University of Alberta

CHANNEL TRAINING AND DECODING FOR MIMO RELAY NETWORKS

by

Sun Sun

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of

Master of Science
in
Communications

Department of Electrical and Computer Engineering

Fall 2011
Edmonton, Alberta

Abstract

To accomplish coherent relaying schemes in cooperative relay networks, accurate channel state information (CSI) is essential. To get such CSI, channel training is employed in practice. In this thesis project, we perform theoretical analysis on channel training design and training-based decoding for multiple-input-multiple-output (MIMO) relay networks, which is in general very challenging for relay networks.

The objective of channel training for MIMO relay networks is to obtain global CSI at the receiver. To perform training, training scheme design, training code design, training time design, and power allocation, are discussed respectively. Employing obtained channel estimations, two coherent training-based decodings are studied for distributed space-time coding (DSTC) MIMO relay networks: mismatched decoding and matched decoding. For both decodings, the diversity and complexity performance are investigated and compared with each other.

Acknowledgement

I owe my deepest gratitude to my supervisor, Prof. Yindi Jing. This thesis would not have been possible without her encouragement and guidance. For me, it is an honor to have the chance to work with Prof. Jing. I learned from her in diverse ways, especially how to organize thoughts and information when analyzing problems. She led me to develop an understanding of this exciting subject, and inspired me to continue my research.

I would like to show my gratitude to my parents, Shumin Sun and Juan Wang, for their unconditional supports. I am also indebted to my husband, Yaoliang Yu, who gives me endless support in both life and academic research.

Contents

1 Introduction to Cooperative Network and Background **1**

1.1 Wireless Channel . . . 1

1.2 Diversity . . . 3

1.3 Multiple-Antenna System . . . 5

1.4 MIMO Relay Network Model and Review on DSTC . . . 7

1.5 Summary . . . 14

2 Preliminary on Training-Based Multiple-Antenna System **15**

2.1 Channel Training for Single-Antenna System . . . 15

2.2 Channel Training for Multiple-Antenna System . . . 17

2.3 Decoding with Imperfect Channel Estimations . . . 20

2.4 Contribution of This Thesis . . . 21

2.5 Summary . . . 22

3 Channel Training in MIMO Relay Networks **24**

3.1 Introduction and Literature Review . . . 24

3.2 Training of the Relay-Receiver Channel Matrix $\mathbf{G}$ . . . 25

3.3 Training of the Transmitter-Relay Channel Vector $\mathbf{f}$ with Perfect $\mathbf{G}$ . . . 26

3.3.1 Pilot Design . . . 27

3.3.2 Lower Bound on the Training Time . . . 30

3.4 Training of the Transmitter-Relay Channel Vector $\mathbf{f}$ with Estimated $\mathbf{G}$ . . . 30

3.4.1 Pilot Design . . . 33

3.4.2 Lower Bound on the Training Time . . . 33

3.4.3 Estimation Error on Quality of $\mathbf{f}$ . . . 34

3.4.4 Discussion on Extension . . . 36

3.5 Training of the End-to-End Channel Matrix $\mathbf{H}$ . . . 36

3.5.1 Separate Training of $\mathbf{H}$ . . . 37

3.5.2 Direct Training of $\mathbf{H}$ . . . 37

3.5.3 Comparison on Separate Training and Direct Training . . . 39

3.6 Discussions . . . 40

3.6.1 Discussion on Channel Correlation . . . 40

3.6.2 Discussion on Applications . . . 41

3.7 Summary . . . 41

3.8 Appendixes . . . 42

3.8.1 Proof of Theorem 3.1 . . . 42

3.8.2 Proof of Theorem 3.3 . . . 42

3.8.3 Proof of Theorem 3.4 . . . 46

4 Training-Based Mismatched Decoding and Diversity Analysis **48**

4.1 System Model . . . 48

4.2 Mismatched Decoding with Perfect $\mathbf{G}$ . . . 50

4.2.1 Diversity Analysis . . . 50

4.2.2 Adaptive Training Time Design . . . 53

4.2.3 Simulation Results . . . 55

4.3 Mismatched Decoding with Imperfect $\mathbf{G}$ . . . 59

4.3.1 Diversity Analysis . . . 60

4.3.2 Simulation Results . . . 62

4.4 Summary 64
4.5 Appendices 65
4.5.1 Useful Results on Wishart Matrix 65
4.5.2 Proof of Lemma 4.1 66
4.5.3 Proof of Theorem 4.2 68
4.5.4 Proof of Theorem 4.3 73
5 Training-Based Matched Decoding and Diversity Analysis **76**
5.1 Matched Decoding 76
5.2 Adaptive Decoding 78
5.3 Complexity Analysis 79
5.4 Simulation Results 82
5.5 Summary 86
6 Summary and Future Work **87**
References **90**

List of Tables

5.1 Numbers of flops in coherent decodings with $R = 2$ 81

5.2 Numbers of flops in coherent decodings with $R = 3$ 81

List of Figures

1.1 Multiple-antenna system. 5

1.2 Cooperative network with three nodes. 8

1.3 MIMO relay network. 9

1.4 DEC_0 vs $\text{DEC}_{0,simp}$ for network with $M = 1, R = N = 2$. . . 13

2.1 Single-antenna system. 16

3.1 $\text{MSE}(\hat{\mathbf{f}})$ under perfect $\mathbf{G}$ and estimated $\mathbf{G}$ for the network with $R = 2$. 35

4.1 Comparison of training code designs. Network 1: $M = 1, R = N = 2$; Network 2: $M = 2, N = R = 2$. 55

4.2 Comparison of direct training, separate training, and the perfect CSI case. Network 1: $M = N = 1, R = 2$; Network 2: $M = 1, R = N = 2$; Network 3: $M = R = N = 2$. 56

4.3 BLER performance of the training-based scheme with $R \leq N$ and $N_p = N_{p,l}$. 57

4.4 Adaptive-N_p design for network with $M = 1, N = R = 2$. . . 58

4.5 Adaptive-N_p design for network with $M = R = 2, N = 3$. . . . 59

4.6 DEC_1 for the network with $M = 1, R = N = 2$. 63

4.7 DEC_1 for the network with $M = 1, R = N = 3$. 63

5.1 Mismatched decoding, matched decoding, and perfect CSI decoding for the network with $M = 1, R = N = 2$. 82

5.2 Mismatched decoding, matched decoding, and perfect CSI decoding for the network with $M = 1, R = N = 3$. 83

5.3 Adaptive decoding for the network with $R = 2$. 84

5.4 Adaptive decoding for the network with $R = 3$. 85

List of Symbols

$\overline{\mathbf{A}}$	conjugate of $\mathbf{A}$
$\mathbf{A}^t$	transpose of $\mathbf{A}$
$\mathbf{A}^*$	conjugate transpose of $\mathbf{A}$
$\mathbf{A}^{-1}$	inverse of $\mathbf{A}$
$\mathbf{A}^\perp$	unitary complement of $\mathbf{A}$
	If $\mathbf{A}$ is $m \times n$, $\mathbf{A}^\perp$ is the $m \times (m-n)$ matrix such that $[\mathbf{A}\ \mathbf{A}^\perp]$ is unitary.
$\overrightarrow{\mathbf{A}}$	column vectorization of $\mathbf{A}$
$\mathbf{a}_i$	the i-th component of vector $\mathbf{a}$
$\mathbf{A}_{ij}$	the component of matrix $\mathbf{A}$ in the i-th row j-th column
$\|\mathbf{A}\|_F$	Frobenius norm of $\mathbf{A}$
$\mathrm{tr}(\mathbf{A})$	trace of $\mathbf{A}$
$\det(\mathbf{A})$	determinant of $\mathbf{A}$
$\mathrm{diag}\{\mathbf{a}_1, \cdots, \mathbf{a}_n\}$	block diagonal matrix with i-th diagonal block $\mathbf{a}_i$
$\mathbf{A} \succ \mathbf{B}$	$\mathbf{A} - \mathbf{B}$ is a positive definite matrix
$\mathbb{I}_R$	$R \times R$ identity matrix
$\mathbf{0}_{m,n}$	$m \times n$ matrix with all zero entries
$\otimes$	Kronecker product
mod	modulo operation
	When $a \bmod b = 0$, redefine it as $a \bmod b = b$.
$\mathbb{P}$	probability
$\mathbb{E}\ (\cdot)$	expectation
$\lceil x \rceil$	smallest integer that is larger than x
$\lfloor x \rfloor$	largest integer that is smaller than x
$\lvert x \rvert$	absolute value of x
$\Re(x)$	real part of x

$\min(x, y)$	the smaller one of x and y
$\max(x, y)$	the bigger one of x and y
$\|\mathcal{C}\|$	cardinality of the set $\mathcal{C}$
$O(x)$	order of x
$o(x)$	lower order of x
$\mathcal{CN}(c, \sigma^2)$	complex Gaussian distributed with mean c and variance σ^2

List of Abbreviations

AF	amplify-and-forward
AWGN	additive white Gaussian noise
BER	bit error rate
BLER	block error rate
BLUE	best linear unbiased estimator
CDF	cumulative distribution function
CSI	channel state information
D-BLAST	diagonal Bell-labs layered space-time architecture
DF	decode-and-forward
DSTC	distributed space-time coding
i.i.d.	independent and identically distributed
LMMSE	linear minimum mean square error
LSE	least squares estimator
MI	mutual information
MIMO	multiple-input-multiple-output
ML	maximum likelihood
MRC	maximum ratio combiner
MSE	mean square error
OSTBC	orthogonal space-time block code
PDF	probability density function
SNR	signal-to-noise ratio
TDMA	time division multiple access
V-BLAST	vertical Bell-labs layered space-time architecture

1 Introduction to Cooperative Network and Background

1.1 Wireless Channel

Wired channel communication enjoys reliable transmission due to the stable channel condition between terminals. The drawback is, its application is quite limited in geography. In contrast, the applications of wireless channel communication, such as cellphone, allow more freedom in communication, and become more and more popular. Since the appearance of the first generation cellular phone in early 1980s, the interest in communication with wireless channel has been fuelled and a lot of researches have been started.

The characterization of the wireless channel is quite different from the characterization of the wired channel. Due to the multi-path propagation that arises from direct transmission, reflection, diffraction, and scattering between the transmitter and the receiver, multiple versions of the transmit signal are created at the receiver, which can decrease the received power in different ways. Two of these that deserve special treatments are large-scale fading and small-scale fading [1].

- Large-scale fading: It corresponds to the degradation of the signal power over a large distance or the time-average behaviour of the signal.

  Under this model, the received power is usually modelled by

  $$P_r = \beta d^{-v} P_t, \tag{1.1}$$

  where P_r is the received power, P_t is the transmitted power, β is the coefficient related to the carrier frequency and other factors, and v is the path loss exponent typically ranging from 2 to 6.

- Small-scale fading: It corresponds to the characterization of the signal over a short distance or a short time interval.

Denote the bandwidth of the signal as B_s. Also, denote the coherence bandwidth of the channel as B_c, over which the channel can be considered as flat. The small-scale fading channels can then be classified as:

- Flat fading ($B_s < B_c$);
- Frequency selective fading ($B_s > B_c$).

In this work, we only consider flat fading, which characterizes the narrowband wireless channels in indoor and urban areas. More specifically, the block-wise flat fading channel model is adopted, where the channel is assumed to be invariant in one transmission block and change independently the next block. Such model is widely used in time division multiple access (TDMA) or frequency-hopping systems [2].

To describe the amplitude of the received signal through flat fading channel, one popular statistical model is Rayleigh fading model [1], which is used in this thesis. In the Rayleigh fading model, the channel amplitude follows Rayleigh distribution, i.e.,

$$p(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}}, \quad r \geq 0, \tag{1.2}$$

and the channel response, or the channel coefficient, is described as Gaussian with zero-mean. This channel model is suitable for the richly scattered environment without dominant propagation path. When such dominant path exists between the transmitter and receiver, the Ricean fading model [1] can be applied, where the channel amplitude follows Ricean distribution, i.e.,

$$p(r) = \frac{r}{\sigma^2} e^{\frac{-(r^2+D^2)}{2\sigma^2}} I_0\left(\frac{Dr}{\sigma^2}\right), \quad r \geq 0, D \geq 0, \tag{1.3}$$

with D the peak amplitude of the dominant signal and $I_0(.)$ the modified Bessel function of the first kind and of zero-order, and the channel coefficient is described as Gaussian with non-zero mean.

1.2 Diversity

In wired channel communication, since the channel between terminals is stable, the transmission can be performed reliably. When a white Gaussian noise is added at the receiver, described as the additive white Gaussian noise (AWGN) channel model, the bit error rate (BER) is shown to decrease exponentially with the average signal-to-noise ratio (SNR), i.e., e^{-SNR} [3]. In wireless communication, however, when the block-wise Rayleigh flat fading channel is considered, the BER is merely proportional to the inverse of the average SNR, i.e., SNR^{-1} [3]. The reason for such significant degradation in BER is, there is a high probability that the channel is in deep fade.

To combat the fading effect and hence improve the reliability of wireless communication, the technology of diversity is suggested. The idea of diversity is to create different versions of transmit signal at the receiver, which go through independent fading. Since the corresponding channels fade independently, the probability that the entire channels in deep fading drops significantly. Therefore, the signal can be conveyed more reliably. For example, if M such replicas of transmit signal are collected, the order of the error probability drops to SNR^{-M}, compared to SNR^{-1} for the case where only a single replica is received. The achieved gain, reflected by the exponent of the SNR in the reliability scaling, by employing diversity technology is called diversity gain.

There are several schemes to achieve diversity. In the following, three popular diversity schemes are listed and described.

- Temporal diversity;
- Frequency diversity;
- Spatial diversity, or antenna diversity.

Denote the coherence time of the channel as T_c, during which the channel keeps roughly unchanged. To achieve temporal diversity, identical signals are sent over different time slots, which are separated more than the coherence time

to ensure independent fading. Similarly, to obtain frequency diversity, the identical signals are sent over different frequencies that are separated more than the coherence bandwidth of the channel. Due to the redundancy in transmission, these two schemes are, however, inefficient in bandwidth.

Spatial diversity is a promising technology in that it does not suffer from the problem of bandwidth inefficiency. Compared to the former two schemes, it utilizes spatial dimension to create diversity by sending or receiving replicas of signal over multiple antennas. Based on the positions where the antennas are located, spatial diversity is categorized into transmit diversity and receive diversity. To guarantee independent fading, the antennas should be separated more than half of the wavelength. For the large infrastructure like the base station, multiple antennas can be deployed to gain spatial diversity. Nevertheless, for small device like cellphone, due to its limitations in both size and power, and considerations such as the cost, the deployment of multiple antennas is sometimes impractical. Therefore, in the cellular communication system, by deploying multiple antennas in the base station, we can achieve transmit diversity in the downlink (from base station to cellphone), and receive diversity in the uplink (from cellphone to base station), respectively. In this work, spatial diversity is adopted to combat fading. Different spatial diversity schemes for multiple-antenna system will be introduced in the next subsection.

