_S^1 = [\mathbf{I}_{N_S}, \mathbf{0}_{N_S\times n_r}, \mathbf{V}_S^1] \in \mathbb{C}^{N_S\times T_{SD}}, \tag{4.47}$$

where $n_r \leq \min\{N_S, N_D - N_S\}$, $\mathbf{V}^1 \in \mathbb{C}^{N_S\times(T_{SD}-n_r-N_S)}$.

The relay estimates $\mathbf{H}_{SR}$ and sends

$$\mathbf{X}_R^1 = [\mathbf{0}_{n_r \times N_S}, \mathbf{I}_{n_r}, \mathbf{V}_R^1] \in \mathbb{C}^{n_r \times T_{SD}}, \tag{4.48}$$

where $\mathbf{V}_R^1 \in \mathbb{C}^{n_r \times (T_{SD} - n_r - N_S)}$. The received signal at the destination is

$$\mathbf{Y}_D^1 = [\mathbf{H}_{SD}^1,\ \mathbf{H}_{RD}] \Bigg[\mathbf{I}_{(N_S+n_r)}, \ \begin{bmatrix} \mathbf{V}_S^1 \\ \mathbf{V}_R^1 \end{bmatrix} \Bigg] + \mathbf{W}_D^1. \tag{4.49}$$

The destination estimates $[\mathbf{H}_{SD}^1,\ \mathbf{H}_{RD}]$ and decodes $\mathbf{V}_S^1$ and $\mathbf{V}_R^1$. Then every T_{SD} time slots, the source sends the signal

$$\mathbf{X}_S^k = \mathbf{U}^{k-1}[\mathbf{I}_{N_S}, \mathbf{0}_{N_S \times n_r}, \mathbf{V}_S^k] \in \mathbb{C}^{N_S \times T_{SD}}, \tag{4.50}$$

where $\mathbf{U}^{k-1} \in \mathbb{C}^{N_S \times N_S}$,$\mathbf{V}_S^k \in \mathbb{C}^{N_S \times (T_{SD} - n_r - N_S)}$. The relay decodes $\mathbf{U}^{k-1}$ and sends

$$\mathbf{X}_R^k = [\mathbf{0}_{n_r \times N_S}, \mathbf{V}_R^k] \in \mathbb{C}^{n_r \times T_{SD}}, \tag{4.51}$$

where $\mathbf{V}_R^k \in \mathbb{C}^{n_r \times (T_{SD} - N_S)}$. The received signal at the destination is

$$\mathbf{Y}_D^k = [\mathbf{H}_{SD}^k \mathbf{U}^k, \mathbf{H}_{RD}] \Bigg[\mathbf{I}_{N_S}, \begin{bmatrix} \mathbf{V}_S^k \\ \mathbf{V}_R^k \end{bmatrix} \Bigg] + \mathbf{W}_D^k. \tag{4.52}$$

The destination can decode $\mathbf{V}_S^k$ and $\mathbf{V}_R^k$ which respectively provide degrees of freedom $N_S(T - N_S)$ and $n_r(T - N_S)$ and $\mathbf{U}^k$ can provide degrees of freedom $N_S^* N_S$.

In the remaining sub-block of length T_{SR}, the relay-destination channel keeps constant and it has already been estimated by the destination. Therefore, the relay does not need to send pilots. Then every T_{SD} time slots, the transmitted signals at the source are:[1]

$$\mathbf{X}_S^1 = [\mathbf{I}_{N_S}, \mathbf{V}_S^k], \tag{4.53}$$

[1]The following expression represents the signaling structure, the information carrying matrices $\mathbf{U}^k$ and $\mathbf{V}_S^k$ are independent across different sub-blocks, but for convenience, we use the same notation across different sub-blocks.

$$\mathbf{X}_S^k = \mathbf{U}^k[\mathbf{I}_{N_S}, \mathbf{V}_S^k], 2 \le k \le K, \tag{4.54}$$

where $n_r \le \min\{N_S, N_D - N_S\}$, $\mathbf{V}_S^k \in \mathbb{C}^{N_S \times (T_{SD} - n_r - N_S)}$. The relay decodes $\mathbf{U}^k$ and sends (4.51), with the codeword representing the message of the latest decoded $\mathbf{U}^k$. In this way, the destination can decode $\mathbf{V}_S^k$ and $\mathbf{V}_R^k$.

The degrees of freedom the relay can decode are $\frac{K'}{K}(K-1)N_S^* N_S$. The source-destination link achieves total degrees of freedom $N_S(K'T_{SD} - n_r - K'N_S)$. The relay-destination link achieves total degrees of freedom $n_r(K'T_{SD} - n_r - K'N_S)$. Adding it up with the degrees of freedom the source-destination link achieves and optimizing the number of relay transmit antennas, it results in the achievable degrees of freedom in (4.40).

In the second case, $T_{SR} = (K/K')T_{RD} = KT_{SD}$, where (K/K') is an integer. The transmission block from the source has length T_{SR} and is divided into sub-blocks with length T_{RD}. During the first sub-block, from time slot 1 to T_{SD}, the source sends the signal

$$\mathbf{X}_S^1 = [\mathbf{I}_{N_S}, \mathbf{0}_{N_S \times n_r}, \mathbf{V}_S^1] \in \mathbb{C}^{N_S \times T_{SD}}, \tag{4.55}$$

where $n_r \le \min\{N_S, N_D - N_S\}$,$\mathbf{V}^1 \in \mathbb{C}^{N_S \times (T_{SD} - n_r - N_S)}$.

The relay estimates $\mathbf{H}_{SR}$ and sends

$$\mathbf{X}_R^1 = [\mathbf{0}_{n_r \times N_S}, \mathbf{I}_{n_r}, \mathbf{V}_R^1] \in \mathbb{C}^{n_r \times T_{SD}}, \tag{4.56}$$

where $\mathbf{V}_R^1 \in \mathbb{C}^{n_r \times (T_{SD} - n_r - N_S)}$. The received signal at the destination is

$$\mathbf{Y}_D^1 = [\mathbf{H}_{SD}^1, \mathbf{H}_{RD}] \left[\mathbf{I}_{(N_S + n_r)}, \quad \begin{bmatrix} \mathbf{V}_S^1 \\ \mathbf{V}_R^1 \end{bmatrix} \right] + \mathbf{W}_D^1. \tag{4.57}$$

The destination estimates $[\mathbf{H}_{SD}^1, \mathbf{H}_{RD}]$ and decodes $\mathbf{V}_S^1$ and $\mathbf{V}_R^1$. Then every T_{SD} time slots, the source sends the signal

$$\mathbf{X}_S^k = \mathbf{U}^{k-1}[\mathbf{I}_{N_S}, \mathbf{V}_S^k] \in \mathbb{C}^{N_S \times T_{SD}}, \tag{4.58}$$

where $\mathbf{U}^{k-1} \in \mathbb{C}^{N_S \times N_S}$, $\mathbf{V}_S^k \in \mathbb{C}^{N_S \times (T_{SD}-N_S)}$. The relay decodes $\mathbf{U}^{k-1}$ and sends

$$\mathbf{X}_R^k = [\mathbf{0}_{n_r \times N_S}, \mathbf{V}_R^k] \in \mathbb{C}^{n_r \times T_{SD}}, \tag{4.59}$$

where $\mathbf{V}_R^k \in \mathbb{C}^{n_r \times (T_{SD}-N_S)}$. The received signal at the destination is

$$\mathbf{Y}_D^k = [\mathbf{H}_{SD}^k \mathbf{U}^k, \mathbf{H}_{RD}] \left[\mathbf{I}_{N_S}, \begin{bmatrix} \mathbf{V}_S^k \\ \mathbf{V}_R^k \end{bmatrix} \right] + \mathbf{W}_D^k. \tag{4.60}$$

During each remaining sub-block of length T_{RD}, from time slot 1 to T_{SD}, the source sends the signal

$$\mathbf{X}_S^1 = \mathbf{U}^1 [\mathbf{I}_{N_S}, \mathbf{0}_{N_S \times n_r}, \mathbf{V}_S^1], \tag{4.61}$$

where $n_r \leq \min\{N_S, N_D - N_S\}$, $\mathbf{V}^1 \in \mathbb{C}^{N_S \times (T_{SD}-n_r-N_S)}$.