To quantitively measure diversity, one commonly used definition of diversity order, or diversity, is [1]

$$G_d = -\lim_{SNR\to\infty} \frac{\log \mathbb{P}_e}{\log SNR}, \tag{1.4}$$

where $\mathbb{P}_e$ denotes the error probability. From this definition, we know that diversity typically describes the log-log relationship between the error rate and SNR at the high SNR region.

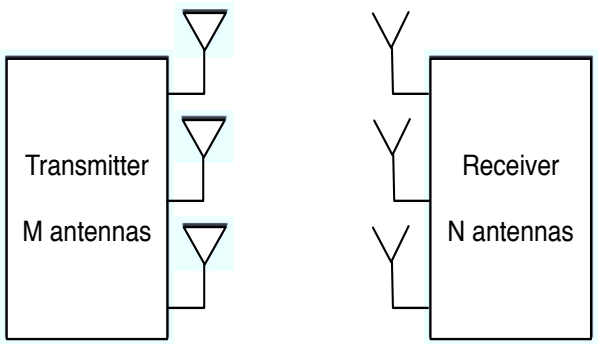

Figure 1.1: Multiple-antenna system.

1.3 Multiple-Antenna System

The demand for higher data rate and lower error probability in wireless communication system never stops. Compared to the single-antenna system, multiple-antenna system [4] attracts widespread attentions in both industrial and academic areas, for it can achieve considerable performance improvement in these aspects. Besides, it can also be employed to suppress co-channel interference, which is one of the biggest challenges in multi-user cellular wireless communication system.

A typical multiple-input-multiple-output (MIMO) system is shown in Fig. 1.1, where multiple antennas are equipped at both the transmitter and the receiver. Over the last two decades, aiming at different objectives, several novel techniques were proposed in multiple-antenna system which can be mainly divided into three categories [5]:

- Spatial multiplexing;
- Smart antennas;
- Spatial diversity.

Denote the transmitter-side channel state information (CSI) as Tx-CSI, and the receiver-side CSI as Rx-CSI. To complete these schemes, different types of CSI are required: To perform spatial multiplexing and spatial diversity,

the Rx-CSI is needed; while to apply smart antennas, both the Tx-CSI and Rx-CSI are required. In the following, we first give brief introductions to spatial multiplexing and smart antennas, then focus on the schemes of spatial diversity that are adopted in this work.

Spatial multiplexing aims at high data rate. Assume that there are M transmit antennas. Compared to the single-antenna system, in which a single sequence is sent from the transmitter, by spatial multiplexing, M independent sequences can be sent simultaneously from the transmit antennas which results in M-fold increase in data rate. Examples of spatial multiplexing technique are the well-known vertical Bell-Labs Layered Space-time Architecture (V-BLAST) [6, 7] and diagonal Bell-Labs Layered Space-time Architecture (D-BLAST) [8].

The technique associated with smart antennas is beamforming [9, 10]. The idea of beamforming is to steer the transmit/receive beam pattern to the directions of dominant paths in the multi-path propagation. So, the received SNR can be largely boosted compared to not optimizing the transmit/receive directions. Also, if there are multiple users in the cellular system, beamforming technique can be used to suppress the co-channel interference by nulling the directions where significant interference signals are present.

Spatial diversity is to improve the error rate performance in wireless communication. To implement transmit diversity, several schemes are proposed, e.g. [5]. One that is adopted in this work is (generalized) orthogonal space-time codes (OSTBCs) [1, 11]. There are two important properties associated with (generalized) OSTBCs. First, they can achieve full diversity with respect to the numbers of transmit antennas and receive antennas. Second, they can provide separate maximum likelihood (ML) detection, whose complexity is linear with the data rate and the number of transmit antennas. Due to these advantages, OSTBCs attract lots of attentions since 1998, when Alamouti code [12] was first introduced for two transmit antennas. In the following, the designing rules for (generalized) OSTBCs with real entries are discussed. The designing

rules when the code is with complex entries are analogous, which can be found in [1]. Consider the OSTBCs with real entries. An $M \times M$ real OSTBC $\mathbf{S}$ with entries $s_1, -s_1, \cdots, s_M, -s_M$ is designed such that

$$\mathbf{S}^t\mathbf{S} = (s_1^2 + \cdots + s_M^2)\mathbb{I}_M. \tag{1.5}$$

To transmit $\mathbf{S}$, the ij-th component of $\mathbf{S}$ is sent at the i-th time slot by the j-th transmit antenna. So, the rate of the code is 1, which is the full rate. However, such real OSTBCs only exist for $M = 2, 4, 8$. For other network settings, an $M \times M$ matrix which satisfies (1.5) cannot be found. Thus, generalized real OSTBCs can be applied. A $T \times M$ generalized real OSTBC $\mathbf{S}$ with entries $s_1, -s_1, \cdots, s_k, -s_k$ is designed such that

$$\mathbf{S}^t\mathbf{S} = c(s_1^2 + \cdots + s_k^2)\mathbb{I}_M, \tag{1.6}$$

where c is a constant. The rate of the code is k/T.

To implement receive diversity, maximum ratio combiner (MRC) [13] can be used to effectively combine the received signals from each path. Suppose that the SNR of the i-th path is SNR_i, and there are M independent paths. Using MRC, the effective SNR can be shown to be $\sum_{i=1}^{M} SNR_i$. So, if each path has the same SNR, by MRC, there is M-fold increase in SNR, which leads to a diversity of M.

1.4 MIMO Relay Network Model and Review on DSTC

In the previous section, it was shown that multiple-antenna system can have much better performance compared to single-antenna system in several aspects. However, due to the limitations of both the cost and size, not all devices can afford the deployment of multiple antennas and enjoy the associated benefits. This motivates the idea of cooperative network [14–17], in which a virtual multiple-antenna system can be created by exploiting the pos-

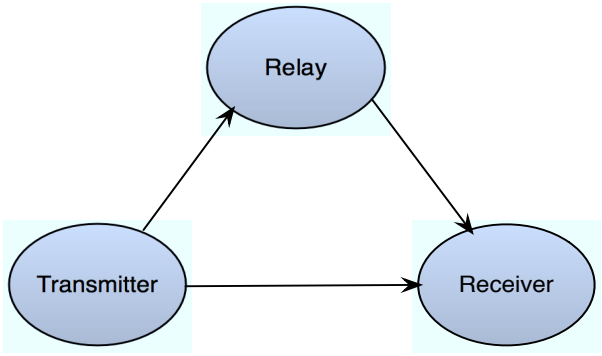

Figure 1.2: Cooperative network with three nodes.

sible cooperation among the distributed nodes in the network, regardless of the number of antennas equipped at the nodes.

To illustrate the basic idea behind cooperative network, let us look at the network in Fig. 1.2, where the relay is appropriately assigned to help the communication. Assume each node is equipped with a single antenna which can transmit and receive signals. By sending the information from the transmitter to the receiver directly, there is no diversity achieved and the received signal power could be decreased significantly due to the fading. However, with the help of the relay, i.e., a replica of the intended signal is sent by the relay to the receiver, spatial diversity can be achieved provided that, the signals from the transmitter and the relay go through independent fading.

To make the relay help the transmission, the transmitter needs to send a copy of the signal to the relay. According to the operation that the relay takes, the network can be categorized into decode-and-forward (DF) relay network and amplify-and-forward (AF) relay network.

In DF relay network [17], after receiving the signal from the transmitter, the relay will decode it and retransmit it to the receiver. This transmission strategy actually splits the relay network into two systems, one of which is the transmitter-relay system and the other of which is the relay-receiver system. Since both systems are essentially multiple-antenna systems, many analysis taken in the multiple-antenna system can be applied directly in DF relay

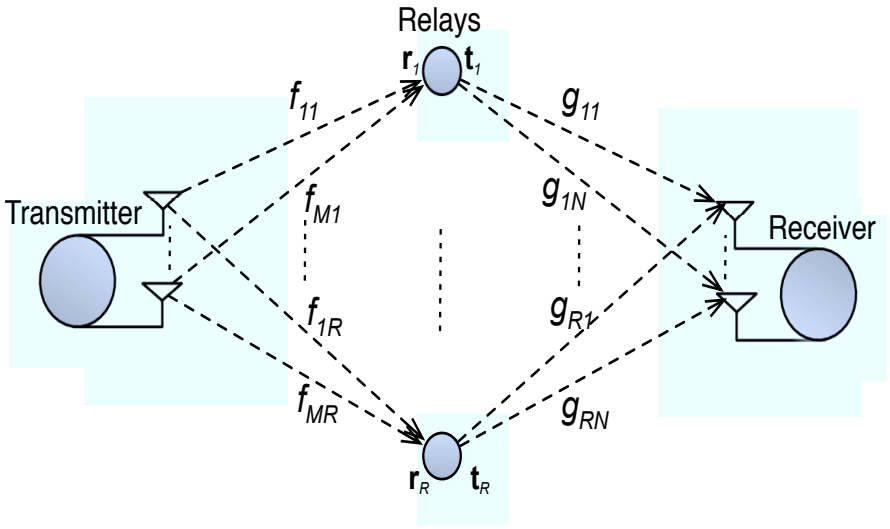

Figure 1.3: MIMO relay network.

network.

In AF relay network [17], the relay amplifies what it receives and forwards the amplified signal to the receiver. Note that, although the transmitted signal by the relay is a noisy version of the intended signal, it can help the receiver achieve a better performance if it goes through independent fading. Compared to DF relay network, no decoding is required at the relay in AF relay network. So, AF relay network is more attractive since at the relay, the processing is simpler, and less resource is consumed.

In the above, we briefly described two transmission strategies in a basic sing-relay single-antenna network. In practice, multiple relays can be assigned to help the communication. It is also possible that multiple antennas are equipped at the transmitter/relay/receiver nodes, which forms a MIMO relay network, see Fig. 1.3. Several transmission strategies are proposed in such relay networks [18–37]. For example,

- Relay selection, e.g. [18–20];
- Cooperative beamforming, e.g. [21, 22];
- Distributed space-time coding (DSTC), e.g. [23–37].

The main challenging of the investigation on MIMO relay network over that on sing-relay single-antenna network stems from the curse of dimensionality. As discussed in [25], the existence of multiple transmit and receive antennas can largely complicate the performance analysis.

In this thesis, DSTC is employed for MIMO relay network. This scheme can be used for networks with both AF and DF relays. The DSTC scheme proposed in [24,25] is for networks with any number of non-regenerative relays and transmit/receive antennas. The key idea is to design the transmit signal of each relay antenna as a linear function of its received signal and its conjugate, so that a linear space-time codeword is formed at the receiver. The scheme is proved to achieve the optimal diversity. In addition, it has the advantages of simple relay signal processing and no channel information requirement at the relays. The differential use of DSTC in non-coherent relay networks is introduced in [30–32]. Its diversity-multiplexing tradeoff is analyzed in [34]. Its use in asynchronous networks is discussed in [36, 37]. In [11, 33], revised DSTC schemes are proposed for networks with partial channel information at relays. Specific distributed space-time code designs are provided in [11,27,28].

We now give a review on transmission protocol and decoding metric for DSTC network with perfect CSI. Consider a wireless relay network with M antennas at the transmitter, N antennas at the receiver, and $R(R \geq 2)$ relays each with a single antenna, as in Fig. 1.3. Perfect synchronization among nodes is assumed. Denote the channel vector from the transmitter to the ith relay as $\mathbf{f}_i = [f_{1i} \cdots f_{Mi}]^t$, and the channel vector from the ith relay to the receiver as $\mathbf{g}_i = [g_{i1} \cdots g_{iN}]$. Let $\mathbf{f} = [\mathbf{f}_1^t \cdots \mathbf{f}_R^t]^t$, which is the $MR \times 1$ vector of the transmitter-relay channels, and $\mathbf{G} = [\mathbf{g}_1^t \cdots \mathbf{g}_R^t]^t$, which is the $R \times N$ matrix of the relay-receiver channels. The $MR \times N$ end-to-end channel matrix is thus

$$\mathbf{H} = [(\mathbf{f}_1\mathbf{g}_1)^t \cdots (\mathbf{f}_R\mathbf{g}_R)^t]^t = \begin{bmatrix} f_{11}g_{11} & f_{11}g_{12} & \cdots & f_{11}g_{1N} \\ \vdots & \vdots & \ddots & \vdots \\ f_{M1}g_{11} & f_{M1}g_{12} & \cdots & f_{M1}g_{1N} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ f_{1R}g_{R1} & f_{1R}g_{R2} & \cdots & f_{1R}g_{RN} \\ \vdots & \vdots & \ddots & \vdots \\ f_{MR}g_{R1} & f_{MR}g_{R2} & \cdots & f_{MR}g_{RN} \end{bmatrix}, \quad (1.7)$$

with the kl-th entry the end-to-end channel from the (k mod M)-th transmit antenna to the l-th receive antenna via the $\lceil \frac{k}{M} \rceil$-th relay. All channel coefficients, i.e., the entries of $\mathbf{f}$ and $\mathbf{G}$, are assumed to be *i.i.d.* block-wise Rayleigh flat fading with the distribution $\mathcal{CN}(0,1)$. All noises at each node are assumed as *i.i.d.* Gaussian following $\mathcal{CN}(0,1)$. Denote the transmit powers at the transmitter and each relay as P and P_1 respectively.

In the two-step DSTC protocol in [25], the information is encoded into a $T \times M$ matrix $\mathbf{B}$, with the constraint $\mathbb{E}\ (\mathrm{tr}(\mathbf{B}^*\mathbf{B})) = M$. To send $\mathbf{B}$ from the transmitter to the receiver, $2T$ symbol periods are dedicated with T symbol periods in each step. For the first step, $\mathbf{B}$ is sent by the transmitter. Denote the signal received at the ith relay by $\mathbf{r}_i$, we have

$$\mathbf{r}_i = \sqrt{\frac{PT}{M}}\mathbf{B}\mathbf{f}_i + \mathbf{v}_i, \quad (1.8)$$

where $\mathbf{v}_i$ is the noise at the ith relay. In the second step, Relay i sends $\mathbf{t}_i$, which is designed as a linear function of its received signal, i.e., $\mathbf{t}_i = \alpha\mathbf{A}_i\mathbf{r}_i$, where $\alpha \triangleq \sqrt{\frac{P_1}{1+P}}$, and $\mathbf{A}_i$ is a pre-designed $T \times T$ unitary matrix. For the simplicity of the presentation, we also assume that $P = RP_1$, which is shown to be optimal in the sense of maximizing the received SNR [25]. Hence, $\alpha = \sqrt{\frac{P}{R(P+1)}}$.

The received matrix at the receiver can be calculated to be

$$\mathbf{X} = \beta \mathbf{S}\mathbf{H} + \mathbf{W}, \tag{1.9}$$

where $\beta \triangleq \sqrt{\frac{P^2 T}{MR(P+1)}}$, $\mathbf{S} = [\mathbf{A}_1\mathbf{B} \cdots \mathbf{A}_R\mathbf{B}]$ is the distributed space-time codeword formed at the receiver, $\mathbf{H}$ is the $MR \times N$ end-to-end channel matrix in (1.7), and

$$\mathbf{W} = \alpha \left[\mathbf{A}_1\mathbf{v}_1 \cdots \mathbf{A}_R\mathbf{v}_R\right] \mathbf{G} + \mathbf{W}_{Rx} \tag{1.10}$$

is the equivalent noise matrix with $\mathbf{W}_{Rx}$ the noise matrix at the receiver. The covariance matrix of $\mathbf{W}$ can be derived to be $\mathbf{R}_{\mathbf{W}} = \mathbb{I}_N + \alpha^2 \overline{\mathbf{G}^* \mathbf{G}}$.