The relay decodes $\mathbf{U}^1$ and sends

$$\mathbf{X}_R^1 = [\mathbf{0}_{n_r \times N_S}, \mathbf{I}_{n_r}, \mathbf{V}_R^1] \in \mathbb{C}^{n_r \times T_{SD}}, \tag{4.62}$$

where $\mathbf{V}_R^1 \in \mathbb{C}^{n_r \times (T_{SD}-n_r-N_S)}$. The received signal at the destination is

$$\mathbf{Y}_D^1 = [\mathbf{H}_{SD}^1, \mathbf{H}_{RD}] \left[\mathbf{I}_{(N_S+n_r)}, \begin{bmatrix} \mathbf{V}_S^1 \\ \mathbf{V}_R^1 \end{bmatrix} \right] + \mathbf{W}_D^1. \tag{4.63}$$

The destination estimates $[\mathbf{H}_{SD}^1, \mathbf{H}_{RD}]$ and decodes $\mathbf{V}_S^1$ and $\mathbf{V}_R^1$. Then every T_{SD} time slots, the source and the relay sends the signal with the same structure as (4.58),(4.59). The received signal at the destination is

$$\mathbf{Y}_D^k = [\mathbf{H}_{SD}^k \mathbf{U}^k, \mathbf{H}_{RD}] \left[\mathbf{I}_{N_S}, \begin{bmatrix} \mathbf{V}_S^k \\ \mathbf{V}_R^k \end{bmatrix} \right] + \mathbf{W}_D^k. \tag{4.64}$$

The degrees of freedom the relay can decode are $(K-1)N_S^* N_S$. The source-destination link achieves total degrees of freedom $N_S(KT_{SD} - \frac{K}{K'} n_r - KN_S)$. The relay-destination link can provide total degrees of freedom $n_r(KT_{SD} - \frac{K}{K'} n_r - KN_S)$. Take the minimum of the degrees of freedom the relay can decode and can transmit. Adding up the degrees of freedom the source-destination link achieves and optimizing the number of relay transmit antennas, it results in the achievable degrees of freedom in (4.40). This completes the second part of the theorem.

When $T_{SD} = K'T_{RD}$, our source transmission block has length T_{SR} and is divided into sub-blocks with length T_{SD}. In the first sub-block, during time slot 1 to T_{RD}, the source sends

$$\mathbf{X}_S^1 = [\mathbf{I}_{N_S}, \mathbf{0}_{N_S \times n_r}, \mathbf{V}_S^1] \in \mathbb{C}^{N_S \times T_{RD}}. \tag{4.65}$$

The relay estimates $\mathbf{H}_{SR}$ and sends

$$\mathbf{X}_R^1 = [\mathbf{0}_{n_r \times N_S}, \mathbf{I}_{n_r}, \mathbf{V}_R^k] \in \mathbb{C}^{n_r \times T_{RD}}. \tag{4.66}$$

During the remaining $(K'-1)T_{RD}$ time slots, every T_{RD} time slots, the source sends

$$\mathbf{X}_S^k = [\mathbf{0}_{N_S \times n_r}, \mathbf{V}_S^k] \in \mathbb{C}^{N_S \times T_{RD}}, 2 \le k \le K', \tag{4.67}$$

and the relay sends

$$\mathbf{X}_R^k = [\mathbf{I}_{n_r}, \mathbf{V}_R^k] \in \mathbb{C}^{n_r \times T_{RD}}, \tag{4.68}$$

where $n_r \le \min\{N_S, N_D - N_S\}$, $\mathbf{V}_R^1 \in \mathbb{C}^{N_S \times (T_{RD} - n_r - N_s)}$ and $\mathbf{V}_R^k \in \mathbb{C}^{N_S \times (T_{RD} - n_r)}$, $k = 2, \ldots, K$.

During the first T_{RD} time slots, the received signal at the destination is

$$\mathbf{Y}_D^1 = [\mathbf{H}_{SD}, \mathbf{H}_{RD}^1] \left[\mathbf{I}_{(N_S+n_r)}, \begin{bmatrix} \mathbf{V}_S^1 \\ \mathbf{X}_D^1 \end{bmatrix} \right] + \mathbf{W}_D^1.$$

The destination estimates $[\mathbf{H}_{SD}, \mathbf{H}_{RD}^1]$ during the first $(N_S + n_r)$ time slots and decodes $\mathbf{V}_S^1$ and $\mathbf{V}_R^1$. Then every T_{RD} time slots, the received signal at the destination is

$$\mathbf{Y}_D^k = \left[\mathbf{H}_{RD}^k,\ [\mathbf{H}_{SD}, \mathbf{H}_{RD}^k] \begin{bmatrix} \mathbf{V}_S^k \\ \mathbf{V}_R^k \end{bmatrix} \right] + \mathbf{W}_D^k.$$

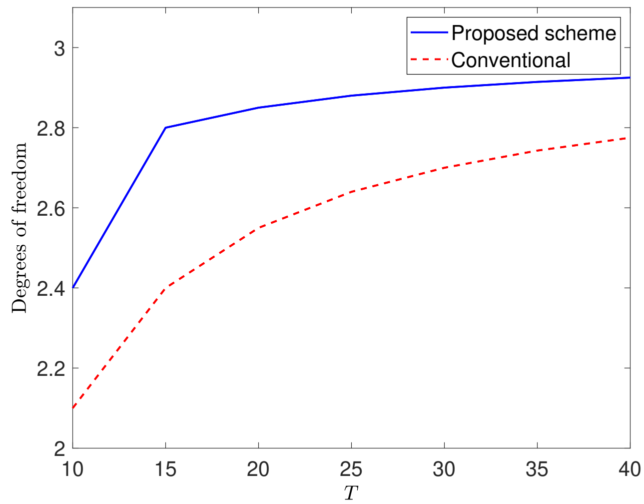

Figure 4.3. DoF for $T_{SR} = \infty$, $T_{RD} = T_{SD} = T$

In the following $(K-1)$ sub-block of length T_{SD}, the source-relay channel $\mathbf{H}_{SR}$ keeps constant and has already been estimated by the relay. Therefore, we copy the transmission strategy in the proof of Theorem 9, when the relay knows the channel $\mathbf{H}_{SR}$. The source and the relay send the signals (4.21), (4.22), (4.24), and (4.25).

The degrees of freedom the relay can decode are $(K-1)N_S^* N_S$. The source-destination link can provide total degrees of freedom $KN_S(K'T_{RD} - K'n_r - N_S)$. The relay-destination link can provide total degrees of freedom $Kn_r(K'T_{RD} - K'n_r - N_S)$. Adding it up with the degrees of freedom the source-destination link achieves and optimizing the number of the relay transmit antennas, the achievable degrees of freedom in (4.41) are obtained. This completes the proof. $\square$

Figure 4.3 compares the performance of the proposed scheme with a conventional transmission strategy which designs signals according to the shortest coherence time, demonstrating the gains in degrees of freedom. The antenna configuration is $N_S = N_R = 3$ and $N_D = 5$. The coherence intervals are $T_{SR} = \infty$, $T_{RD} = T_{SD} = T$. The proposed scheme has a significant gain in degrees of freedom over the conventional transmission. Figure 4.4 considers the

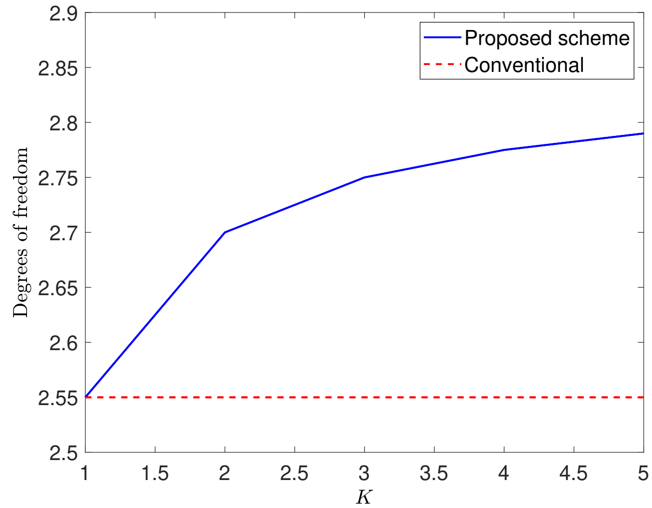

Figure 4.4. DoF for $T_{SR} = KT_{RD} = KT_{SD} = KT$, $T = 10$

case where $T_{SR} = KT_{RD} = KT_{SD} = KT$,$T = 10$ for different K. When $K = 1$, i.e., all links have identical coherence times, and there is no degrees of freedom gain to be obtained; when K grows, the gain achieved by the proposed scheme increases.

4.3 Achievable DoF with Relay Scheduling

In this section, a new scheme combining product superposition and relay scheduling is introduced. The following theorem highlights the main result of this section. For convenience and compact expression of the results, we define:

$$d_1 \triangleq N_S(T_{SD} - n_r - N_S),$$

$$d_2 \triangleq n_r(T_{SD} - n_r - N_S),$$

$$d_3 \triangleq N_S \min\{N_S, N_R\},$$

Theorem 11. *In a relay channel with link coherence times $T_{SR} = \infty$, $T_{SD} = T_{SR}$ and antenna configuration $N_S < N_D$, under aligned coherence blocks,*

- *If $d_2 \leq d_3$, the degrees of freedom $d = \frac{1}{T_{SD}} \max_{n_r}(d_1 + d_2)$ are achievable.*

- *If $d_2 > d_3$, the following degrees of freedom are achievable,*

$$d = \frac{1}{T_{SD}} \max_{n_r} \left(\frac{d_2 - d_3}{d_2} N_S (T_{SD} - N_S) + \frac{d_3}{d_2} (d_1 + d_2) \right). \tag{4.69}$$

Proof. If $d_2 \leq d_3$, the achievable degrees of freedom follow Theorem 9. When $d_2 > d_3$, the transmit scheme with relay scheduling has two phases, each of them lasting an integer multiple of the coherence interval T. In both phases, product superposition is used at the source, but the relay action is different in the two phases. We transmit for $d_2 - d_3$ coherence intervals in Phase 1, followed by transmitting d_3 coherence intervals in Phase 2.