With perfect CSI of both the relay-receiver channel matrix $\mathbf{G}$ and the end-to-end channel matrix $\mathbf{H}$, the ML decoding is

$$DEC_0 : \arg \min_{\mathbf{B}} \operatorname{tr}\Big(\mathbf{X} - \beta\mathbf{S}\mathbf{H}\Big) \overline{\mathbf{R}}_{\mathbf{W}}^{-1} \Big(\mathbf{X} - \beta\mathbf{S}\mathbf{H}\Big)^*. \tag{1.11}$$

Stacking the columns of $\mathbf{X}$ into one column vector, from (1.9), we have

$$\overrightarrow{\mathbf{X}} = \beta(\mathbf{G}^t \otimes \mathbb{I}_T)\tilde{\mathbf{S}}\mathbf{f} + \overrightarrow{\mathbf{W}} = \beta \mathbf{Z}\mathbf{f} + \overrightarrow{\mathbf{W}}, \tag{1.12}$$

where

$$\tilde{\mathbf{S}} \triangleq \operatorname{diag}\{\mathbf{A}_1\mathbf{B}, \cdots, \mathbf{A}_R\mathbf{B}\}, \tag{1.13}$$

an equivalent format for the distributed space-time codeword, and $\mathbf{Z} \triangleq (\mathbf{G}^t \otimes \mathbb{I}_T)\tilde{\mathbf{S}}$. From (1.12), the ML decoding with perfect $\mathbf{G}$ and $\mathbf{f}$ can be rewritten as:

$$DEC_0 : \arg \min_{\mathbf{B}} \left(\overrightarrow{\mathbf{X}} - \beta\mathbf{Z}\mathbf{f}\right)^* \mathbf{R}_{\overrightarrow{\mathbf{W}}}^{-1} \left(\overrightarrow{\mathbf{X}} - \beta\mathbf{Z}\mathbf{f}\right), \tag{1.14}$$

where $\mathbf{R}_{\overrightarrow{\mathbf{W}}} \triangleq \mathbf{R}_{\mathbf{W}} \otimes \mathbb{I}_T$.

To use (1.11) for decoding, the knowledge of $\mathbf{G}$ and $\mathbf{H}$ is required at the receiver; while to use (1.14), the knowledge of $\mathbf{G}$ and $\mathbf{f}$ is needed. It is note-

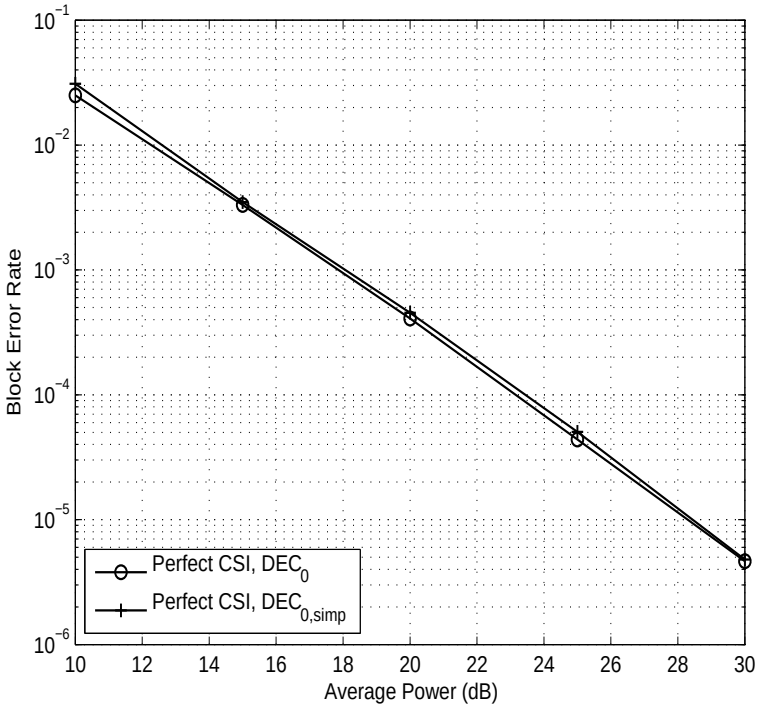

Figure 1.4: DEC_0 vs $\mathrm{DEC}_{0,simp}$ for network with $M = 1, R = N = 2$.

worthy that the decoding rules in (1.11) and (1.14) are equivalent although with different representations. It is proved in [25] that with the above ML decoding, full diversity can be achieved. The decoding complexity, however, is high since it requires the joint decoding of all the information symbols in $\mathbf{B}$ even with orthogonal design of $\mathbf{S}$ such that $\mathbf{S}^*\mathbf{S} = \mathbb{I}_{MR}$ [11].

To reduce the complexity, a simplification of DEC_0 is proposed as

$$\mathrm{DEC}_{0,simp} : \arg\max_{\mathbf{B}} \Re\mathrm{tr}(\mathbf{H}\mathbf{X}^*\mathbf{S}), \tag{1.15}$$

which can be derived straightforwardly from either (1.11) or (1.14) by replacing $\mathbf{G}^t\overline{\mathbf{G}}$ with its expectation $R\mathbb{I}_N$ in $\mathbf{R}_{\mathbf{W}}$ and considering the orthogonal structure of $\mathbf{S}$. This decoding can be performed symbol-by-symbol [1], thus has much lower complexity. Simulation shows that $\mathrm{DEC}_{0,simp}$ performs almost the same as the optimal decoding DEC_0 in BLER performance. As an example, in

Fig. 1.4, the BLER performance of DEC_0 and $\mathrm{DEC}_{0,simp}$ for the network with $M = 1, R = N = 2$ are shown. Alamouti code with BPSK modulated information symbols is applied for the data transmission [11]. We see that $\mathrm{DEC}_{0,simp}$ has almost the same behaviour as DEC_0.

As a special case, we consider the network with $M = 1$, and $R = N \geq 2$. Since the signal sent by the transmitter is now a vector with the dimension $T \times 1$, a lowercase variable $\mathbf{s} \triangleq [\mathbf{s}_1 \cdots \mathbf{s}_T]$ is adopted instead of $\mathbf{B}$. Let $\mathbf{T} \triangleq \mathbf{H}\mathbf{X}^*[\mathbf{A}_1 \cdots \mathbf{A}_R]$, which is $R \times RT$. Decompose $\mathbf{T}$ into blocks of size $1 \times T$ and denote the block of the i-th row and the j-th column as $\mathbf{T}_{b,ij}$ for $i, j = 1, \cdots, R$. It can be shown with straightforward algebra that the simplified decoding is:

$$\mathrm{DEC}_{0,simp} : \arg \max_{s_j} \Re\, s_j \left(\sum_{i=1}^{R} \mathbf{T}_{b,ii} \right)_j, \quad j = 1, \cdots, T. \tag{1.16}$$

1.5 Summary

The applications of wireless communication, such as cellphone, are very popular nowadays, and can be seen almost everywhere. In contrast to the stable channel condition in wired communication, the wireless channel, however, suffers from fading due to the multi-path propagation. To improve reliability of wireless communication, the technique called diversity was suggested, and one efficient way to create diversity is through multiple-antenna system. In addition to providing diversity and hence to improving error rate performance, multiple-antenna system can also increase data rate and suppress co-channel interference. However, due to the limitation of resources, not all devices can afford the deployment of multiple antennas. The cooperative network is then attractive, for it can create a virtual multiple-antenna system by exploiting the cooperation of nodes in the network, regardless of the number of antennas equipped at each node. One cooperative scheme, called DSTC, was reviewed.

2 Preliminary on Training-Based Multiple-Antenna System

To implement DSTC in cooperative network, the Rx-CSI is required. To get such CSI, one commonly adopted method is channel training, where known pilot signal is sent to track the channels. In this chapter, we provide preliminary on channel training for multiple-antenna system. The reason for including this is two-fold: First, such topic is well studied, e.g. [2, 38–46]; second, there are some similarities between channel training for cooperative system and that for multiple-antenna system, which helps our investigation later, although the former is much more difficult and challenging.

In the following discussions, assume that the channel coefficients are Gaussian following $\mathcal{CN}(0, \rho^2)$, and all noises are Gaussian following $\mathcal{CN}(0, \sigma^2)$[1]. To perform channel training, the coherence time T_c is split into two phases: training phase and data transmission phase. In Section 2.1, we first illustrate the idea of channel training by describing channel training for single-antenna system, which is straightforward. Then, in Section 2.2, channel training for multiple-antenna system is presented. In Section 2.3, decoding with imperfect channel estimations for multiple-antenna system is discussed. In Section 2.4, the contribution of this thesis is included.

2.1 Channel Training for Single-Antenna System

Consider a single-antenna system in Fig. 2.1, where the receiver needs to know the channel coefficient h. In training phase, a power constraint pilot s_p is sent from the transmitter to track the channel, where the subscript "p" is used to indicate training phase. Since s_p is a scalar, without loss of generality, let $s_p = 1$. At the receiver, a noisy observation of the pilot is obtained:

$$x_p = h + w_p, \tag{2.1}$$

[1]Except in Chapter 2, ρ and σ are considered to be 1.

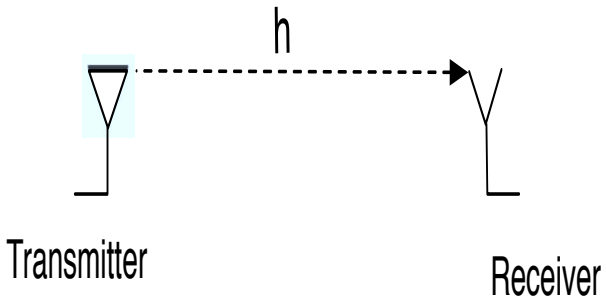

Figure 2.1: Single-antenna system.

where w_p is the noise. Denote the estimation of h as $\hat{h}$, and the estimation error as $\Delta h \triangleq h - \hat{h}$. To get $\hat{h}$, the mean square error (MSE) of $\hat{h}$ is employed as the criterion, which is defined as

$$\text{MSE}(\hat{h}) \triangleq \mathbb{E}_{h,w_p}(|\Delta h|^2). \tag{2.2}$$

By minimizing MSE($\hat{h}$), the minimum MSE (MMSE) estimation of h is $\hat{h} = \mathbb{E}(h|x_p)$. Generally, $\mathbb{E}(h|x_p)$ is a non-linear function of x_p. But, under the condition that h and x_p are jointly Gaussian, $\mathbb{E}(h|x_p)$ is shown to be linear with x_p [52]. So, the MMSE estimation of h coincides with the linear MMSE (LMMSE) estimation, and is derived as

$$\hat{h} = \mathbb{E}(h|x_p) = \frac{\rho^2}{\rho^2 + \sigma^2} x_p. \tag{2.3}$$

In data transmission phase, the estimation of h in (2.3) is employed. Suppose that s_d is transmitted, where the subscript "d" is used to indicate data transmission phase. Given $\hat{h}$, the received signal can be represented as

$$x_d = s_d\hat{h} + s_d\Delta h + w_p = s_d\hat{h} + w_p', \tag{2.4}$$

where $w_p' \triangleq s_d\Delta h + w_p$ is the equivalent noise including both the estimation error and the noise.

2.2 Channel Training for Multiple-Antenna System

Channel training for single-antenna system, as we have seen, is straightforward since all variables are scalars (only the temporal dimension is induced). For multiple-antenna system, where both the spatial and temporal dimensions are induced, however, more efforts are needed for channel training.

Consider a point-to-point multiple-antenna system with M transmit antennas and N receive antennas. Denote the MIMO channels as an $M \times N$ matrix $\mathbf{H}$,

$$\mathbf{H} = \begin{bmatrix} h_{11} & \cdots & h_{1N} \\ \vdots & \ddots & \vdots \\ h_{M1} & \cdots & h_{MN} \end{bmatrix}, \tag{2.5}$$

where h_{ij} is the channel coefficient from the ith transmit antenna to the jth receive antenna. Assume that all h_{ij}'s are independent.

Decompose $T_c = T_p + T_d$, where T_p is the number of symbol intervals dedicated for training, and T_d is the number of symbol intervals for data transmission. In training phase, a $T_p \times M$ pilot matrix $\sqrt{\frac{\rho_p}{M}}\mathbf{S}_p$ is sent from the transmitter, with ij-th entry the signal sent at i-th time slot from j-th transmit antenna. Note that ρ_p is the corresponding SNR in the training phase, and $\mathbf{S}_p$ satisfies the power constraint $\mathrm{tr}(\mathbf{S}_p\mathbf{S}_p^*) = MT_p$. At the receiver, the observation signal $\mathbf{X}_p$ is received. The whole training process can be represented as [2]

$$\mathbf{X}_p = \sqrt{\frac{\rho_p}{M}}\mathbf{S}_p\mathbf{H} + \mathbf{W}_p, \quad s.t. \quad \mathbf{S}_p \in \mathcal{C}^{T_p \times M},\ \mathrm{tr}(\mathbf{S}_p\mathbf{S}_p^*) = MT_p, \tag{2.6}$$

where $\mathbf{W}_p$ is the noise at the receiver. Stacking the columns of $\mathbf{X}_p$ into one column vector, we have

$$\overrightarrow{\mathbf{X}}_p = \left[\sqrt{\frac{\rho_p}{M}}(\mathbb{I}_N \otimes \mathbf{S}_p)\right]\overrightarrow{\mathbf{H}} + \overrightarrow{\mathbf{W}}_p = \sqrt{\frac{\rho_p}{M}}\mathbf{L}_p\overrightarrow{\mathbf{H}} + \overrightarrow{\mathbf{W}}_p, \tag{2.7}$$

where $\mathbf{L}_p \triangleq \mathbb{I}_N \otimes \mathbf{S}_p$.

Denote the estimation of $\overrightarrow{\mathbf{H}}$ as $\hat{\overrightarrow{\mathbf{H}}}$ and its estimation error as $\Delta\overrightarrow{\mathbf{H}} \triangleq \overrightarrow{\mathbf{H}} - \hat{\overrightarrow{\mathbf{H}}}$. By minimizing MSE of $\hat{\overrightarrow{\mathbf{H}}}$, defined as $\mathrm{MSE}(\hat{\overrightarrow{\mathbf{H}}}) \triangleq \mathbb{E}_{\mathbf{H},\mathbf{W}_p}(\|\Delta\overrightarrow{\mathbf{H}}\|_F^2)$, the MMSE (also LMMSE) estimation of $\overrightarrow{\mathbf{H}}$ can be derived by Bayesian Gauss-Markov theorem [52] as

$$\hat{\overrightarrow{\mathbf{H}}} = \sqrt{\frac{\rho_p}{M}}\left[\left(\frac{\sigma}{\rho}\right)^2 \mathbb{I}_{MN} + \frac{\rho_p}{M}\mathbf{L}_p^*\mathbf{L}_p\right]^{-1} \mathbf{L}_p^*\overrightarrow{\mathbf{X}}_p. \tag{2.8}$$

In data transmission phase, suppose that the information signal $\mathbf{S}_d$ is sent from the transmitter with the constraint $\mathbb{E}\ \mathrm{tr}(\mathbf{S}_d\mathbf{S}_d^*) = MT_d$. The received signal is then

$$\mathbf{X}_d = \sqrt{\frac{\rho_d}{M}}\mathbf{S}_d\mathbf{H} + \mathbf{W}_d, \tag{2.9}$$

where ρ_d is the corresponding SNR in data transmission, and $\mathbf{W}_d$ is the noise at the receiver. Given the estimation $\hat{\overrightarrow{\mathbf{H}}}$ in (2.8), or equivalently, $\hat{\mathbf{H}}$, by straightforward transformation from (2.8), the received signal in (2.9) can be rewritten as

$$\mathbf{X}_d = \sqrt{\frac{\rho_d}{M}}\mathbf{S}_d\hat{\mathbf{H}} + \sqrt{\frac{\rho_d}{M}}\mathbf{S}_d\Delta\mathbf{H} + \mathbf{W}_d = \sqrt{\frac{\rho_d}{M}}\mathbf{S}_d\hat{\mathbf{H}} + \mathbf{W}'_d, \tag{2.10}$$

where $\mathbf{W}'_d \triangleq \sqrt{\frac{\rho_d}{M}}\mathbf{S}_d\Delta\mathbf{H} + \mathbf{W}_d$ denotes the equivalent noise including both the estimation error and the noise.

The time and power constraints for the training-based multiple-antenna system can be summarized as

$$T_c = T_p + T_d \quad \text{and} \quad \rho T_c = \rho_p T_p + \rho_d T_d, \tag{2.11}$$

where ρ is the SNR during the whole transmission phase.

To perform channel training, we need to solve the following problems:

- How to design the pilot signal?

- How to allocate the transmission resources, e.g., power and time, between training and data transmission?

From (2.7), the pilot signal design refers to the design of $\mathbf{S}_p$. Unlike the pilot for single-antenna system which is a scalar, here, $\mathbf{S}_p$ is two-dimensional. So, the design of $\mathbf{S}_p$ refers to the design of pilots sent at each time slot and each transmit antenna during the training phase. The allocation of the resources, however, refers to the optimization over the training parameters T_p, T_d, ρ_p, and ρ_d.

It is plain to see that: Since the pilot signal does not carry desired information for the user, allocating more resources to training seems like a kind of "waste". However, insufficient training could result in inaccurate channel estimation, which in turn would degrade the communication performance. Therefore, there is a tradeoff between allocating more resources to data transmission for higher data rate, and allocating more resources to training for more accurate channel estimation.

To optimally solve the above training problems, several criteria have been posed in the literature, such as MSE of the channel estimations, mutual information (MI), and block error rate (BLER). The MSE criterion is performed by minimizing the estimation error over the pilot signal $\mathbf{S}_p$ and training parameters. The advantage of such criterion is that, it is decoupled from the data transmission phase, and thus is independent of the transmission scheme the system applies. Such MSE-based training design can be found in [44]. In contrast, the MI and BLER-based training designs depend on the data transmission schemes. Since it is usually difficult to get the closed-form expressions of MI and BLER, the corresponding bounds are instead employed as the criteria, which are expressed as the functions of $\mathbf{S}_p$ and training parameters. The optimal training designs are then obtained by maximizing or minimizing the bounds of MI or BLER, respectively, e.g. [2, 48, 49].

In this work, to design the pilot signal, we employ the criterion of MSE; to optimize the resource allocation, we aim at achieving full diversity in data

transmission, which is one the most important indices in wireless communications. From (1.4), we know that diversity is essentially a measurement based on BLER.

2.3 Decoding with Imperfect Channel Estimations

One interesting study surrounding the training-based system is, how the imperfect channel estimations would affect the data transmission performance? In this thesis project, we mainly consider the impact of imperfect channels on diversity in data transmission.

In most literatures, the estimated channels are treated as if perfect in decoding, i.e., the estimation errors are ignored, for simplicity of analysis. This decoding strategy is termed as mismatched decoding [50], for its mismatch with the data equation in (2.9). In general, mismatched decoding is suboptimal, since the estimation error could be large. In [51], a sufficient condition for the full diversity performance of mismatched decoding is proved, which is restated as follows.

Consider the system equation in data transmission as

$$\mathbf{X}_d = \mathbf{K}\mathbf{S}_d + \mathbf{W}_d, \tag{2.12}$$

where $\mathbf{K}$ is an $N \times M$ channel matrix, $\mathbf{S}_d$ is an $M \times T$ transmit information matrix drawn from a finite constellation of matrices, $\mathbf{W}_d$ is the noise matrix, and $\mathbf{X}_d$ is the received signal. Denote the estimation of $\mathbf{K}$ as $\hat{\mathbf{K}}$ with estimation error $\Delta\mathbf{K} \triangleq \hat{\mathbf{K}} - \mathbf{K}$.

- Assumption 1: All entries of the channel matrix $\mathbf{K}$ and the noise matrix $\mathbf{W}$ are independent, and are distributed as Gaussian with mean zero and variance ρ^2 and σ^2, respectively.

- Assumption 2: The difference of any distinct information matrices $\mathbf{S}_{1,d}$ and $\mathbf{S}_{2,d}$ has rank M. In other words, the information signal is designed

as fully diverse.

- Assumption 3: The columns of $\Delta\mathbf{K}$ are independently Gaussian distributed with mean zero and covariance matrix $\gamma^2\sigma^2\mathbb{I}$, where γ is some constant (This covariance constraint actually says that the estimation error has the same "size" as the noise.).

Theorem 2.1. *[51] Under the assumptions shown above, full diversity MN can be achieved by applying the mismatched decoding*

$$\arg\min_{\mathbf{S}_d} \|\mathbf{X}_d - \hat{\mathbf{K}}\mathbf{S}_d\|. \tag{2.13}$$

In [43], the authors also study the decoding where the estimation errors are taken into account. We will call it matched decoding to emphasize the distinction from the mismatched decoding. It is expected that matched decoding has a better performance than mismatched decoding, but with a higher computational complexity.

2.4 Contribution of This Thesis

Cooperative relay network is very attractive as discussed in Section 1.4. However, to perform most of the proposed cooperative schemes, accurate CSI is required at the receiver, and therefore, channel training is needed. In this thesis project, we perform the theoretical analysis on channel training and training-based decodings for MIMO relay networks with multiple single-antenna relays, and multiple transmit and receive antennas, which is in general very challenging for relay networks.

In Chapter 3, channel training is first investigated. The objective is to estimate the transmitter-relay channel vector $\mathbf{f}$, the relay-receiver channel matrix $\mathbf{G}$, and the end-to-end channel matrix $\mathbf{H}$ at the receiver with no estimator requirement at the relays, since such global CSI is commonly required at the receiver by the transmission schemes such as AF and DSTC. We show that,

to estimate the transmitter-relay channel vector $\mathbf{f}$, the knowledge of the relay-receiver channel matrix $\mathbf{G}$ is needed. In other words, training of $\mathbf{f}$ is coupled with training of $\mathbf{G}$. The associated training design, which includes training scheme design, training code design, training time design, and power allocation, are discussed respectively.

Employing the channel estimations provided in Chapter 3, the effect of channel estimation errors on the network diversity performance is studied for DSTC network. Two coherent decodings are considered: mismatched decoding (Chapter 4) in which channel estimations are treated as if perfect, and matched decoding (Chapter 5) in which estimation error is taken into account.

For mismatched decoding, we first show that with the shortest training time, full diversity cannot always be achieved in data transmission. Then, an upper bound on training time is given to ensure full diversity. However, since a long training time is undesirable in practice, to shorten the training time while maintain the full diversity, a novel training scheme, called adaptive training, is provided, whose training time length is adaptive to the quality of the relay-receiver channels.

For matched decoding, we show that it can achieve full diversity with the shortest training time. So, matched decoding shortens the requirement on training time without sacrificing diversity. However, its complexity is prohibitively high. A modified matched decoding, adaptive decoding, is hence introduced by switching between the simplified mismatched decoding and matched decoding to balance the performance and complexity.

2.5 Summary

Before preceding to the investigation on training-based MIMO relay networks, the preliminary on training-based multiple-antenna system was reviewed. To give an idea, the channel training for single-antenna system was first introduced. Then, for the multiple-antenna system, the training-based system model, the training problem statement, and the training design criteria were

presented in detail. In addition, the decoding with imperfect channel estimations was also discussed. In the end, the contribution of this thesis was summarized.

3 Channel Training in MIMO Relay Networks

3.1 Introduction and Literature Review

For cooperative networks with relaying, research activities on the channel training and estimation issues are growing, e.g. [55–62]. Unlike DF relay networks, for AF relay networks, often, we cannot appeal to the results of the point-to-point multiple-antenna system directly, because the end-to-end channels are often concatenations of channels of multiple communication stages. For AF relay networks with single-antenna nodes and single relay, the channel training is studied in [55–57]. In both [55] and [56], channel estimation is performed at the receiver. In [57], the authors consider estimating the Tx-Relay channel at the relay, and having the relay forward the channel information to the receiver. The disadvantages are, the quantization error is induced, and the complexity at the relay is increased. In [58], the channel training problem is investigated for networks with multiple relays but single antenna at each node, where the end-to-end channel coefficients are estimated at the receiver directly.

To our best knowledge, there is little work on channel training for more general MIMO relay networks with multiple relays and multiple antennas at the transmitter and receiver. The main difficulties are two-fold: First, the curse of dimensionality, which is the inherent difficult in MIMO relay networks; second, the coupling of channel trainings for different communication stages: Since there are multiple communication stages, it is possible that the channel training for one stage is coupled with the channel trainings for other stages, which results in a non-Gaussian estimation model making analysis even more involved.

In this chapter, we will investigate channel training for general MIMO relay networks (The detailed system model is refered to Section 1.4.). The objective is to estimate the transmitter-relay channel vector $\mathbf{f}$, the relay-receiver channel matrix $\mathbf{G}$, and the end-to-end channel matrix $\mathbf{H}$ at the receiver, since such

global CSI is commonly required at the receiver for transmission schemes such as AF and DSTC. Note that, with our training scheme, the relays are not required to be equipped with estimators. In other words, the channel estimation is only conducted at the receiver.

To complete channel training, we investigate training scheme design, pilot signal design, and training time design respectively in detail. The power allocation between training phase and data transmission phase is pre-designed, since the optimization over power allocation is beyond the scope. In particular, we set the average power of each node in the network to be the same for both training and data transmission phases. The proposed schemes and analysis results can be straightforwardly extended to other power settings.

In Section 3.2, the training of the relay-receiver channel matrix $\mathbf{G}$ is first discussed. For the training of $\mathbf{f}$, as we will see, since the estimation of $\mathbf{f}$ depends on the knowledge of $\mathbf{G}$, in Section 3.3, it is first investigated by assuming the estimation of $\mathbf{G}$ as error-free; then, in Section 3.4, it is revisited by considering the estimation error of $\mathbf{G}$ for a specific network with $M = 1, R = N \geq 2$. In Section 3.5, the training of $\mathbf{H}$ is investigated and two different schemes are proposed to estimate $\mathbf{H}$. In Section 3.6, further discussions with respect to channel model and applications are included.

3.2 Training of the Relay-Receiver Channel Matrix G

The training of $\mathbf{G}$ is straightforward as the relay-receiver link is a virtual multiple-antenna system, whose training has been well-investigated. Using results in [2], we use the following training design and channel estimation. The pilot $\sqrt{RP_1}\mathbb{I}_R$ is sent from the relays, where P_1 is the average transmit power at each relay for each transmission for both the training and data transmission phases. Denote the received matrix at the receiver as $\mathbf{Y}_p$. We have

$$\overrightarrow{\mathbf{Y}}_p = \sqrt{RP_1}\overrightarrow{\mathbf{G}} + \overrightarrow{\mathbf{W}}_g, \tag{3.1}$$

where $\mathbf{W}_g$ is the $R \times N$ noise matrix at the receiver. By Bayesian Gauss-Markov theorem [52], the LMMSE estimation of $\overrightarrow{\mathbf{G}}$ at the receiver is thus:

$$\hat{\overrightarrow{\mathbf{G}}} = \frac{\sqrt{RP_1}}{1+RP_1}\overrightarrow{\mathbf{Y}}_p. \tag{3.2}$$

Since $\overrightarrow{\mathbf{G}}$ and $\overrightarrow{\mathbf{Y}}_p$ are jointly Gaussian, (3.2) is also the MMSE estimation [52].

Due to the Gaussian model, it also follows that $\hat{\overrightarrow{\mathbf{G}}} \sim \mathcal{CN}\left(\mathbf{0}, \frac{RP_1}{1+RP_1}\mathbb{I}_{RN}\right)$. Let $\Delta\mathbf{G} \triangleq \mathbf{G} - \hat{\mathbf{G}}$, which is the estimation error on $\mathbf{G}$. Invoking the geometric property of the LMMSE estimator that $\Delta\mathbf{G}$ and $\hat{\mathbf{G}}$ are uncorrelated, we have that $\Delta\overrightarrow{\mathbf{G}} \sim \mathcal{CN}\left(\mathbf{0}, \frac{1}{1+RP_1}\mathbb{I}_{RN}\right)$. This training stage takes R symbol intervals.

3.3 Training of the Transmitter-Relay Channel Vector f with Perfect G

In this section, the training of the transmitter-relay channel vector $\mathbf{f}$ is discussed, including the pilot design, and a lower bound on the training time. As will be explained soon, to estimate $\mathbf{f}$, the receiver needs to know $\mathbf{G}$. Thus, this training stage is performed after the training of $\mathbf{G}$, which was presented in Section 3.2. That is, the receiver has already obtained $\hat{\mathbf{G}}$. Since the estimation of $\mathbf{f}$ is coupled with the estimation of $\mathbf{G}$, for simplicity of analysis, in this section we assume that the estimated $\mathbf{G}$ is perfect, i.e., $\hat{\mathbf{G}} = \mathbf{G}$.