During Phase 1, the relay transmission is deactivated, but the source continues to transmit via product superposition. In this phase, in each coherence interval of length T_{SD}, the source delivers to the destination data rates corresponding to its point-to-point degrees of freedom bound, which is $N_S(T_{SD} - N_s)$, while delivering *additional* data to the relay with degrees of freedom d_3. We transmit in Phase 1 for $d_2 - d_3$ coherence intervals, therefore, the normalized (per-symbol) average degrees of freedom contribution of this phase is $\frac{d_2-d_3}{d_2}\frac{1}{T_{SD}} N_S(T_{SD} - N_S)$.

During Phase 2, the relay is activated and the source sends the product superposition signal

$$\mathbf{X}_S = \mathbf{U}[\mathbf{I}_{N_S}, \mathbf{0}_{N_S \times n_r}, \mathbf{V}_S], \tag{4.70}$$

where $n_r \leq \min\{N_S, N_D - N_S\}$, $\mathbf{U} \in \mathbb{C}^{N_S \times N_S}$ and $\mathbf{V}_S \in \mathbb{C}^{N_S \times (T_{SD} - n_r - N_S)}$.

The relay knows $\mathbf{H}_{SR}$ and decodes $\mathbf{U}$. The relay uses n_r antennas for transmission, sending

$$\mathbf{X}_R = [\mathbf{0}_{n_r \times N_S}, \mathbf{I}_{n_r}, \mathbf{V}_R] \in \mathbb{C}^{n_r \times T_{SD}}, \tag{4.71}$$

where $\mathbf{V}_R \in \mathbb{C}^{n_r \times (T_{SD} - n_r - N_S)}$.

The destination estimates the equivalent channel $[\mathbf{H}_{SD}\mathbf{U},\ \mathbf{H}_{RD}]$ during the first $(N_S + n_r)$ time slots and then decodes its messages. Destination receives: $\mathbf{V}_S$ from the source and $\mathbf{V}_R$

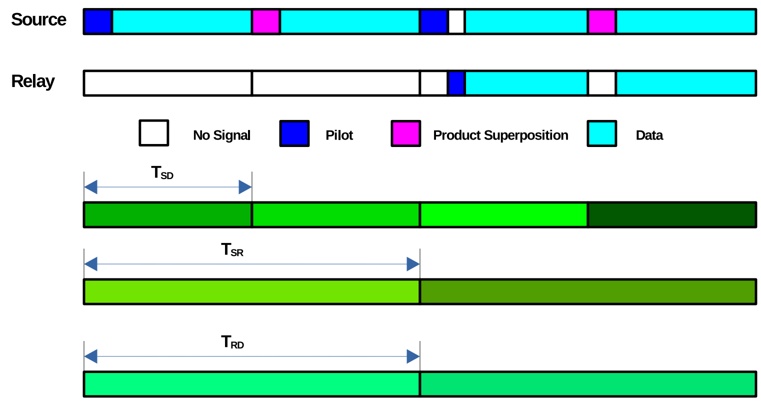

Figure 4.5. Signaling structure with relay scheduling

from the relay, providing degrees of freedom d_1 and d_2, respectively. Phase 2 consists of d_3 coherence intervals; further, recall that the relay has stored data available from Phase 1 in addition to the data it is receiving during Phase 2. Therefore, the relay can send data with degrees of freedom d_2 to the destination. Hence, during phase 2, the normalized per-symbol degrees of freedom are $\frac{1}{T_{SD}}\frac{d_3}{d_2}(d_1 + d_2)$.

Adding the degrees of freedom achieved in Phase 1 and Phase 2 and optimizing the number of relay transmit antennas to be activated produces (4.69). This completes the proof. □

Remark 4. *For comparison, we also mention the degrees of freedom* without *relay scheduling. For a relay with the following setup, $T_{SR} = \infty$, $T_{SD} = T_{SR} = T$ and $N_S < N_D$. From Theorem 9, the following degrees of freedom are achievable:*

$$d = \frac{1}{T_{SD}} \max_{n_r} \min\{d_1 + d_2, d_1 + d_3\}.$$

Figure 4.5 shows the signaling structure of the proposed scheme combining product superposition and relay scheduling. Figure 4.6 shows the comparison between the achievable degrees of freedom of product superposition alone and with relay scheduling when $N_S = 3, N_D = 5$ for different T.

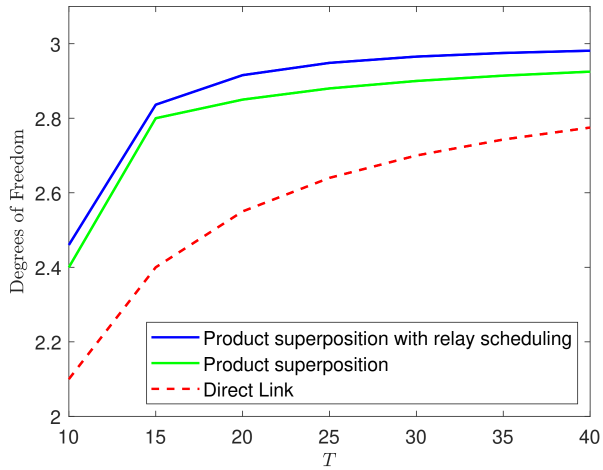

Figure 4.6. Achievable DoF in Theorem 11

4.4 General Coherence Times

4.4.1 Unaligned Coherence Blocks

We now consider the scenario when the coherence blocks are not perfectly aligned. To build intuition and motivation for the proposed approach, we begin with an unaligned counterpart to the toy example in Section 4.2.2. Then we generalize the result to arbitrary coherence times.

The unaligned toy example is as follows: the source and relay are equipped with 2 antennas and the destination is equipped with 3 antennas. The coherence times of the three links are as follows: $T_{SR} = \infty$, i.e., the source-relay channel is static, therefore the cost of training over this link is amortized over a large number of samples and we can assume the relay knows $\mathbf{H}_{SR}$. Furthermore we assume $T_{SD} = T_{RD} = 8$. The coherence block of the channel $\mathbf{H}_{RD}$ starts from the 5th time slot of $\mathbf{H}_{SR}$.

The source uses product superposition, sending

$$\mathbf{X}_S = \mathbf{U}[\mathbf{I}_2, \mathbf{0}_{2\times 1}, \mathbf{V}_S], \tag{4.72}$$

where $\mathbf{U} \in \mathbb{C}^{2\times 2}$ and $\mathbf{V}_S \in \mathbb{C}^{2\times 5}$.

At the relay, the received signal is

$$\mathbf{Y}_R = \mathbf{H}_{SR}\mathbf{X}_S + \mathbf{W}_R = \mathbf{H}_{SR}\mathbf{U}[\mathbf{I}_2, \mathbf{0}_{2\times 1}, \mathbf{V}_S] + \mathbf{W}_R. \quad (4.73)$$

The received signal at the first two time slots is

$$\mathbf{Y}'_R = \mathbf{H}_{SR}\mathbf{U} + \mathbf{W}'_R. \quad (4.74)$$

The relay knows $\mathbf{H}_{SR}$ and decodes $\mathbf{U}$. Assume the signal decoded by the relay in the previous block is $\mathbf{U}'$ and the two rows of $\mathbf{U}'$ are $\mathbf{u}'_1, \mathbf{u}'_2 \in \mathbb{C}^{1\times 2}$.

The relay uses one antenna for transmission and sends

$$\mathbf{X}_R = [\mathbf{0}_{1\times 2}, 1, \mathbf{u}'_1, \mathbf{u}'_2, 0] \in \mathbb{C}^{1\times 8}. \quad (4.75)$$

Now in one coherence block of $\mathbf{H}_{SD}$, because of the unaligned blocks of $\mathbf{H}_{RD}$, the received signal at the destination will experience two realizations of $\mathbf{H}_{RD}$ in the first 4 time slots,

$$\mathbf{Y}_D = [\mathbf{H}_{SD}\mathbf{U}, \mathbf{H}_{RD1}]\left[\mathbf{I}_3, \begin{bmatrix} \mathbf{V}_S^1 \\ \mathbf{u}'_1(1) \end{bmatrix}\right] + \mathbf{W}_D. \quad (4.76)$$

The destination estimates the equivalent channel $[\mathbf{H}_{SD}\mathbf{U}, \mathbf{H}_{RD1}]$ in the first three time slots and decodes $\mathbf{V}_S, \mathbf{u}'_1(1)$.