To estimate $\mathbf{f}$ at the receiver, we employ the two-step DSTC scheme [25,63]. The explanations of notation follow those in Section 1.4. This training stage takes $2N_p$ symbol intervals, with N_p symbol intervals for each step. Let $\mathbf{B}_p$ be the $N_p \times M$ pilot vector satisfying $\mathrm{tr}(\mathbf{B}_p^*\mathbf{B}_p) = M$. Let $\alpha_p \triangleq \sqrt{\frac{P}{R(P+1)}}$ and $\beta_p \triangleq \sqrt{\frac{P^2N_p}{RM(P+1)}}$, where P is the transmit power of the transmitter and also the total transmit power of the relays. Following the derivation of (1.9) in Section 1.4, at the receiver we have

$$\mathbf{X}_p = \beta_p\mathbf{S}_p\mathbf{H} + \mathbf{W}_p, \tag{3.3}$$

where $\mathbf{S}_p \triangleq [\mathbf{A}_{1,p}\mathbf{B}_p \cdots \mathbf{A}_{R,p}\mathbf{B}_p]$, which is the distributed space-time codeword formed at the receiver,

$$\mathbf{W}_p \triangleq \alpha_p \left[\mathbf{A}_{1,p}\mathbf{v}_{1,p} \cdots \mathbf{A}_{R,p}\mathbf{v}_{R,p}\right] \mathbf{G} + \mathbf{W}_{r,p}, \tag{3.4}$$

which is the equivalent noise matrix with $\mathbf{W}_{r,p}$ the noise matrix at the receiver. Stacking the columns of $\mathbf{X}_p$ into one column, we have

$$\overrightarrow{\mathbf{X}}_p = \beta_p[(\mathbf{G}^t \otimes \mathbb{I}_{N_p})\tilde{\mathbf{S}}_p]\mathbf{f} + \overrightarrow{\mathbf{W}}_p = \beta_p \mathbf{Z}_p \mathbf{f} + \overrightarrow{\mathbf{W}}_p, \tag{3.5}$$

where $\tilde{\mathbf{S}}_p \triangleq \text{diag}\{\mathbf{A}_{1,p}\mathbf{B}_p, \cdots, \mathbf{A}_{R,p}\mathbf{B}_p\}$, and the $NN_p \times MR$ matrix

$$\mathbf{Z}_p \triangleq (\mathbf{G}^t \otimes \mathbb{I}_{N_p})\tilde{\mathbf{S}}_p. \tag{3.6}$$

If $\mathbf{G}$ is perfectly known at the receiver, from (3.5), $\mathbf{Z}_p$ can function as the pilot signal in estimating $\mathbf{f}$. From the Bayesian Gauss-Markov theorem [52], the LMMSE estimation of $\mathbf{f}$ can be obtained as

$$\hat{\mathbf{f}} = \beta_p(\mathbb{I}_{MR} + \beta_p^2 \mathbf{Z}_p^* \mathbf{R}^{-1}_{\overrightarrow{\mathbf{W}}_p} \mathbf{Z}_p)^{-1} \mathbf{Z}_p^* \mathbf{R}^{-1}_{\overrightarrow{\mathbf{W}}_p} \overrightarrow{\mathbf{X}}_p, \tag{3.7}$$

where

$$\mathbf{R}_{\overrightarrow{\mathbf{W}}_p} \triangleq (\mathbb{I}_N + \alpha_p^2 \overline{\mathbf{G}^*\mathbf{G}}) \otimes \mathbb{I}_{N_p}, \tag{3.8}$$

is the covariance matrix of $\overrightarrow{\mathbf{W}}_p$. Since $\mathbf{f}$ and $\overrightarrow{\mathbf{X}}_p$ are jointly Gaussian, (3.7) is also the MMSE estimation. Let $\Delta\mathbf{f} \triangleq \mathbf{f} - \hat{\mathbf{f}}$ be the estimation error vector. It is thus Gaussian distributed with mean zero and covariance matrix

$$\mathbf{R}_{\Delta\mathbf{f}} = \left(\mathbb{I}_{MR} + \beta_p^2 \mathbf{Z}_p^* \mathbf{R}^{-1}_{\overrightarrow{\mathbf{W}}_p} \mathbf{Z}_p\right)^{-1}. \tag{3.9}$$

3.3.1 Pilot Design

In this subsection, we investigate the optimal pilot design, i.e., the designs of $\mathbf{B}_p$ and $\mathbf{A}_{i,p}$'s, based on the MSE of the channel estimation, which is also the

Algorithm 1 Optimal training code design for $N_p \geq MR$.

1: Let $\mathbf{B}_p = [\mathbb{I}_M \ \mathbf{0}_{M,N_p-M}]^t$.
2: Generate an $N_p \times MR$ unitary matrix $\mathbf{S}_{po}$ (e.g., $\mathbf{S}_{po} = [\mathbb{I}_{MR} \ \mathbf{0}_{MR,N_p-MR}]^t$).
3: Generate the $MR \times MR$ permutation matrix $\mathbf{U}_i$ by switching the first M columns with the $(i-1)M+1, \cdots, iM$-th columns of the identity matrix. Let $\mathbf{A}_{i,p} = [\mathbf{S}_{po}\mathbf{U}_i \ (\mathbf{S}_{po}\mathbf{U}_i)^{\perp}]$.

power of the estimation error. The design problem can be represented formally as

$$\min_{\mathbf{A}_{i,p},\mathbf{B}_p} \operatorname{tr}(\mathbf{R}_{\Delta\mathbf{f}}) \quad \text{s.t. } \operatorname{tr}(\mathbf{B}_p^*\mathbf{B}_p) = M \text{ and } \mathbf{A}_{i,p}^*\mathbf{A}_{i,p} = \mathbb{I}_{N_p}, \tag{3.10}$$

where $\mathbf{R}_{\Delta\mathbf{f}}$ is given in (3.9).

We consider the cases $N_p \geq MR$ and $N_p < MR$ separately.

Theorem 3.1. *If $N_p \geq MR$, $tr(\mathbf{R}_{\Delta\mathbf{f}})$ is minimized when $\mathbf{S}_p^*\mathbf{S}_p = \mathbb{I}_{MR}$.*

Proof. See Appendix 3.8.1. □

Theorem 1 says that when $N_p \geq MR$, the optimal pilot design is to make the distributed space-time codeword $\mathbf{S}_p$ unitary, which is consistent with the pilot design in point-to-point MIMO system [2]. Our pilot design problem is thereby reduced to finding $\mathbf{B}_p$ and $\mathbf{A}_{i,p}$'s such that $\mathbf{S}_p$ is an $N_p \times MR$ unitary matrix. We propose a pilot design method in Algorithm 1.

Now we show that the design in Algorithm 1 will result in an optimal code. Note that $N_p \geq MR \geq M$, so Step 1 of Algorithm 1 is always valid. Divide $\mathbf{S}_{po}$ into R blocks each with dimension $N_p \times M$: $\mathbf{S}_{po} = [\mathbf{S}_{p1} \cdots \mathbf{S}_{pR}]$. From Step 3, the effect of right-multiplying $\mathbf{U}_i$ with $\mathbf{S}_{po}$ is to switch $\mathbf{S}_{p1}$ and $\mathbf{S}_{pi}$, making $\mathbf{A}_{i,p} = [\mathbf{S}_{pi} \ \mathbf{S}_{pi}^{\perp}]$. Along with the $\mathbf{B}_p$ generated in Step 1, $\mathbf{A}_{i,p}\mathbf{B}_p = [\mathbf{S}_{pi} \ \mathbf{S}_{pi}^{\perp}][\mathbb{I}_M \ \mathbf{0}_{M,N_p-M}]^t = \mathbf{S}_{pi}$. Hence, $\mathbf{S}_p = [\mathbf{A}_{1,p}\mathbf{B}_p \ \cdots \ \mathbf{A}_{R,p}\mathbf{B}_p] = [\mathbf{S}_{p1} \cdots \mathbf{S}_{pR}] = \mathbf{S}_{po}$ which is unitary. Thus, Algorithm 1 provides an optimal pilot design.

When $N_p < MR$, $\mathbf{S}_p^*\mathbf{S}_p$ is not full rank and $\mathbf{S}_p^*\mathbf{S}_p = \mathbb{I}_{MR}$ cannot be satisfied. For this case, we propose Algorithm 2, which first (Steps 1 and 2) considers

Algorithm 2 Optimal training code design for $M \leq N_p < MR$.

1: Let $\tilde{R} = \lfloor N_p/M \rfloor$.
2: Use Algorithm 1 to generate $\mathbf{B}_p$ and $\mathbf{A}_{i,p}$ for $i = 1, \ldots, \tilde{R}$.
3: For $i = \tilde{R}+1, \ldots, R$, let $\mathbf{A}_{i,p} = \mathbf{A}_{i \bmod \tilde{R},p}$.

a network whose number of relays is $\tilde{R} \triangleq \lfloor N_p/M \rfloor$. For this smaller network, we have $N_p \geq M\tilde{R}$. Thus, Algorithm 1 can be used to optimally design $\mathbf{S}'_p \triangleq [\mathbf{A}_{1,p}\mathbf{B}_p \cdots \mathbf{A}_{\tilde{R},p}\mathbf{B}_p]$ for such network. Then, in Step 3, we repeat the design of $\mathbf{S}'_p$. So, the pilot $\mathbf{S}_p$ has the structure $[\mathbf{S}'_p\ \mathbf{S}'_p\ \cdots]$. Note that in Step 1 of Algorithm 2, we need $\tilde{R} \geq 1$ which implies that $N_p \geq M$. In the next subsection, we will show that this condition is actually an necessary condition for reliable training.

From the above illustration, Algorithm 2 is optimal in the sense of each smaller network with $\tilde{R}$ relays. In the following, we will show that Algorithm 2 also minimizes an upper bound on the power of the estimation error. Let

$$\tilde{\mathbf{Z}}_p = (\mathbf{G}^t \otimes \mathbb{I}_{N_p})\text{diag}\{\mathbf{A}_{1,p}\mathbf{B}_p, \cdots, \mathbf{A}_{\tilde{R},p}\mathbf{B}_p\}, \tag{3.11}$$

which is the first $M\tilde{R}$ columns of $\mathbf{Z}_p$. Order the eigenvalues of

$$[\tilde{\mathbf{Z}}_p\ \mathbf{0}_{NN_p,MR-M\tilde{R}}]^* \mathbf{R}^{-1}_{\overrightarrow{\mathbf{W}}_p} [\tilde{\mathbf{Z}}_p\ \mathbf{0}_{NN_p,MR-M\tilde{R}}] \tag{3.12}$$

as $\tilde{\delta}_1 \geq \cdots \geq \tilde{\delta}_{MR} \geq 0$. Order the eigenvalues of $\mathbf{Z}_p^* \mathbf{R}^{-1}_{\overrightarrow{\mathbf{W}}_p} \mathbf{Z}_p$ as $\delta_1 \geq \cdots \geq \delta_{MR} \geq 0$. Using the interlacing property of matrices [68], we have $\delta_i \geq \tilde{\delta}_i$, for $i = 1, \ldots, MR$, which results in $\frac{1}{1+\beta_p^2\delta_i} \leq \frac{1}{1+\beta_p^2\tilde{\delta}_i}$. Hence,

$$\begin{aligned} \text{tr}\,(\mathbf{R}_{\Delta\mathbf{f}}) &\leq \text{tr}\,(\mathbb{I}_{MR} + \beta_p^2 [\tilde{\mathbf{Z}}_p\ \mathbf{0}]^* \mathbf{R}^{-1}_{\overrightarrow{\mathbf{W}}_p} [\tilde{\mathbf{Z}}_p\ \mathbf{0}])^{-1} && (3.13) \\ &= \text{tr}\left(\mathbb{I}_{M\tilde{R}} + \beta_p^2 \tilde{\mathbf{Z}}_p^* \mathbf{R}^{-1}_{\overrightarrow{\mathbf{W}}_p} \tilde{\mathbf{Z}}_p\right)^{-1} + M(R-\tilde{R}). && (3.14) \end{aligned}$$

Since $\mathbf{S}'_p = [\mathbf{A}_{1,p}\mathbf{B}_p \cdots \mathbf{A}_{\tilde{R},p}\mathbf{B}_p]$ is designed as unitary, based on Theorem 3.1, this upper bound is minimized as the first term is minimized.

3.3.2 Lower Bound on the Training Time

In this subsection, the lower bound on the training time is discussed. The total length of the training, including both the training of $\mathbf{G}$ and the training of $\mathbf{f}$, is $R + 2N_p$. From the estimation $\hat{\mathbf{f}}$ in (3.7), there is no requirement on N_p to conduct the LMMSE estimation. For good performance, however, the number of independent training equations in (3.5) should be no less than that of the unknowns. An interesting thing is that the equivalent pilot in the training equation, $\mathbf{Z}_p = (\mathbf{G}^t \otimes \mathbb{I}_{N_p})\tilde{\mathbf{S}}_p$, is a random matrix whose property depends on the realization of $\mathbf{G}$. We thus consider the best scenario of $\mathbf{G}$ being full rank to obtain lower bounds on N_p and on the total length of training.

When $\mathbf{G}$ is full rank, with a properly designed $\tilde{\mathbf{S}}_p$, the number of independent equations in (3.5) is $\min(NN_p,\, RN_p, MR)$. Since the number of unknowns in $\mathbf{f}$ is MR, for reliable training we need

$$N_p \geq \max(\lceil MR/N \rceil, M) \triangleq N_{p,l}, \tag{3.15}$$

which is thus a lower bound on N_p. Hence, a lower bound on the total training length is $R + 2N_{p,l}$.

For the training-based point-to-point MIMO system, such derived lower bound can guarantee full diversity in the data transmission with mismatched decoding [43, 51]. For the MIMO relay network, however, we find that due to the complicated concatenation of the two layers of channels, this result does not always hold considering different network settings. The detailed discussions will be found in Chapter 4.

3.4 Training of the Transmitter-Relay Channel Vector f with Estimated G

In Section 3.3, the training of $\mathbf{f}$ when $\mathbf{G}$ is known perfectly at the receiver is discussed. In reality, only an estimation of $\mathbf{G}$, shown in (3.2), is available.

In this section, we will revisit the training of $\mathbf{f}$ when only the estimated $\mathbf{G}$ given in (3.2) is available at the receiver. The challenging of this case over the previous one is, now, to estimate $\mathbf{f}$, the estimation error of $\mathbf{G}$ needs to be considered. For tractable analysis, we restrict our attention to the network with $M = 1$, and $R = N(R \geq 2)$, i.e., the network with a single-antenna transmitter, R single-antenna relays, and a R-antenna receiver. The extension to the general network will be discussed in Subsection 3.4.4. Since $M = 1$, the pilot sent from the transmitter is with the dimension $N_p \times 1$, which is a vector. To emphasize this point, we employ a lowercase variable $\mathbf{s}_p$ to denote the pilot sent from the transmitter.

The following theorem on the LMMSE estimation of $\mathbf{f}$ is proved.

Theorem 3.2. *Given $\hat{\mathbf{G}}$ in (3.2), the LMMSE estimation of $\mathbf{f}$ with the observation model in (3.5) is*

$$\hat{\mathbf{f}} = \beta_p \left(\mathbb{I}_R + \beta_p^2 \hat{\mathbf{Z}}_p^* \mathbf{R}^{-1}_{\overrightarrow{\mathbf{W}}_{e,p}} \hat{\mathbf{Z}}_p \right)^{-1} \hat{\mathbf{Z}}_p^* \mathbf{R}^{-1}_{\overrightarrow{\mathbf{W}}_{e,p}} \overrightarrow{\mathbf{X}}_p, \tag{3.16}$$

where $\hat{\mathbf{Z}}_p \triangleq (\hat{\mathbf{G}}^t \otimes \mathbb{I}_{N_p}) \tilde{\mathbf{S}}_p$, and

$$\mathbf{R}_{\overrightarrow{\mathbf{W}}_{e,p}} \triangleq \frac{\beta_p^2}{1+P} \left(\mathbb{I}_R \otimes \mathbf{S}_p \mathbf{S}_p^* \right) + \frac{R\alpha_p^2}{1+P} \mathbb{I}_{RN_p} + \left(\mathbb{I}_R + \alpha_p^2 \hat{\mathbf{G}}^t \overline{\hat{\mathbf{G}}} \right) \otimes \mathbb{I}_{N_p}. \tag{3.17}$$

Let $\Delta\mathbf{f} \triangleq \mathbf{f} - \hat{\mathbf{f}}$, which is the estimation error on $\mathbf{f}$. $\Delta\mathbf{f}$ has zero-mean i.e, $\mathbb{E}(\Delta\mathbf{f}) = \mathbf{0}$. Its covariance matrix is

$$\mathbf{R}_{\Delta\mathbf{f}} \triangleq \left(\mathbb{I}_R + \beta_p^2 \hat{\mathbf{Z}}_p^* \mathbf{R}^{-1}_{\overrightarrow{\mathbf{W}}_{e,p}} \hat{\mathbf{Z}}_p \right)^{-1}. \tag{3.18}$$

Proof. Replacing $\mathbf{G}$ with $\hat{\mathbf{G}} + \Delta\mathbf{G}$ in (3.5), we get

$$\overrightarrow{\mathbf{X}}_p = \beta_p \hat{\mathbf{Z}}_p \mathbf{f} + \beta_p [(\Delta\mathbf{G}^t \otimes \mathbb{I}_{N_p}) \tilde{\mathbf{S}}_p] \mathbf{f} + \overrightarrow{\mathbf{W}}_{g1,p} + \overrightarrow{\mathbf{W}}_{g2,p} + \overrightarrow{\mathbf{W}}_{r,p}, \tag{3.19}$$

where

$$\mathbf{W}_{g1,p} \triangleq \alpha_p \left[\mathbf{A}_{1,p}\mathbf{v}_{1,p} \cdots \mathbf{A}_{R,p}\mathbf{v}_{R,p}\right] \hat{\mathbf{G}} \tag{3.20}$$

and

$$\mathbf{W}_{g2,p} \triangleq \alpha_p \left[\mathbf{A}_{1,p}\mathbf{v}_{1,p} \cdots \mathbf{A}_{R,p}\mathbf{v}_{R,p}\right] \Delta\mathbf{G}. \tag{3.21}$$

Define

$$\overrightarrow{\mathbf{W}}_{e,p} \triangleq \beta_p[(\Delta\mathbf{G}^t \otimes \mathbb{I}_{N_p})\tilde{\mathbf{S}}_p]\mathbf{f} + \overrightarrow{\mathbf{W}}_{g1,p} + \overrightarrow{\mathbf{W}}_{g2,p} + \overrightarrow{\mathbf{W}}_{r,p}, \tag{3.22}$$

which is the equivalent noise in the training equation. It has zero-mean. Its covariance matrix can be calculated as follows.

$$\begin{aligned}
\mathbf{R}_{\overrightarrow{\mathbf{W}}_{e,p}} =& \mathbb{E}\ (\overrightarrow{\mathbf{W}}_{e,p}\overrightarrow{\mathbf{W}}^*_{e,p}) \\
=& \mathbb{E}\ \Big(\beta_p^2[(\Delta\mathbf{G}^t \otimes \mathbb{I}_{N_p})\tilde{\mathbf{S}}_p]\mathbf{f}\mathbf{f}^*[(\Delta\mathbf{G}^t \otimes \mathbb{I}_{N_p})\tilde{\mathbf{S}}_p]^* + \overrightarrow{\mathbf{W}}_{g1,p}\overrightarrow{\mathbf{W}}^*_{g1,p} \\
& \qquad + \overrightarrow{\mathbf{W}}_{g2,p}\overrightarrow{\mathbf{W}}^*_{g2,p} + \overrightarrow{\mathbf{W}}_{r,p}\overrightarrow{\mathbf{W}}^*_{r,p}\Big) \\
=& \frac{\beta_p^2}{1+P}(\mathbb{I}_R \otimes \mathbf{S}_p\mathbf{S}_p^*) + \frac{R\alpha_p^2}{1+P}\mathbb{I}_{RN_p} + (\mathbb{I}_R + \alpha_p^2\hat{\mathbf{G}}^t\overline{\hat{\mathbf{G}}}) \otimes \mathbb{I}_{N_p},
\end{aligned}$$

which is (3.17). The second equality is derived since $[(\Delta\mathbf{G}^t \otimes \mathbb{I}_{N_p})\tilde{\mathbf{S}}_p]\mathbf{f}$, $\overrightarrow{\mathbf{W}}_{g1,p}$, $\overrightarrow{\mathbf{W}}_{g2,p}$, and $\overrightarrow{\mathbf{W}}_{r,p}$ are mutually uncorrelated. Note that although $\mathbf{f}$, which is to be estimated, is contained in $\overrightarrow{\mathbf{W}}_{e,p}$, it introduces no trouble in the LMMSE estimator since $\overrightarrow{\mathbf{W}}_{e,p}$ is uncorrelated with $\mathbf{f}$. This can be seen by verifying that $\mathbb{E}\ (\overrightarrow{\mathbf{W}}_{e,p}\mathbf{f}^*) = \mathbb{E}\ (\overrightarrow{\mathbf{W}}_{e,p})\mathbb{E}\ (\mathbf{f}^*) = \mathbf{0}$. By Bayesian Gauss-Markov theorem [52], the LMMSE estimation of $\mathbf{f}$ is given in (3.16), and the covariance matrix of $\Delta\mathbf{f}$ is in (3.18). □

Note that since $\mathbf{f}$ and $\overrightarrow{\mathbf{X}}_p$ are not jointly Gaussian, the LMMSE estimation of $\mathbf{f}$ in (3.16) is not the MMSE estimation, and the estimation error $\Delta\mathbf{f}$ is not Gaussian.

3.4.1 Pilot Design

The pilot design problem is to find $\mathbf{s}_p$ and $\mathbf{A}_{i,p}$'s such that the power of the estimation error of $\hat{\mathbf{f}}$ is minimized, i.e.,

$$\min_{\mathbf{A}_{i,p},\mathbf{s}_p} \operatorname{tr}(\mathbf{R}_{\Delta\mathbf{f}}) \quad \text{s.t. } \mathbf{s}_p^*\mathbf{s}_p = 1 \text{ and } \mathbf{A}_{i,p}^*\mathbf{A}_{i,p} = \mathbb{I}_{N_p}, \tag{3.23}$$

where $\mathbf{R}_{\Delta\mathbf{f}}$ is given in (3.18). In Subsection 3.3.1, where $\mathbf{G}$ is assumed to be perfectly known at the receiver, it is shown that if $N_p \geq R$, the power of the estimation error is minimized when $\mathbf{S}_p^*\mathbf{S}_p = \mathbb{I}_R$ and an algorithm, Algorithm 1, is proposed to gain such a pilot; if $1 \leq N_p < R$, Algorithm 2 is proposed. However, when only $\hat{\mathbf{G}}$ is available at the receiver, the expression of $\mathbf{R}_{\Delta\mathbf{f}}$ in (3.18) is even more involved than that in (3.9), which largely complicates the optimization problem in (3.23). For simplicity, we adopt the pilot designs in Subsection 3.3.1. When the estimation on $\mathbf{G}$ has good quality, for example, when the average power P is large, $\Delta\mathbf{G}$ is small compared with $\mathbf{G}$ and it can be expected that these pilot designs for perfect $\mathbf{G}$ perform close to optimal.

A case of special interest is when $N_p = 1$. In this case, $\mathbf{s}_p$ and $\mathbf{A}_{i,p}$ reduce to scalars. An obvious code design is $\mathbf{s}_p = \mathbf{A}_{i,p} = 1$ to satisfy the power constraints.

3.4.2 Lower Bound on the Training Time

A lower bound on the training time can be derived similarly as that in Subsection 3.3.2, where the number of independent training equations in (3.19) should be no less than that of the unknowns. The difference is, instead of the perfect $\mathbf{G}$, the estimation of $\mathbf{G}$ is available in estimating $\mathbf{f}$. A lower bound on N_p is thus derived by assuming $\hat{\mathbf{G}}$ as full rank. Since the number of unknowns in $\mathbf{f}$ is R, a lower bound on N_p is 1, i.e., $N_p \geq 1$, which is innocuous. A lower bound on the total training length is thus $R + 2$, which includes both the training of $\mathbf{G}$ and the training of $\mathbf{f}$.

3.4.3 Estimation Error on Quality of f

In the following, we investigate the estimation error on $\mathbf{f}$ for the network with $M = 1$, and $R = N(R \geq 2)$. Define the mean square error of the estimation $\hat{\mathbf{f}}$ as $\text{MSE}(\hat{\mathbf{f}}) \triangleq \mathbb{E}_{\hat{\mathbf{G}}}(\text{tr}(\mathbf{R}_{\Delta\mathbf{f}}))$, which is the average power of the estimation error. Note that this definition is slightly different from the conventional one due to the extra average over $\hat{\mathbf{G}}$. This is because in the training of $\mathbf{f}$, the equivalent pilot is a function of the random matrix $\hat{\mathbf{G}}$, while conventionally the pilot is a fixed scalar or matrix. We analyze the behaviour of $\text{MSE}(\hat{\mathbf{f}})$ under the aforementioned training and pilot designs. The results are essential to the diversity derivations in Chapter 4 and Chapter 5 .

Theorem 3.3. *With the estimation of* $\mathbf{f}$ *in (3.16) and the aforementioned pilot designs, when* $N_p = 1$, $MSE(\hat{\mathbf{f}}) = 2R^2 \log P/P + O(1/P)$, *when* $N_p \geq R$, $MSE(\hat{\mathbf{f}}) = O(1/P)$.

Proof. See Appendix 3.8.2. □

The different scalings of $\text{MSE}(\hat{\mathbf{f}})$ for different values of N_p is due to the randomness of the equivalent pilot in (3.19) and can be intuitively explained as follows. When $N_p = 1$, with the proposed pilot designs, (3.19) can be rewritten as $\overrightarrow{\mathbf{X}}_p = \beta_p \hat{\mathbf{G}}^t \mathbf{f} + \overrightarrow{\mathbf{W}}_{e,p}$. When $\hat{\mathbf{G}}$ is full rank, there are R independent equations. When $\hat{\mathbf{G}}$ is rank deficient, the number of independent equations is less than R, which consequently leads to inaccurate estimations since there are R unknown parameters. Although $\hat{\mathbf{G}}$ is full rank with probability 1, it can get infinitely closed to singular with non-zero probability, which on average affects the MSE behaviour. When $N_p \geq R$, the number of independent training equations in (3.19) is at least R regardless of the rank of $\hat{\mathbf{G}}$. Since there are R unknown elements in $\mathbf{f}$, reliable estimation is expected.

When $N_p = 1$, other than the scalings with respect to P discussed above, we can also see from Theorem 3.3 that $\text{MSE}(\hat{\mathbf{f}})$ is proportional to R^2 when $\log P \gg 1$. Since there are R relays in total, the average $\text{MSE}(\hat{\mathbf{f}})$ per relay

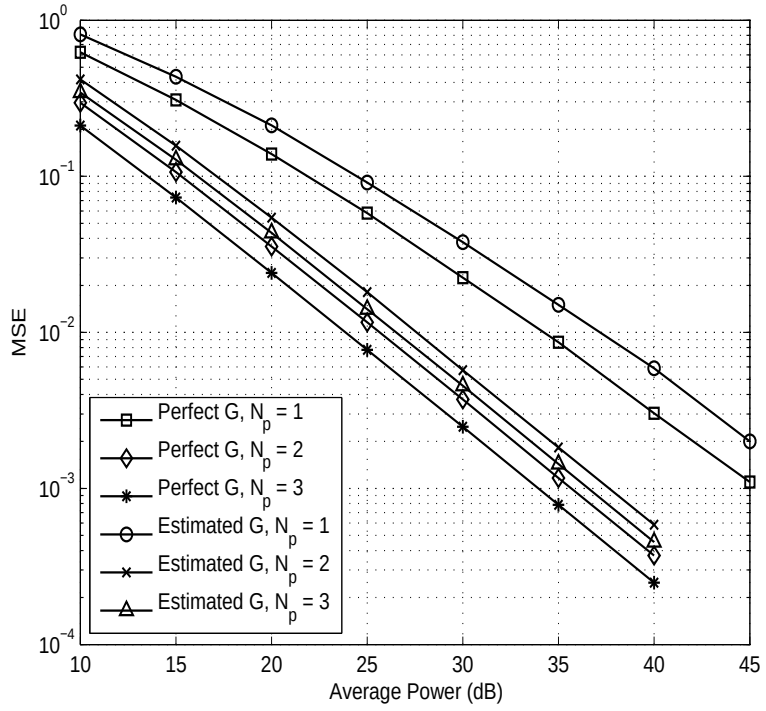

Figure 3.1: MSE($\hat{\mathbf{f}}$) under perfect $\mathbf{G}$ and estimated $\mathbf{G}$ for the network with $R = 2$.

is linear in R. This implies that, as the network has more relays and receive antennas, the estimation quality gets worse.

When $\mathbf{G}$ is perfectly known at the receiver, we also briefly show in Appendix 3.8.2 that MSE($\hat{\mathbf{f}}$) has similar behaviour: when $N_p = 1$, MSE($\hat{\mathbf{f}}$) $= R^2 \log P/P + O(1/P)$, when $N_p \geq R$, MSE($\hat{\mathbf{f}}$) $= O(1/P)$. When $N_p = 1$, comparing MSE($\hat{\mathbf{f}}$) under estimated $\mathbf{G}$ with that under perfect $\mathbf{G}$, there is approximately 3dB degradation due to the estimation error on $\mathbf{G}$.

In Fig. 3.1, to verify our results in Theorem 3.3, MSE($\hat{\mathbf{f}}$) for $N_p = 1, 2$, and 3 are shown for the network with $R = 2$. For comparison, MSE($\hat{\mathbf{f}}$) under perfect $\mathbf{G}$ is also exhibited. It is observed that given the same N_p, MSE($\hat{\mathbf{f}}$) under perfect $\mathbf{G}$ is about 3dB lower than that under estimated $\mathbf{G}$. It should be emphasized that the 3dB difference is on MSE($\hat{\mathbf{f}}$) not the BLER performance of the network. With the proposed training for $\mathbf{G}$, from our simulation, the difference on the network BLER of the two cases is very small in DSTC net-

work. The analytical results in Theorem 3.3, when $N_p = 1$ MSE($\hat{\mathbf{f}}$) scales as $\log P/P$, and when $N_p \geq 2$ MSE($\hat{\mathbf{f}}$) scales as $1/P$, are confirmed by Fig. 3.1 by reading the slopes of the MSE curves. We can also see that in the high power region, to achieve the same level of MSE($\hat{\mathbf{f}}$), by increasing N_p from 1 to 2, 10dB power can be saved in training. But by further increasing N_p from 2 to 3, only 0.5dB power can be saved.