In the next 4 time slots, the received signal is:

$$\mathbf{Y}_D = [\mathbf{H}_{SD}\mathbf{U}, \mathbf{H}_{RD2}]\begin{bmatrix} \mathbf{V}_S \\ \mathbf{u}'_1(2), \mathbf{u}'_2, 0 \end{bmatrix} + \mathbf{W}_D. \quad (4.77)$$

The first part of the equivalent channel $\mathbf{H}_{SD}\mathbf{U}$ is already estimated. The second part $\mathbf{H}_{RD2}$ will be estimated in the next transmit block. Therefore, the destination decodes $\mathbf{V}_S^2, \mathbf{u}'_1(2)$ and $\mathbf{u}'_2$. This shows that when the coherence blocks from the source and relay to the destination are unaligned, the proposed scheme can still be used. The destination achieves the same degrees of freedom $d = (2 \times 5 + 2 \times 1 \times 2)/8 = 1.75$ when the coherence blocks are aligned and with the same coherence times. Recall that for a conventional technique that trains all links according to the shortest coherence interval, the degrees of freedom are $d' = 2 \times (8-2)/8 = 1.5$.

A similar reasoning can be used to verify that when the coherence blocks from the source to the relay and the source to destination are unaligned, the offset of these coherence blocks will not affect the achievability of our proposed scheme.

4.4.2 Arbitrary Coherence Times

The following theorem states the achievable degrees of freedom with arbitrary coherence times and Figure 4.7 illustrates the signaling structure for the achievable scheme.

Theorem 12. *In a relay channel with link coherence times satisfying $T_{SR} > T_{SD}$, $T_{RD} > T_{SD}$. and antenna configuration $N_S, N_R < N_D$, the following degrees of freedom are achievable:*

$$\begin{aligned} d = \frac{1}{T_{SR}T_{SD}T_{RD}} \max_{n_r} \{ & N_S(T_{SR}T_{SD}T_{RD} - N_S T_{SR}T_{RD} - n_r T_{SR}T_{SD}) + \\ & \min\{N_S^* N_S(T_{SR}T_{RD} - T_{SD}T_{RD}), n_r(T_{SR}T_{SD}T_{RD} - N_S T_{SR}T_{RD} - n_r T_{SR}T_{SD})\}\}. \end{aligned} \tag{4.78}$$

where $N_S^* \triangleq \min\{N_S, N_R\}$.

Proof. Design the pilot-based achievable scheme in the following manner:

- On the multiple-access side, pilots sent from the relay and the source will be allocated in different time slots, such that they will not interfere with each other. In addition, during these time slots no data is sent, avoiding pilot contamination.

- On the broadcast side, the source-relay link needs fewer pilots than the source-destination. Thus, product superposition enables transmission of additional data to the relay.

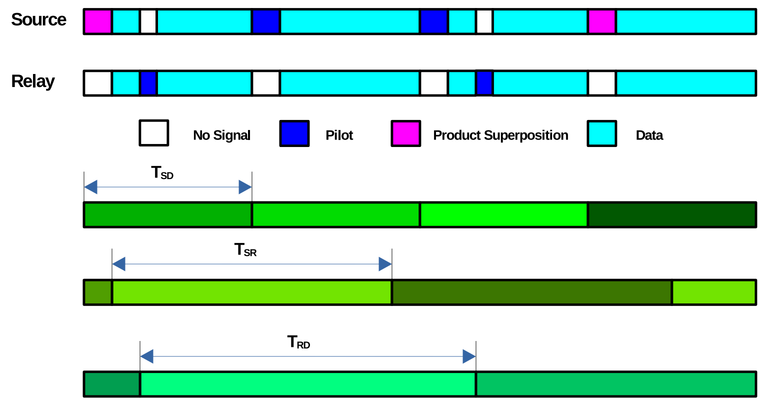

Figure 4.7. Signaling structure for relay with arbitrary coherence times

In the following, we consider a super-interval of length $T_{SR}T_{RD}T_{SD}$, after which the coherence intervals will come back to their original alignment. The achievable degrees of freedom are calculated as follows:

In each source-destination coherence interval T_{SD}, N_S pilot symbols are transmitted. We call the pilot symbols in each coherence block a *pilot sequence*.

Therefore, for source-destination link, we repeat the length-N_S pilot sequence $T_{SR}T_{RD}$ times over the length-$T_{SR}T_{RD}T_{SD}$ super-interval. Having coherence time $T_{SR}T_{RD}$, the relay needs $T_{SD}T_{RD}$ pilot sequences. Hence, product superposition can be applied during $T_{SR}T_{RD} - T_{SD}T_{RD}$ pilot sequences of length N_S to send data to the relay. Data with N_S^* degrees of freedom per symbol can be sent.

Over each super-interval, the relay-destination link needs $T_{SR}T_{SD}$ pilot sequences of length n_r. The pilot slots will be non-overlapping with pilots transmitted from the source terminal.

In each super-interval, the source and the relay each have $(T_{SR}T_{SD}T_{RD} - N_S T_{SR}T_{RD} - n_r T_{SR}T_{SD})$ time slots available for sending data. The source has N_S degrees of freedom available per transmission, and the relay n_r degrees of freedom per transmission.

The relay can decode at most $N_S^* N_S(T_{SR}T_{RD} - T_{SD}T_{RD})$ degrees of freedom, therefore, it provides $\min\{N_S^* N_S(T_{SR}T_{RD} - T_{SD}T_{RD}), n_r(T_{SR}T_{SD}T_{RD} - N_S T_{SR}T_{RD} - n_r T_{SR}T_{SD})\}$ degrees of freedom, the minimum of the degrees of freedom the relay can receive and can transmit.

We can now add the degrees of freedom of the source transmission (subject to relay constraints) with the degrees of freedom provided by the relay transmission, and optimize the number of relay antennas to be activated. This concludes the proof. □

The following corollary shows the achievable degrees of freedom when using relay scheduling with arbitrary coherence times.

Corollary 5. *Define the following notation:*

$$\begin{aligned}
d_1 &\triangleq N_S(1 - \frac{N_S}{T_{SD}} - \frac{n_r}{T_{RD}}), \\
d_2 &\triangleq n_r(1 - \frac{N_S}{T_{SD}} - \frac{n_r}{T_{RD}}), \\
d_3 &\triangleq N_S^* N_S(\frac{1}{T_{SD}} - \frac{1}{T_{RD}}).
\end{aligned}$$

In a relay with coherence diversity,

- *If $d_2 \leq d_3$, the degrees of freedom $d = \max_{n_r}(d_1 + d_2)$ are achievable.*
- *If $d_2 > d_3$, the following degrees of freedom are achievable.*

$$d = \max_{n_r}(\frac{d_2 - d_3}{d_2} N_S(1 - \frac{N_S}{T_{SD}}) + \frac{d_3}{d_2}(d_1 + d_2)), \tag{4.79}$$

Proof. If $d_2 \leq d_3$, the achievable degrees of freedom follows Theorem 12. When $d_2 > d_3$. the transmit scheme with relay scheduling has two phases. In both phases, product superposition is used at the source, but the relay action is different in the two phases, as described in the sequel. We propose to transmit for $d_2 - d_3$ coherence intervals in Phase 1, followed by transmitting d_3 coherence intervals in Phase 2.

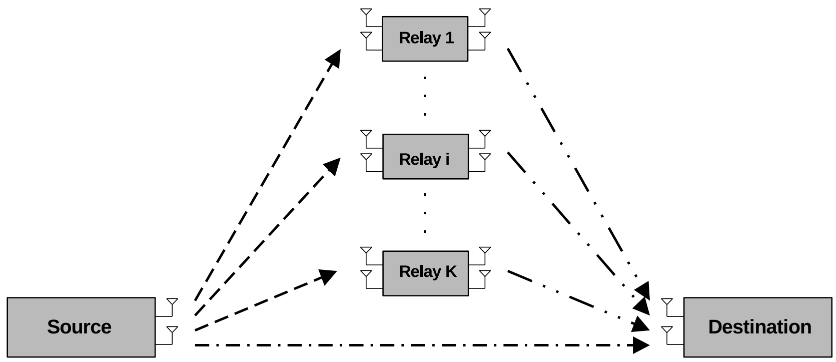

Figure 4.8. Channel with multiple parallel relays

During Phase 1, the relay transmission is deactivated but the source continues to transmit via product superposition. In this phase, the source delivers to the destination data rates corresponding to its point-to-point degrees of freedom bound, following the result in [52] which is $N_S(T_{SD} - N_S)$, while delivering *additional* data to the relay with degrees of freedom d_3. We transmit in Phase 1 for $d_2 - d_3$ coherence intervals, therefore, the normalized (per-symbol) average degrees of freedom contribution of this phase is $\frac{d_2-d_3}{d_2}\frac{1}{T_{SD}}N_S(T_{SD} - N_S)$.

During Phase 2, following the strategy from the proof of Theorem 11, the relay has stored the data available from Phase 1 in addition to the data it is receiving in Phase 2. Therefore, the relay can send data with degrees of freedom d_2 to the destination. Hence, during phase 2, the normalized per-symbol degrees of freedom are $\frac{d_3}{d_2}(d_1 + d_2)$.

Adding the degrees of freedom achieved in Phase 1 and Phase 2 and optimizing the number of relay transmit antennas to be activated produces (4.79). This completes the proof. □

4.5 Multiple Relays in Parallel

This section studies the MIMO relay channel with K full-duplex relays under coherence diversity. The source and destination are equipped with N_S and N_D antennas, respectively.

Relay k has $N_R(k)$ receive antennas and uses $n_R(k) \leq N_R(k)$ antennas for transmission. Figure 4.8 shows the structure of the system. The received signals at the relays and the destination are:

$$\mathbf{y}_R(k) = \mathbf{H}_{SR}(k)\mathbf{x}_S + \mathbf{w}_R(k), \qquad k = 1, \ldots, K \tag{4.80}$$

$$\mathbf{y}_D = \mathbf{H}_{SD}\mathbf{x}_S + \sum_{k=1,\ldots,K} \mathbf{H}_{RD}(k)\mathbf{x}_R(k) + \mathbf{w}_D, \tag{4.81}$$

where $\mathbf{x}_S$ and $\mathbf{x}_R(k)$ are signals transmitted from the source and Relay k. $\mathbf{w}_R$ and $\mathbf{w}_D$ are i.i.d. zero-mean Gaussian noise and $\mathbf{H}_{SR}(k)$, $\mathbf{H}_{RD}(k)$ and $\mathbf{H}_{SD}$ are the channel gain matrices, whose entries are i.i.d. Gaussian. We assume there is no free channel state information at the destination and no CSIT at the source or relay. In the parallel relay geometry, there are no inter-relay links. Denote the coherence time of the link between the source and Relay k as $T_{SR}(k)$ and the coherence time of the link between Relay k and the destination as $T_{RD}(k)$.

4.5.1 Achievable DoF for Two Parallel Relays

Consider the following channel with two parallel relays. $T_{SR}(2) = K_2 T_{SR}(1) = K_2 K_1 T_{SD} = K_2 K_1 T$ and the destination knows the channel state of $\mathbf{H}_{RD}(1)$ and $\mathbf{H}_{RD}(2)$, i.e., $T_{RD}(1) = T_{RD}(2) = \infty$. Denote $N_S^*(i) = \min\{N_S, N_R(i)\}$. If Relay 1 or Relay 2 is activated alone, the achievable degrees of freedom are

$$d_i = \max_{n_R(i)} \Big\{ N_S(1 - \frac{N_S}{T}) + \min\big\{(1 - \frac{1}{K_i})\frac{N_S^*(i)N_S}{T}, n_R(i)(1 - \frac{N_S}{T})\big\}\Big\}. \tag{4.82}$$

When Relay 1 and Relay 2 are both activated, consider a transmission interval of length K_2K_1T. During each coherence interval of length K_2K_1T, Relay 1 and Relay 2 send the messages they decoded in the previous interval. The transmitted signal from Relay 1 and 2 over each sub-interval of length T has the following structure and is repeated K_2K_1 times:

$$\mathbf{X}_R(i) = [\mathbf{0}_{n_R(i)\times N_S}, \mathbf{V}_{Ri}], \quad i = 1, 2. \tag{4.83}$$

During the first coherence interval of length K_1T, in the first sub-interval of length T, the source sends $\mathbf{X}_S = [\mathbf{I}_{N_S}, \mathbf{X}_D]$. Relay 1 and Relay 2 estimate their channels. The signal at the destination is

$$\begin{aligned}\mathbf{Y}_D &= [\mathbf{H}_{SD}, \mathbf{H}_{RD}(1), \mathbf{H}_{RD}(2)] \begin{bmatrix} \mathbf{I}_{N_S}, \mathbf{V}_D \\ \mathbf{0}_{n_R(1)\times N_S}, \mathbf{V}_{R1} \\ \mathbf{0}_{n_R(2)\times N_S}, \mathbf{V}_{R2} \end{bmatrix} + \mathbf{W}_D \\ &= \left[\mathbf{H}_{SD}, [\mathbf{H}_{SD}, \mathbf{H}_{RD}(1), \mathbf{H}_{RD}(2)] \begin{bmatrix} \mathbf{V}_D \\ \mathbf{V}_{R1} \\ \mathbf{V}_{R2} \end{bmatrix}\right] + \mathbf{W}_D.\end{aligned} \tag{4.84}$$

The destination estimates $\mathbf{H}_{SD}$ and decodes the messages in $\mathbf{V}_D, \mathbf{V}_{R1}$ and $\mathbf{V}_{R2}$, which respectively provide degrees of freedom $N_S, n_R(1)$ and $n_R(2)$ per symbol over this interval of length $(T - N_S)$.

In the remaining $K_1 - 1$ intervals of length T, the source sends the signal

$$\mathbf{X}_S = \mathbf{U}_1^i[\mathbf{I}_{N_S}, \mathbf{V}_D^i], \quad i = 1, 2, \ldots, K_1 - 1, \tag{4.85}$$

where $\mathbf{U}_1^i \in \mathbb{C}^{N_S \times N_S}$. Relay 1 has already estimated its channel in the first interval of length T. It can decode $\mathbf{U}_1^i$, achieving degrees of freedom $N_S^*(1)N_S$. The total degrees of freedom Relay 1 can decode are $(K_1 - 1)N_S^*(1)N_S$. The received signal at the destination is

$$\mathbf{Y}_D = \left[\mathbf{H}_{SD}\mathbf{U}_1^i, [\mathbf{H}_{SD}\mathbf{U}_1^i, \mathbf{H}_{RD}(1), \mathbf{H}_{RD}(2)] \begin{bmatrix} \mathbf{V}_D \\ \mathbf{V}_{R1} \\ \mathbf{V}_{R2} \end{bmatrix}\right] + \mathbf{W}_D. \tag{4.86}$$

The destination estimates $\mathbf{H}_{SD}\mathbf{U}_1^i$ and decodes $\mathbf{V}_D$,$\mathbf{V}_{R1}$, and $\mathbf{V}_{R2}$.

During each of the remaining $K_2 - 1$ coherence intervals of length K_1T, the transmitter sends a signal with the same structure as the first sub-interval of length T, multiplying it from the left by $\mathbf{U}_2^j$, which contains the message for Relay 2. During each interval of length K_1T, the transmitted signal from the source has the following structure

$$\mathbf{X}_S = \mathbf{U}_2^j\Bigg[[\mathbf{I}_{N_S}, \mathbf{X}_D^1], \mathbf{U}_1^1[\mathbf{I}_{N_S}, \mathbf{X}_D^2], \mathbf{U}_1^2[\mathbf{I}_{N_S}\mathbf{X}_D^3], \ldots, \mathbf{U}_1^{(K_1-1)}[\mathbf{I}_{N_S}, \mathbf{X}_D^{K_1}]\Bigg]. \tag{4.87}$$

During these $K_2 - 1$ coherence intervals with length K_1T, the channel $\mathbf{H}_{SR}(2)$ remains the same as in the first sub-interval of length K_1T. Therefore, in each interval of length K_1T, Relay 2 can achieve degrees of freedom $N_S^*(2)N_S$. The total degrees of freedom Relay 2 can decode are $(K_2 - 1)N_S^*(2)N_S$ over coherence interval of length K_2K_1T.

At Relay 1, the first N_S symbols received during the first sub-interval of length K_1T are

$$\mathbf{Y}_R(1) = \mathbf{H}_{SR}^j\mathbf{U}_2^j + \mathbf{W}_R(1). \tag{4.88}$$

The first N_S symbols during the remaining sub-interval of length K_1T received at Relay 1 are

$$\mathbf{Y}_R(1) = \mathbf{H}_{SR}^j\mathbf{U}_2^j\mathbf{U}_1^i + \mathbf{W}_R(1),\ i = 1, \ldots, K_1 - 1. \tag{4.89}$$

Relay 1 first estimates its equivalent channel

$$\tilde{\mathbf{H}}_{SR}^j(1) = \mathbf{H}_{SR}^j(1)\mathbf{U}_2^j, \tag{4.90}$$

and decodes $\mathbf{X}_R^i(1)$, which provides degrees of freedom $N_S^*(1)N_S$. The total degrees of freedom Relay 1 can decode are $(K_2 - 1)(K_1 - 1)N_S^*(1)N_S$.