3.4.4 Discussion on Extension

In the above discussions, we studied the training of $\mathbf{f}$ with estimated $\mathbf{G}$ by constraining our focus on the relay network with a single-antenna transmitter, R single-antenna relays, and a R-antenna receiver. In this subsection, we will discuss the extension work to the more general network settings.

First, consider the network with a single-antenna transmitter, unequal numbers of relays and receive antennas. For this network, the training scheme of $\mathbf{f}$ and its estimation rule (3.16) can be applied directly. Nevertheless, the analysis of the training properties can be further involved, and more investigation is needed.

Second, consider the network with an M-antenna transmitter, R single-antenna relays, and an N-antenna receiver. Denote the channel vector from the i-th transmit antenna to the relays as $\tilde{\mathbf{f}}_i = [f_{i1} \ \cdots \ f_{iR}]^t$. Since this general network can be transformed into M sub networks each with a single-antenna transmitter, R single-antenna relays, and an N-antenna receiver, training of $\mathbf{f}$ can thus be split into training of $\tilde{\mathbf{f}}_i, i = 1, \cdots, M$, where our training scheme and estimation rule in (3.16) can be applied.

3.5 Training of the End-to-End Channel Matrix H

In this section, to estimate the end-to-end channel matrix $\mathbf{H}$, two training strategies, separate training and direct training, are suggested and compared.

3.5.1 Separate Training of H

In (1.7), the end-to-end channel matrix $\mathbf{H}$ is given as $\mathbf{H} = [(\mathbf{f}_1\mathbf{g}_1)^t \cdots (\mathbf{f}_R\mathbf{g}_R)^t]^t$, which is constructed by $\mathbf{f}$ and $\mathbf{G}$. So, one straightforward method is to estimate $\mathbf{H}$ through the estimations of $\mathbf{f}$ and $\mathbf{G}$. Specifically, the estimation of $\mathbf{H}$ can be obtained as

$$\hat{\mathbf{H}} = [(\hat{\mathbf{f}}_1\hat{\mathbf{g}}_1)^t \cdots (\hat{\mathbf{f}}_R\hat{\mathbf{g}}_R)^t]^t. \tag{3.24}$$

In this case, the analysis of the training of $\mathbf{H}$ follows those discussed in the previous sections.

3.5.2 Direct Training of H

Another method refers to estimate the end-to-end channel matrix $\mathbf{H}$ directly at the receiver. Recall that the DSTC system equation is derived in (3.3) as $\mathbf{X}_p = \beta_p\mathbf{S}_p\mathbf{H} + \mathbf{W}_p$. By stacking the columns of $\mathbf{X}_p$ into one single column vector in a different way, (3.3) can be rewritten as

$$\overrightarrow{\mathbf{X}}_p = \beta_p(\mathbb{I}_N \otimes \mathbf{S}_p)\overrightarrow{\mathbf{H}} + \overrightarrow{\mathbf{W}}_p = \beta_p\mathbf{L}_p\overrightarrow{\mathbf{H}} + \overrightarrow{\mathbf{W}}_p, \tag{3.25}$$

where $\mathbf{L}_p \triangleq \mathbb{I}_N \otimes \mathbf{S}_p$, which can be treated as the equivalent pilot signal in estimating $\mathbf{H}$. Note that, in contrast to the equivalent pilot signal $\mathbf{Z}_p$ in estimating $\mathbf{f}$, with a pre-designed $\mathbf{S}_p$, $\mathbf{L}_p$ is deterministic.

To obtain the LMMSE estimation of $\overrightarrow{\mathbf{H}}$ from (3.25), we need to calculate the mean and covariance matrix of $\overrightarrow{\mathbf{H}}$. It can be proved easily that $\mathbb{E}(\overrightarrow{\mathbf{H}}) = \mathbf{0}$. As all elements of $\mathbf{f}$ and $\mathbf{G}$ are *i.i.d.* $\mathcal{CN}(0,1)$, we can obtain from straightforward calculation that the covariance matrix of $\overrightarrow{\mathbf{H}}$ is $\mathbf{R}_{\overrightarrow{\mathbf{H}}} = \mathbb{I}_{MNR}$. Thus, from the Bayesian Gauss-Markov theorem [52], the LMMSE estimation of $\overrightarrow{\mathbf{H}}$ is

$$\hat{\overrightarrow{\mathbf{H}}} = \beta_p(\mathbb{I}_{MNR} + \beta_p^2\mathbf{L}_p^*\mathbf{R}^{-1}_{\overrightarrow{\mathbf{W}}_p}\mathbf{L}_p)^{-1}\mathbf{L}_p^*\mathbf{R}^{-1}_{\overrightarrow{\mathbf{W}}_p}\overrightarrow{\mathbf{X}}_p, \tag{3.26}$$

where $\mathbf{R}_{\overrightarrow{\mathbf{W}}_p}$ is the covariance matrix of $\overrightarrow{\mathbf{W}}_p$ given in (3.8). Let $\Delta\overrightarrow{\mathbf{H}} = \overrightarrow{\mathbf{H}} - \hat{\overrightarrow{\mathbf{H}}}$ be the estimation error. The mean and covariance matrix of $\Delta\overrightarrow{\mathbf{H}}$ can be derived

to be $\mathbb{E}\,(\Delta\overrightarrow{\mathbf{H}}) = \mathbf{0}$ and $\mathbf{R}_{\Delta\overrightarrow{\mathbf{H}}} \triangleq (\mathbb{I}_{MNR} + \beta_p^2 \mathbf{L}_p^* \mathbf{R}_{\overrightarrow{\mathbf{W}}_p}^{-1} \mathbf{L}_p)^{-1}$, respectively.

From (3.26), the estimation of $\overrightarrow{\mathbf{H}}$ requires the information of $\mathbf{G}$ through $\mathbf{R}_{\overrightarrow{\mathbf{W}}_p}$, where $\mathbf{G}^t\overline{\mathbf{G}}$ is essentially needed. Since $\mathbb{E}\,(\mathbf{G}^t\overline{\mathbf{G}}) = R\,\mathbb{I}_N$, we have $\mathbf{R}_{\overrightarrow{\mathbf{W}}_p} \approx (1 + \alpha_p^2 R)\mathbb{I}_{NN_p}$ by replacing $\mathbf{G}^t\overline{\mathbf{G}}$ with its expectation in $\mathbf{R}_{\overrightarrow{\mathbf{W}}_p}$. This approximation is expected to be tight especially for large R. Hence, without the trainings of $\mathbf{G}$ and $\mathbf{f}$, an estimation of $\overrightarrow{\mathbf{H}}$ can be derived as

$$\hat{\overrightarrow{\mathbf{H}}}_{dir} = \frac{\beta_p}{1 + \alpha_p^2 R} \left(\mathbb{I}_{MNR} + \frac{\beta_p^2}{1 + \alpha_p^2 R} \mathbf{L}_p^* \mathbf{L}_p \right)^{-1} \mathbf{L}_p^* \overrightarrow{\mathbf{X}}_p \tag{3.27}$$

by replacing $\mathbf{R}_{\overrightarrow{\mathbf{W}}_p}$ with its approximation in (3.26). The subscript "dir" in (3.27) indicates the direct training scheme. Similarly, by using the approximation of $\mathbf{R}_{\overrightarrow{\mathbf{W}}_p}$, the covariance matrix of $\Delta\overrightarrow{\mathbf{H}}$ is transformed to be $\mathbf{R}_{\Delta\overrightarrow{\mathbf{H}},dir} \triangleq (\mathbb{I}_{MNR} + \frac{\beta_p^2}{1+\alpha_p^2 R} \mathbf{L}_p^* \mathbf{L}_p)^{-1}$.

Due to the absence of the trainings of $\mathbf{G}$ and $\mathbf{f}$, performing the direct training scheme takes $2N_p$ symbol intervals. Since there are MRN unknown variables in $\mathbf{H}$, to reliably estimate $\mathbf{H}$, the number of independent training equations in (3.25) should be no less than MRN. In other words, the lower bound on N_p is MR, and the lower bound on training time to estimate $\mathbf{H}$ using direct training is $2MR$.

The optimal pilot signal design, which includes the designs of $\mathbf{B}_p$ and $\mathbf{A}_{i,p}$'s, is obtained by minimizing the trace of $\mathbf{R}_{\Delta\overrightarrow{\mathbf{H}},dir}$. In other words, the optimization problem is formulated as

$$\min_{\mathbf{A}_{i,p},\mathbf{B}_p} \mathrm{tr}(\mathbf{R}_{\Delta\overrightarrow{\mathbf{H}},dir}) \quad \text{s.t.} \quad \mathrm{tr}\,(\mathbf{B}_p^*\mathbf{B}_p) = M \quad \text{and} \quad \mathbf{A}_{i,p}^*\mathbf{A}_{i,p} = \mathbb{I}_{N_p}. \tag{3.28}$$

The following theorem on the optimal pilot design is derived.

Theorem 3.4. *If $N_p \geq MR$, $tr(\mathbf{R}_{\Delta\overrightarrow{\mathbf{H}},dir})$ is minimized when $\mathbf{S}_p^*\mathbf{S}_p = \mathbb{I}_{MR}$.*

Proof. See Appendix 3.8.3. □

To generate $\mathbf{B}_p$ and $\mathbf{A}_{i,p}$'s with the constraint $\mathbf{S}_p^*\mathbf{S}_p = \mathbb{I}_{MR}$, Algorithm 1 in Subection 3.3.1 can be used.

3.5.3 Comparison on Separate Training and Direct Training

To estimate $\mathbf{H}$, we provided separate training and direct training in Subsection 3.5.1 and Subsection 3.5.2, respectively. In direct training, the LMMSE estimation of $\mathbf{H}$ is derived by treating the entries of $\mathbf{H}$ as if independent. This is appropriate for networks with single transmit and single receive antenna. For networks with multiple transmit and receive antennas, i.e., for $M > 1$ and $N > 1$, however, due to the special structure of $\mathbf{H}$ as shown in (1.7), entries of $\mathbf{H}$ are related. The number of unknowns in $\mathbf{H}$ is actually $R(M + N - 1)$, which is less than MNR. Treating them as if independent can result in longer training time than necessary and suboptimal performance. In contrast, in separate training, the estimation of $\mathbf{H}$ is derived from the estimations of $\mathbf{f}$ and $\mathbf{G}$, where the structure property of $\mathbf{H}$ is exploited. Hence, for large networks, separate training is preferred.

In the following, we show heuristically why the separate training is preferred by comparing the lower bounds on the training time for both schemes. From the previous investigations, we know that for separate training, the lower bound on the training time is

$$R + 2N_{p,l} = R + 2\max(\lceil MR/N \rceil, M), \tag{3.29}$$

while for direct training, the corresponding lower bound is $2MR$. Define the difference between these two as $\Delta_l \triangleq 2MR - (R + 2N_{p,l})$. We have

$$\Delta_l \begin{cases} \approx \frac{R(2MN-N-2M)}{N}, & R \geq N \\ = 2MR - R - 2M, & R < N \end{cases}. \tag{3.30}$$

It is plain to see that, as the network size gets larger, Δ_l gets bigger.

Therefore, in this work, we employ separate training to estimate $\mathbf{H}$, while the direct training is used as a benchmark.

3.6 Discussions

3.6.1 Discussion on Channel Correlation

In the previous discussions, we focused on the investigation of the channel training and estimation problems in networks with independent channels. In some network applications, channel correlation may exist due to, for example, lack of enough spacing between antennas. One common model for correlated channel is the Kronecker MIMO channel model [64], using which the relay-receiver channel matrix $\mathbf{G}$ and the transmitter-relay channel vector $\mathbf{f}$ can be written as

$$\mathbf{G} = \mathbf{R}_{Relay}^{1/2} \mathbf{G}_w \left(\mathbf{R}_{Rx}^{1/2}\right)^t \tag{3.31}$$

and

$$\mathbf{f} = \left[\mathbf{R}_{Relay}^{1/2} \otimes \mathbf{R}_{Tx}^{1/2}\right] \mathbf{f}_w, \tag{3.32}$$

where $\mathbf{R}_{Tx}$, $\mathbf{R}_{Relay}$, and $\mathbf{R}_{Rx}$ are the correlation matrices of the transmitter, relays, and receiver respectively, and entries of $\mathbf{G}_w$ and $\mathbf{f}_w$ are i.i.d following $\mathcal{CN}(0,1)$. If the correlation matrices are known at the receiver, to estimate $\mathbf{G}$ and $\mathbf{f}$ is equivalent to estimate $\mathbf{G}_w$ and $\mathbf{f}_w$. The proposed training scheme in Sections 3.2, 3.3, and 3.4 can be applied directly. The training equations for $\mathbf{G}_w$ and $\mathbf{f}_w$ under the correlation channel model can be written as:

$$\mathbf{Y}'_p = \mathbf{Y}_p \left(\mathbf{R}_{Rx}^{-1/2}\right)^t = \sqrt{RP_1}\mathbf{R}_{Relay}^{1/2}\mathbf{G}_w + \mathbf{W}_g \left(\mathbf{R}_{Rx}^{-1/2}\right)^t, \tag{3.33}$$

and

$$\overrightarrow{\mathbf{X}}_p = \beta_p[(\mathbf{G}^t \otimes \mathbb{I}_{N_p})\tilde{\mathbf{S}}_p]\mathbf{K}\mathbf{f}_w + \overrightarrow{\mathbf{W}}_p, \tag{3.34}$$

where $\mathbf{K} = \mathbf{R}_{Relay}^{1/2} \otimes \mathbf{R}_{Rx}^{1/2}$. The LMMSE estimations of $\mathbf{G}_w$ and $\mathbf{f}_w$ can be derived straightforwardly using our proposed scheme by treating $\mathbf{R}_{Relay}^{1/2}$ and $[(\mathbf{G}^t \otimes \mathbb{I}_{N_p})\tilde{\mathbf{S}}_p]\mathbf{K}$ as pilots. However, for the training code design, and training time investigation, more involved investigation are needed.

If the channel correlation matrices are unknown at the receiver, the chan-

nel estimation problem includes the estimation of the correlation matrices in addition to the estimation of $\mathbf{G}_w$ and $\mathbf{f}_w$. Design and results in the previous discussion cannot be applied directly, further investigation are required.

3.6.2 Discussion on Applications

In the training scheme we proposed, all channels, including the transmitter-relay channels, the relay-receiver channels, and the end-to-end channels, are estimated at the receiver. In many non-regenerative cooperative schemes such as AF [17], DSTC [25], and beamforming [21], such global CSI information is required at the receiver, and our proposed training design can be applied. It is noteworthy that the proposed scheme requires neither feedback nor communication of channel coefficients among network nodes.

3.7 Summary

In this chapter, for the MIMO relay network with M antennas at the transmitter, N antennas at the receiver, and $R(R \geq 2)$ relays each with a single antenna, a training scheme was proposed to estimate the transmitter-relay channel vector $\mathbf{f}$, the relay-receiver channel matrix $\mathbf{G}$, and the end-to-end channel matrix $\mathbf{H}$ at the receiver. To estimate $\mathbf{G}$, the training results in the point-to-point multiple-antenna system can be employed directly. To estimate $\mathbf{f}$, the DSTC scheme was used. Since the equivalent pilot signal in estimating $\mathbf{f}$ needs the information of $\mathbf{G}$, the training of $\mathbf{f}$ was first discussed for the MIMO relay network by assuming the estimated $\mathbf{G}$ as perfect. Then, considering the estimation error of $\mathbf{G}$, the training of $\mathbf{f}$ was revisited for the specific network with $M = 1, R = N \geq 2$. To estimate $\mathbf{H}$, two training schemes, called separate training and direct training, were provided and compared. In addition, the application of our proposed training scheme, and generalization of it to the network with correlated channel models were also discussed.