At the destination, the received signal during the first sub-interval of length K_1T is

$$\mathbf{Y}_D = \left[\mathbf{H}_{SD}\mathbf{U}_2^j, [\mathbf{H}_{SD}\mathbf{U}_2^j, \mathbf{H}_{RD}(1), \mathbf{H}_{RD}(2)]\begin{bmatrix}\mathbf{V}_D\\ \mathbf{V}_{R1}\\ \mathbf{V}_{R2}\end{bmatrix}\right] + \mathbf{W}_D, \tag{4.91}$$

and the received signals during each remaining sub-intervals of length K_1T are

$$\mathbf{Y}_D = \left[\mathbf{H}_{SD}\mathbf{U}_2^j\mathbf{U}_1^i, [\mathbf{H}_{SD}\mathbf{U}_2^j\mathbf{U}_1^i, \mathbf{H}_{RD}(1), \mathbf{H}_{RD}(2)]\begin{bmatrix}\mathbf{V}_D\\ \mathbf{V}_{R1}\\ \mathbf{V}_{R2}\end{bmatrix}\right] + \mathbf{W}_D, \tag{4.92}$$

where $i = 1, \ldots, K_1 - 1$, $j = 1, \ldots, K_2 - 1$. The destination estimates the equivalent channel $\mathbf{H}_{SD}\mathbf{U}_2^j$, $\mathbf{H}_{SD}\mathbf{U}_2^j\mathbf{U}_1^i$, and decodes $\mathbf{X}_D, \mathbf{V}_{R1}$ and $\mathbf{V}_{R2}$, which respectively provide degrees of freedom $N_S, n_R(1), n_R(2)$ per symbol over each time interval of length K_1T.

During each interval of length K_2K_1T, the source-destination link can always provide degrees of freedom $N_S(1 - \frac{N_S}{T})$ per symbol. The maximum degrees of freedom decoded at Relay 1 are $(K_1-1)N_S^*(1)N_S + (K_2-1)(K_1-1)N_S^*(1)N_S = K_2(K_1-1)N_S^*(1)N_S$. The degrees of freedom decoded at Relay 2 are $(K_2-1)N_S^*(2)N_S$. During each interval of length K_2K_1T, the number of time slots available to relays for sending data is $K_2K_1(T - N_S)$. The degrees of freedom the relays can provide via the relay-destination links are $n_R(i)K_2K_1(T - N_S), i = 1, 2$. Noting that the emitted data by the relays is limited by what they can decode, we add the degrees of freedom of the two relays, normalize it per symbol, and optimize the number of transmit antennas activated at the relays. The following degrees of freedom are achievable

$$d = \max_{n_R(i)} \Big\{ N_S(1 - \frac{N_S}{T}) + \min\big\{(1 - \frac{1}{K_1})\frac{N_S^*(1)N_S}{T}, n_R(1)\frac{T - N_S}{T}\big\} + \min\big\{\frac{K_2 - 1}{K_1K_2}\frac{N_S^*(2)N_S}{T}, n_R(2)\frac{T - N_S}{T}\big\}\Big\}. \tag{4.93}$$

4.5.2 Achievable DoF for K Parallel Relays

We now extend the ideas and techniques that were developed in the two-relay framework to the K-relay case. In the interest of economy of expression, the parts that are similar to the earlier discussions are condensed or omitted.

Denote with $\mathbf{T}_{SR}$ and $\mathbf{T}_{RD}$ the size-K vectors containing, respectively, source-relay and relay-destination coherence times, and $\mathbf{N}_R, \mathbf{n}_R$ the number of receive and activated transmit antennas at the relays. Also, we allow a subset k of relays to be used. We denote the coherence times of selected relays with size-k vectors $\mathbf{T}', \mathbf{T}''$ and the number of receive and activated transmit antennas in selected relays with size-k vector $\mathbf{N}', \mathbf{n}'$. The following result shows the achievable degrees of freedom, which is maximized over selected relays and their activated transmit antennas. We define a selection matrix $\mathbf{P}_{k\times K}$ containing k rows of the identity matrix $\mathbf{I}_{K\times K}$, corresponding to the k indices of the selected relays.

Theorem 13. *For the multi-relay system* (4.80) *and* (4.81)*, the following degrees of freedom are achievable:*

$$d = \max_{\mathbf{P},\mathbf{n}',k} \Big\{ N_S(1-\frac{N_S}{T_{SD}} - \sum_{i=1}^{k}\frac{n'_i}{T''_i}) + \sum_{i=1}^{k}\min\big\{N_i^* N_S(\frac{1}{T'_{i-1}}-\frac{1}{T'_i}), n'_i(1-\frac{N_S}{T_{SD}}-\sum_{j=1}^{k}\frac{n'_j}{T''_j})\big\}\Big\},$$

$$\text{subject to:}\quad [\mathbf{T}'\ \mathbf{T}''\ \mathbf{N}'\ \mathbf{n}'] = \mathbf{P}[\mathbf{T}_{SR}\ \mathbf{T}_{RD}\ \mathbf{N}_R\ \mathbf{n}_R], \tag{4.94}$$

where $T'_0 \triangleq T_{SD}$*,* $\mathbf{P}$ *is a selection matrix consisting of* k *rows of the identity matrix of size* K*, and* $N_i^* = \min\{N_S, N'_i\}$*.*

Proof. The transmit scheme is designed in the same spirit as Theorem 12: On the multiple-access side, pilots sent from the relays and the source are allocated in different time slots; on the broadcast side, product superposition enables transmission of additional data to the relays. Throughout this proof, we index only the activated relays, e.g., Relay i refers to i-th activated relay. Without loss of generality, $T'_1 \le T'_2 \le \cdots \le T'_k$. Define $T_1 \triangleq \prod_{i=1}^{k} T'_i$ and $T_2 \triangleq \prod_{i=1}^{k} T''_i$. In the following, we consider a super-interval of length $T_1T_2T_{SD}$,

During each coherence interval of length T'_i, Relay i needs $T_{SD}T_2T_1/T'_i$ pilot sequences each of length N_S for channel estimation. Relay $(i-1)$ needs $T_{SD}T_2T_1/T'_{i-1}$ pilot sequences each of length N_S. Therefore, product superposition can be applied during $(T_{SD}T_2T_1/T'_{i-1} - T_{SD}T_2T_1/T'_i)$ pilot sequences each of length N_S to send data to Relay i, providing N_i^* degrees of freedom per symbol.

During each coherence interval of length T_{SD} in the source-destination link, N_S pilot symbols are transmitted. In each super-interval (see above) T_1T_2 pilot sequences of length N_S are transmitted.

For channel estimation between Relay i and the destination, during the super-interval of length $T_1T_2T_{SD}$, the destination needs $T_1T_2T_{SD}/T''_i$ pilot sequences of length n'_i.

Therefore, In each super-interval, the source and relays can use $(T_{SD}T_1T_2 - N_ST_1T_2 - \sum_{i=1}^{k} \frac{n'_i}{T''_i} T_1T_2T_{SD})$ time slots to send data. The source has N_S degrees of freedom available per transmission, and Relay i has n'_i degrees of freedom per transmission.

The decodable degrees of freedom Relay i are at most $N'_iN_S(T_{SD}T_2T_1/T'_{i-1} - T_{SD}T_2T_1/T'_i)$. Therefore, the degrees of freedom Relay i can provide are:

$$\min\left\{N'_iN_S(\frac{T_{SD}T_2T_1}{T'_{i-1}} - \frac{T_{SD}T_2T_1}{T'_i}), n'_i(T_{SD}T_1T_2 - N_ST_1T_2 - \sum_{i=1}^{k} \frac{n'_i}{T''_i} T_1T_2T_{SD})\right\}.$$

It is the minimum of the degrees of freedom Relay i can receive and can transmit.

We can now add up the degrees of freedom of the source and the relays and normalize it per symbol. Optimizing the relays to be activated (over k and $\mathbf{P}$) and the number of transmit antennas at the relays $\mathbf{n}'$, the degrees of freedom in (4.94) are achieved. This concludes the proof. $\square$

Figure 4.9 and 4.10 show the achievable degrees of freedom with two parallel relays equipped with $N_S = 3$, $N_D = 6$, $N_R(1) = N_R(2) = 1$ antennas. In Figure 4.9, $T_{SD} = 5$, $T_{RD}(1) = T_{RD}(2) = \infty$, $\frac{T_{SR}(1)}{T_{SR}(2)} = \frac{2}{3}$. In Figure 4.10, $T_{SD} = 5$, $T_{RD}(1) = T_{RD}(2) = \infty$, $T_{SR}(1) = 6$ and different $\frac{T_{SR}(2)}{T_{SR}(1)}$.

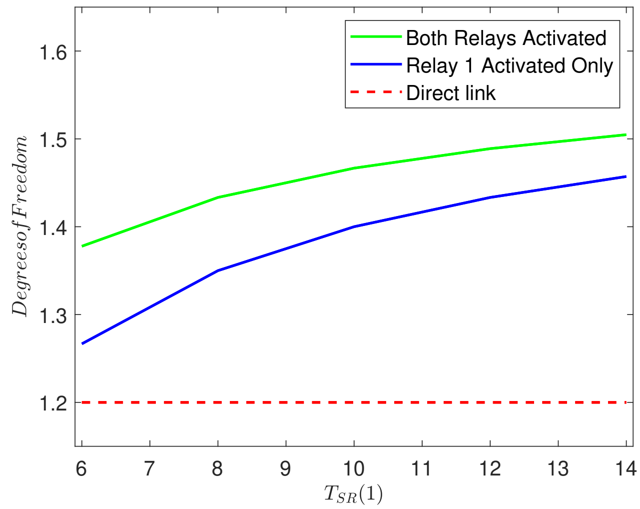

Figure 4.9. DoF subject to $\frac{T_{SR}(1)}{T_{SR}(2)} = \frac{2}{3}$

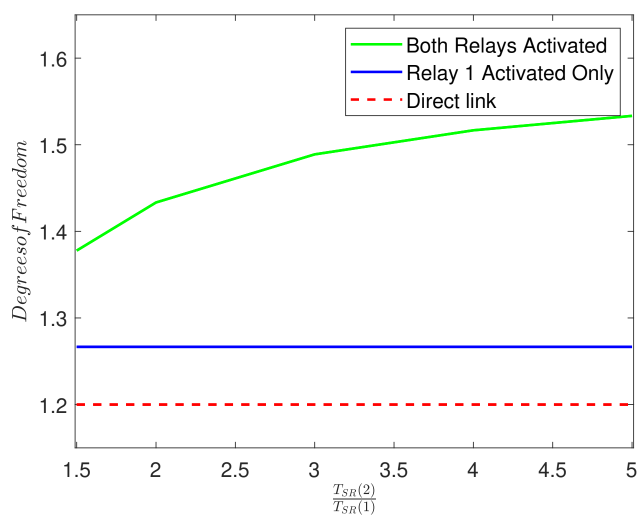

Figure 4.10. DoF subject to $T_{SR}(1) = 6$ with different $\frac{T_{SR}(2)}{T_{SR}(1)}$

CHAPTER 5

CONCLUSION

This dissertation investigates wireless channels with multiple users when they have coherence disparity in space and time. Correlation and coherence diversity gains are explored in broadcast and relay channels.

Broadcast channels with spatial correlation disparity are investigated where the links have non-identical correlations. Both the degrees of freedom and rate results are demonstrated. Results are presented for two cases where CSIR are given and are not free. Applying product superposition and rate splitting, gains are presented in correlation diversity scenarios.

In massive MIMO, in order to reduce the resource for training, this paper proposes transmit strategies using product superposition and rate splitting, which make use of the statistical difference between the channel of different users. Sum rate is calculated to evaluate the performance.

Relay channels with non-identical coherence times are studied. The degrees of freedom are analyzed in the following scenarios: difference coherence time configurations and with multiple parallel relay nodes.

REFERENCES

[1] A. Goldsmith, *Wireless communications.* Cambridge university press, 2005.

[2] D. Tse and P. Viswanath, *Fundamentals of wireless communication.* Cambridge university press, 2005.

[3] T. S. Rappaport *et al.*, *Wireless communications: principles and practice.* prentice hall PTR New Jersey, 1996, vol. 2.

[4] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 6, pp. 1211–1226, Aug. 2002.

[5] K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, P. Karlsson, and M. Beach, "Modeling of wide-band MIMO radio channels based on NLoS indoor measurements," *IEEE Trans. Veh. Technol.*, vol. 53, no. 3, pp. 655–665, May 2004.

[6] D.-S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502–513, Mar. 2000.

[7] S. A. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1165–1175, Jul. 2004.

[8] E. A. Jorswieck and H. Boche, "Channel capacity and capacity-range of beamforming in MIMO wireless systems under correlated fading with covariance feedback," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1543–1553, Sep. 2004.

[9] A. M. Tulino, A. Lozano, and S. Verdu, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2491–2509, Jul. 2005.

[10] W. Chang, S. Chung, and Y. H. Lee, "Diversity-multiplexing tradeoff in rank-deficient and spatially correlated MIMO channels," in *IEEE International Symposium on Information Theory (ISIT)*, Jul. 2006, pp. 1144–1148.

[11] E. Dall'Anese, A. Assalini, and S. Pupolin, "On the effect of imperfect channel estimation upon the capacity of correlated MIMO fading channels," in *IEEE Vehicular Technology Conference*, Apr. 2009, pp. 1–5.

[12] A. Soysal, "Tightness of capacity bounds in correlated MIMO systems with channel estimation error," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Sep. 2010, pp. 667–671.

[13] W. Zhou, J. Wu, and P. Fan, "High mobility wireless communications with doppler diversity: Fundamental performance limits," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6981–6992, 2015.

[14] A. Ghazal, Y. Yuan, C.-X. Wang, Y. Zhang, Q. Yao, H. Zhou, and W. Duan, "A non-stationary IMT-advanced MIMO channel model for high-mobility wireless communication systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2057–2068, 2016.

[15] N. Sun and J. Wu, "Maximizing spectral efficiency for high mobility systems with imperfect channel state information," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1462–1470, 2014.

[16] I. Chih-Lin, L. J. Greenstein, and R. D. Gitlin, "A microcell/macrocell cellular architecture for low-and high-mobility wireless users," *IEEE J. Sel. Areas Commun.*, vol. 11, no. 6, pp. 885–891, 1993.

[17] J. Zhang, H. Du, P. Zhang, J. Cheng, and L. Yang, "Performance analysis of 5G mobile relay systems for high-speed trains," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 12, pp. 2760–2772, 2020.

[18] S. Chakraborty and D. Sen, "Joint estimation of MCFOs and channel gains for two-way multi-relay systems with high mobility," *IEEE Wireless Communications Letters*, vol. 6, no. 5, pp. 610–613, 2017.

[19] Y. Khattabi and M. Matalgah, "Improved error performance ZFSTD for high mobility relay-based cooperative systems," *Electronics Letters*, vol. 52, no. 4, pp. 323–325, 2016.

[20] H. Ghazzai, T. Bouchoucha, A. Alsharoa, E. Yaacoub, M.-S. Alouini, and T. Y. Al-Naffouri, "Transmit power minimization and base station planning for high-speed trains with multiple moving relays in OFDMA systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 175–187, 2017.

[21] M.-S. Pan, T.-M. Lin, and W.-T. Chen, "An enhanced handover scheme for mobile relays in LTE-A high-speed rail networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 2, pp. 743–756, 2015.

[22] T. Al-Naffouri, M. Sharif, and B. Hassibi, "How much does transmit correlation affect the sum-rate scaling of MIMO Gaussian broadcast channels?" *IEEE Trans. Commun.*, vol. 57, no. 2, pp. 562–572, Feb. 2009.

[23] A. Adhikary, H. S. Dhillon, and G. Caire, "Massive-MIMO meets HetNet: Interference coordination through spatial blanking," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 6, pp. 1171–1186, Jun. 2015.

[24] A. Abdi and M. Kaveh, "A space-time correlation model for multielement antenna systems in mobile fading channels," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 3, pp. 550–560, Apr. 2002.

[25] J. Nam, J. Y. Ahn, A. Adhikary, and G. Caire, "Joint spatial division and multiplexing: Realizing massive MIMO gains with limited channel state information," in *46th Annual Conference on Information Sciences and Systems (CISS)*, Mar. 2012, pp. 1–6.

[26] J. Nam, "Fundamental limits in correlated fading MIMO broadcast channels: Benefits of transmit correlation diversity," in *IEEE International Symposium on Information Theory (ISIT)*, Jun. 2014, pp. 2889–2893.

[27] J. Nam, A. Adhikary, J. Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: Opportunistic beamforming, user grouping and simplified downlink scheduling," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 876–890, Oct. 2014.

[28] J. Nam, G. Caire, and J. Ha, "On the role of transmit correlation diversity in multiuser MIMO systems," *IEEE Trans. Inf. Theory*, vol. 63, no. 1, pp. 336–354, Jan. 2017.

[29] A. Adhikary and G. Caire, "JSDM and multi-cell networks: Handling inter-cell interference through long-term antenna statistics," in *48th Asilomar Conference on Signals, Systems and Computers*, Nov. 2014, pp. 649–655.

[30] A. Adhikary, E. A. Safadi, and G. Caire, "Massive MIMO and inter-tier interference coordination," in *Information Theory and Applications Workshop (ITA)*, Feb. 2014, pp. 1–10.

[31] J.-. Lee, J.-. Ko, and Y.-. Lee, "Effect of transmit correlation on the sum-rate capacity of two-user broadcast channels," *IEEE Trans. Commun.*, vol. 57, no. 9, pp. 2597–2599, Sep. 2009.

[32] J. W. Lee, H. N. Cho, H. J. Park, and Y. H. Lee, "Sum-rate capacity of correlated multi-user MIMO channels," in *Information Theory and Applications Workshop (ITA)*, Jan. 2010, pp. 1–5.

[33] W. Xu, C. Zhao, and Z. Ding, "Limited feedback multiuser scheduling of spatially correlated broadcast channels," *IEEE Trans. Veh. Technol.*, vol. 58, no. 8, pp. 4406–4418, 2009.

[34] V. Raghavan, J. J. Choi, and D. J. Love, "Design guidelines for limited feedback in the spatially correlated broadcast channel," *IEEE Trans. Commun.*, vol. 63, no. 7, pp. 2524–2540, 2015.

[35] J. Choi, V. Raghavan, and D. J. Love, "Limited feedback design for the spatially correlated multi-antenna broadcast channel," in *2013 IEEE Global Communications Conference (GLOBECOM)*, 2013, pp. 3481–3486.

[36] Z. Jiang, A. F. Molisch, G. Caire, and Z. Niu, "Achievable rates of FDD massive MIMO systems with spatial channel correlation," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2868–2882, 2015.

[37] F. Zhang, M. Fadel, and A. Nosratinia, "Spatially correlated MIMO broadcast channel: Analysis of overlapping correlation eigenspaces," in *IEEE International Symposium on Information Theory (ISIT)*, Jun. 2017, pp. 1097–1101.

[38] F. Zhang and A. Nosratinia, "Spatially correlated MIMO broadcast channel with partially overlapping correlation eigenspaces," in *2018 IEEE International Symposium on Information Theory (ISIT)*, Jun. 2018, pp. 1520–1524.

[39] K. Ngo, S. Yang, and M. Guillaud, "An achievable DoF region for the two-user non-coherent MIMO broadcast channel with statistical CSI," in *2017 IEEE Information Theory Workshop (ITW)*, Nov. 2017, pp. 604–608.

[40] I. Krikidis, J. Thompson, S. McLaughlin, and N. Goertz, "Optimization issues for cooperative amplify-and-forward systems over block-fading channels," *IEEE Trans. Veh. Technol.*, vol. 57, no. 5, pp. 2868–2884, 2008.

[41] I. Krikidis, H. A. Suraweera, S. Yang, and K. Berberidis, "Full-duplex relaying over block fading channel: A diversity perspective," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4524–4535, Dec. 2012.

[42] K. Hu, Q. Gao, Z. Wang, and W. Huang, "Outage performance of two-way decode-and-forward relaying over block fading channels," in *2017 IEEE 9th International Conference on Communication Software and Networks (ICCSN)*, May 2017, pp. 51–55.

[43] V. Jamali, N. Zlatanov, H. Shoukry, and R. Schober, "Achievable rate of the half-duplex multi-hop buffer-aided relay channel with block fading," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6240–6256, Nov. 2015.

[44] M. Shaqfeh, A. Zafar, H. Alnuweiri, and M. Alouini, "Maximizing expected achievable rates for block-fading buffer-aided relay channels," *IEEE Transactions on Wireless Communications*, vol. 15, no. 9, pp. 5919–5931, Sep. 2016.

[45] R. Simoni, V. Jamali, N. Zlatanov, R. Schober, L. Pierucci, and R. Fantacci, "Buffer-aided diamond relay network with block fading and inter-relay interference," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7357–7372, Nov. 2016.

[46] V. Jamali, N. Zlatanov, A. Ikhlef, and R. Schober, "Achievable rate region of the bidirectional buffer-aided relay channel with block fading," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7090–7111, Nov. 2014.

[47] B. Wang, J. Zhang, and A. Host-Madsen, “On the capacity of MIMO relay channels,” *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 29–43, 2005.

[48] Y. Fan and J. Thompson, “MIMO configurations for relay channels: Theory and practice,” *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1774–1786, 2007.

[49] H. Bolcskei, R. U. Nabar, O. Oyman, and A. J. Paulraj, “Capacity scaling laws in MIMO relay networks,” *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1433–1444, 2006.

[50] Y. Li and A. Nosratinia, “Product superposition for MIMO broadcast channels,” *IEEE Trans. Inf. Theory*, vol. 58, no. 11, pp. 6839–6852, Nov. 2012.

[51] ——, “Coherent product superposition for downlink multiuser MIMO,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1746–1754, Mar. 2015.

[52] M. Fadel and A. Nosratinia, “Coherence disparity in broadcast and multiple access channels,” *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7383–7401, Dec. 2016.

[53] ——, “MISO broadcast channel under unequal link coherence times and channel state information,” *Entropy*, vol. 22, no. 9, 2020.

[54] ——, “Frequency-selective multiuser downlink channels under mismatched coherence conditions,” *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 2393–2404, 2019.

[55] B. Hassibi and B. M. Hochwald, “How much training is needed in multiple-antenna wireless links?” *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[56] F. Zhang and A. Nosratinia, “The degrees of freedom of MIMO relay under coherence diversity,” in *2019 IEEE International Symposium on Information Theory (ISIT)*, 2019, pp. 1177–1181.

[57] ——, “Coherence diversity dof in MIMO relays: Generalization, transmission schemes, and multi-relay strategies,” in *2021 IEEE International Symposium on Information Theory (ISIT)*, 2021, (Accepted).

[58] L. Zheng and D. N. C. Tse, “Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel,” *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[59] M. Chiani, M. Z. Win, and A. Zanella, “On the capacity of spatially correlated MIMO Rayleigh-fading channels,” *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2363–2371, Oct. 2003.

[60] E. M. Luks, F. Rákóczi, and C. R. Wright, “Some algorithms for nilpotent permutation groups,” *J. Symb. Comput.*, vol. 23, no. 4, pp. 335–354, Apr. 1997. [Online]. Available: http://dx.doi.org/10.1006/jsco.1996.0092

[61] A. El Gamal and Y.-H. Kim, *Network Information Theory.* New York, NY, USA: Cambridge University Press, 2011.

[62] R. Diestel, *Graph Theory: 5th edition*, ser. Springer Graduate Texts in Mathematics. Springer-Verlag, © Reinhard Diestel, 2017.

[63] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends® in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017. [Online]. Available: http://dx.doi.org/10.1561/2000000093

BIOGRAPHICAL SKETCH

Fan Zhang received his BSc degree in Electrical Engineering from Peking University, China, in 2007 and his MSEE degree in Electrical Engineering from The University of Texas at Dallas, US, in 2015. He is a PhD candidate in Electrical Engineering at The University of Texas at Dallas, Richardson, TX, USA. His research interests include information theory and its application in wireless communications.

CURRICULUM VITAE

Fan Zhang

Summary

An electrical engineer with 6+ years of research experience in wireless communication, communication theory and information theory

Key Skills

- Good knowledge of the foundations of wireless communications
- Theoretical performance analysis of communication systems
- Solid background in mathematics, information theory, statistics, signal processing, and random process
- Detection and estimation theory, and its application in wireless communication

Education

The University of Texas at Dallas, Richardson, TX, US
Ph.D in Electrical Engineering August 2015 - Current
Advisor: Prof. Aria Nosratinia

The University of Texas at Dallas, Richardson, TX, US
M.S. in Electrical Engineering August 2013 - May 2015

Peking University, Beijing, China
B.S. in Electrical Engineering September 2007 - July 2011

Publications

1. F. Zhang and A. Nosratinia, "The Impact of Coherence Diversity on MIMO Relays," submitted to IEEE Transactions on Wireless Communications

2. F. Zhang, K.-H. Ngo, S. Yang and A. Nosratinia, "Transmit Correlation Diversity: Generalization, New Techniques, and Improved Bounds," submitted to IEEE Transactions on Information Theory

3. F. Zhang and A. Nosratinia, "Coherence Diversity DoF in MIMO Relays: Generalization, Transmission Schemes, and Multi-Relay Strategies," IEEE International Symposium on Information Theory (ISIT), Melbourne, Victoria, Australia, 2021, Accepted

4. F. Zhang and A. Nosratinia, “The Degrees of Freedom of MIMO Relay under Coherence Diversity,” IEEE International Symposium on Information Theory (ISIT), Paris, France, 2019, pp. 1177-1181

5. F. Zhang and A. Nosratinia, “Spatially Correlated MIMO Broadcast Channel with Partially Overlapping Correlation Eigenspaces,” IEEE International Symposium on Information Theory (ISIT), Vail, CO, 2018, pp. 1520-1524

6. F. Zhang, M. Fadel, and A. Nosratinia, “Spatially correlated MIMO broadcast channel: analysis of overlapping correlation eigenspaces,” IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 2017, pp. 1097-1101