3.8 Appendixes

3.8.1 Proof of Theorem 3.1

Denote the (i,j)th element of $(\mathbb{I}_N + \alpha_p^2 \mathbf{G}^t \overline{\mathbf{G}})^{-1}$ as $\tilde{g}_{ij}$, after some algebra,

$$\begin{aligned}
&\mathbb{I}_{MR} + \beta_p^2 \mathbf{Z}_p^* \mathbf{R}_{\overrightarrow{\mathbf{W}}_p}^{-1} \mathbf{Z}_p \\
=&\mathbb{I}_{MR} + \beta_p^2 \sum_{i=1}^{N} \sum_{j=1}^{N} \tilde{g}_{ji} \left(\text{diag}\{\overline{g}_{1j}, \cdots, \overline{g}_{Rj}\} \otimes \mathbb{I}_M \right) \mathbf{S}_p^* \mathbf{S}_p \left(\text{diag}\{g_{1j}, \cdots, g_{Rj}\} \otimes \mathbb{I}_M \right).
\end{aligned} \tag{3.35}$$

Recall that $\text{Cov}(\Delta \mathbf{f}) = (\mathbb{I}_{MR} + \beta_p^2 \mathbf{Z}_p^* \mathbf{R}_{\overrightarrow{\mathbf{W}}_p}^{-1} \mathbf{Z}_p)^{-1}$. Using (3.35), we can lower bound $\text{tr}\left(\text{Cov}(\Delta \mathbf{f})\right)$ as:

$$\begin{aligned}
\text{tr}\left(\text{Cov}(\Delta \mathbf{f})\right) \geq & \sum_{i=1}^{MR} \left[\left(\mathbb{I}_{MR} + \beta_p^2 \mathbf{Z}_p^* \mathbf{R}_{\overrightarrow{\mathbf{W}}_p}^{-1} \mathbf{Z}_p \right)_{ii} \right]^{-1} && (3.36) \\
= & \sum_{k=1}^{R} \sum_{i=1}^{M} \frac{1}{1 + \beta_p^2 l_k d_i} \\
\geq & \sum_{k=1}^{R} \frac{M}{1 + \beta_p^2 l_k}, && (3.37)
\end{aligned}$$

where $l_k = \sum_{i=1}^{N} \sum_{j=1}^{N} \tilde{g}_{ji} \overline{g}_{kj} g_{ki}$, and $d_i \geq 0$ is the ith diagonal element of $\mathbf{B}_p^* \mathbf{B}_p$ satisfying $\sum_{i=1}^{M} d_i = M$. For (3.36) to become an equality, $\mathbb{I}_{MR} + \beta_p^2 \mathbf{Z}_p^* \mathbf{R}_{\overrightarrow{\mathbf{W}}_p}^{-1} \mathbf{Z}_p$ should be diagonal, which means $\mathbf{S}_p^* \mathbf{S}_p$ should be diagonal. For (3.37) to be an equality, we need $\mathbf{B}_p^* \mathbf{B}_p = \mathbb{I}_M$. Thus, when $N_p \geq MR$, the optimal design is to have $\mathbf{S}_p^* \mathbf{S}_p = \mathbb{I}_{MR}$, which implies $\mathbf{B}_p^* \mathbf{B}_p = \mathbb{I}_M$ from the unitarity of $\mathbf{A}_{i,p}'s$.

3.8.2 Proof of Theorem 3.3

Recall that from Theorem 3.2, with the imperfect estimation $\hat{\mathbf{G}}$, $\mathbf{R}_{\Delta \mathbf{f}} = \left(\mathbb{I}_R + \beta_p^2 \hat{\mathbf{Z}}_p^* \mathbf{R}_{\overrightarrow{\mathbf{W}}_{e,p}}^{-1} \hat{\mathbf{Z}}_p \right)^{-1}$, where $\hat{\mathbf{Z}}_p \triangleq (\hat{\mathbf{G}}^t \otimes \mathbb{I}_{N_p}) \tilde{\mathbf{S}}_p$ and $\mathbf{R}_{\overrightarrow{\mathbf{W}}_{e,p}} = \frac{\beta_p^2}{1+P} (\mathbb{I}_R \otimes \mathbf{S}_p \mathbf{S}_p^*) + \frac{R \alpha_p^2}{1+P} \mathbb{I}_{RN_p} + (\mathbb{I}_R + \alpha_p^2 \hat{\mathbf{G}}^t \overline{\hat{\mathbf{G}}}) \otimes \mathbb{I}_{N_p}$.

We first consider the case of $N_p = 1$. Using the proposed training code,

$\hat{\mathbf{Z}}_p = \hat{\mathbf{G}}^t$ and $\mathbf{R}_{\overrightarrow{\mathbf{W}}_{e,p}} = k\mathbb{I}_R + \alpha_p^2 \hat{\mathbf{G}}^t \overline{\hat{\mathbf{G}}}$ where $k \triangleq 1 + \frac{R(\alpha_p^2+\beta_p^2)}{1+P} = 2 + O\left(\frac{1}{P}\right)$. Thus

$$\mathbf{R}_{\Delta\mathbf{f}} = \left[\mathbb{I}_R + \beta_p^2 \overline{\hat{\mathbf{G}}} \left(k\mathbb{I}_R + \alpha_p^2 \hat{\mathbf{G}}^t \overline{\hat{\mathbf{G}}}\right)^{-1} \hat{\mathbf{G}}^t\right]^{-1}. \tag{3.38}$$

Using the fact that $\overline{\hat{\mathbf{G}}} \left(k\mathbb{I}_R + \alpha_p^2 \hat{\mathbf{G}}^t \overline{\hat{\mathbf{G}}}\right)^{-1} \hat{\mathbf{G}}^t$ and $\left(k\mathbb{I}_R + \alpha_p^2 \hat{\mathbf{G}}^t \overline{\hat{\mathbf{G}}}\right)^{-1} \hat{\mathbf{G}}^t \overline{\hat{\mathbf{G}}}$ have the same eigenvalues, from (3.38), the eigenvalues of $\mathbf{R}_{\Delta\mathbf{f}}$ are $\frac{k+\alpha_p^2 \hat{\lambda}_{un,i}}{k+(\alpha_p^2+\beta_p^2)\hat{\lambda}_{un,i}}$ for $i = 1, \cdots, R$ with $\hat{\lambda}_{un,i}$'s the unordered eigenvalues of $\hat{\mathbf{G}}\hat{\mathbf{G}}^*$. Thus,

$$\mathrm{MSE}(\hat{\mathbf{f}}) = \mathbb{E}_{\ \hat{\lambda}_{un,i}} \left[\sum_{i=1}^{R} \frac{k + \alpha_p^2 \hat{\lambda}_{un,i}}{k + (\alpha_p^2 + \beta_p^2)\hat{\lambda}_{un,i}}\right]. \tag{3.39}$$

Since entries of $\hat{\mathbf{G}}$ are i.i.d. complex Gaussian, the unordered eigenvalues of the Wishart matrix $\hat{\mathbf{G}}\hat{\mathbf{G}}^*$ have the same PDF, which was derived in [65] to be

$$p_{\hat{\lambda}_{un,i}}(x) = \frac{1}{R}\left(1 + \frac{1}{P}\right) e^{-\left(x+\frac{x}{P}\right)} \left[\sum_{i=0}^{R-1} L_i^2\left(x + \frac{x}{P}\right)\right],$$

where $L_i(x)$ is the Laguerre polynomial of order i defined as

$$L_i(x) = \sum_{k=0}^{i} (-1)^k \binom{i}{i-k} \frac{x^k}{k!}.$$

After some straightforward transformation, when $P \gg 1$, $p_{\hat{\lambda}_{un,i}}(x)$ can be rewritten as

$$p_{\hat{\lambda}_{un,i}}(x) = e^{-x} + \frac{1}{R} e^{-x} \sum_{j=1}^{2(R-1)} r_j x^j + \text{lower order terms in } P, \tag{3.40}$$

where r_j's are constants irrelative of P.

Let $f(x) \triangleq \frac{k+\alpha_p^2 x}{k+(\alpha_p^2+\beta_p^2)x}$. From (3.39), we have

$$\mathrm{MSE}(\hat{\mathbf{f}}) = R\mathbb{E}_{\ \hat{\lambda}_{un,i}} f(\hat{\lambda}_{un,i})$$

$$= R\int_0^\infty f(x)e^{-x}dx + \sum_{j=1}^{2(R-1)} r_j \int_0^\infty f(x)x^j e^{-x}dx + \text{lower order terms in } P. \tag{3.41}$$

First we calculate the second integral in (3.41). For $j \geq 1$,

$$\int_0^\infty f(x)x^j e^{-x}dx = \frac{e^{\frac{k}{\alpha_p^2+\beta_p^2}}\left[k\mathrm{E}_{1+j}\left(\frac{k}{\alpha_p^2+\beta_p^2}\right)\Gamma(1+j) + \alpha_p^2\mathrm{E}_{2+j}\left(\frac{k}{\alpha_p^2+\beta_p^2}\right)\Gamma(2+j)\right]}{\alpha_p^2+\beta_p^2}, \tag{3.42}$$

where $\mathrm{E}_n(x)$ is the exponential integral defined as $\mathrm{E}_n(x) = \int_1^\infty e^{-xt}/t^n\,dt$, and $\Gamma(x)$ is the Gamma function defined as $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt$. Since $\alpha_p^2 = 1/R + O(1/P)$ and $\beta_p^2 = P/R + O(1)$ (notice that $N_p = 1$), it can be shown that when $P \gg 1$ and $j \geq 1$, $\mathrm{E}_{1+j}\left(k/(\alpha_p^2+\beta_p^2)\right) = 1/j + O(1/P)$ and $\mathrm{E}_{2+j}\left(k/(\alpha_p^2+\beta_p^2)\right) = 1/(1+j) + O(1/P)$. So, when $P \gg 1$,

$$\int_0^\infty f(x)x^j e^{-x}dx = \frac{(j-1)!(2R+j)}{P} + O\left(\frac{1}{P^2}\right).$$

Now we calculate the first integral in (3.41).

$$\int_0^\infty f(x)e^{-x}dx = \frac{\alpha_p^2(\alpha_p^2+\beta_p^2) + k\beta_p^2 e^{\frac{k}{\alpha_p^2+\beta_p^2}}\mathrm{E}_1\left(\frac{k}{\alpha_p^2+\beta_p^2}\right)}{(\alpha_p^2+\beta_p^2)^2}. \tag{3.43}$$

When $P \gg 1$, $\mathrm{E}_1\left(\frac{k}{\alpha_p^2+\beta_p^2}\right) = \log P + O(1)$. So,

$$\int_0^\infty f(x)e^{-x}dx = 2R\frac{\log P}{P} + O\left(\frac{1}{P}\right).$$

Applying the results in (3.42) and (3.43) to (3.40), $\mathrm{MSE}(\hat{\mathbf{f}}) = 2R^2\log P/P + O(1/P)$.

Next we consider the case of $N_p \geq R$. We have for $P \gg 1$, $\beta_p^2 = \frac{N_p}{R}P + O(1)$. To prove $\mathrm{MSE}(\hat{\mathbf{f}}) = O(1/P)$, we will find a lower bound and an upper bound on $\mathrm{MSE}(\hat{\mathbf{f}})$ both of order $1/P$.

We start by bounding $\mathbf{R}_{\overrightarrow{\mathbf{W}}_{e,p}}$. It is obvious that $\mathbf{R}_{\overrightarrow{\mathbf{W}}_{e,p}} \succ \mathbb{I}_{RN_p}$. Applying the lower bound on $\mathbf{R}_{\overrightarrow{\mathbf{W}}_{e,p}}$ in $\mathbf{R}_{\Delta\mathbf{f}}$ given in (3.18), $\mathbf{R}_{\Delta\mathbf{f}}$ can be lower bounded as

$$\begin{aligned}\mathbf{R}_{\Delta\mathbf{f}} \succ & (\mathbb{I}_R + \beta_p^2 \hat{\mathbf{Z}}_p^* \hat{\mathbf{Z}}_p)^{-1} \\ = & \text{diag}\left\{(1+\beta_p^2\|\hat{\mathbf{g}}_1\|^2)^{-1}, \cdots, (1+\beta_p^2\|\hat{\mathbf{g}}_R\|^2)^{-1}\right\}. \end{aligned} \tag{3.44}$$

The equality is derived by using the facts that $\hat{\mathbf{Z}}_p$ can be rewritten as $\hat{\mathbf{Z}}_p = (\mathbb{I}_R \otimes \mathbf{S}_p)\left([\hat{\mathbf{G}}_1^t, \cdots, \hat{\mathbf{G}}_R^t]^t\right)$ with $\hat{\mathbf{G}}_i \triangleq \text{diag}\{\hat{g}_{1i}, \cdots, \hat{g}_{Ri}\}$, and $\mathbf{S}_p^*\mathbf{S}_p = \mathbb{I}_R$.

Since $\text{MSE}(\hat{\mathbf{f}}) = \mathbb{E}_{\hat{\mathbf{G}}}(\text{tr}(\mathbf{R}_{\Delta\mathbf{f}}))$, using (3.44), $\text{MSE}(\hat{\mathbf{f}})$ can be lower bounded as follows.

$$\begin{aligned}\text{MSE}(\hat{\mathbf{f}}) &> \mathbb{E}_{\hat{\mathbf{G}}} \sum_{i=1}^{R} \left(1+\beta_p^2\|\hat{\mathbf{g}}_i\|^2\right)^{-1} = \mathbb{E}_{\mathbf{G}} \sum_{i=1}^{R} \left(1+\frac{\beta_p^2 P}{P+1}\|\mathbf{g}_i\|^2\right)^{-1} \\ &= \frac{R^2}{N_p(R-1)}\frac{1}{P} + \text{lower order terms in } P, \end{aligned} \tag{3.45}$$

where (3.45) is derived since $x = \|\mathbf{g}_i\|^2$ has Gamma distribution with degree R ($R \geq 2$). Its PDF is $p_{\|\mathbf{g}_i\|^2}(x) = \frac{1}{\Gamma(R)} x^{R-1} e^{-x}$. Thus, we conclude that the lower bound on $\text{MSE}(\hat{\mathbf{f}})$ scales as $1/P$.

Now, we upper bound $\mathbf{R}_{\overrightarrow{\mathbf{W}}_{e,p}}$. Since $\mathbf{S}_p^*\mathbf{S}_p = \mathbb{I}_R$, we have $\mathbf{S}_p\mathbf{S}_p^* \preceq \mathbb{I}_{N_p}$ and thus $\mathbb{I}_R \otimes (\mathbf{S}_p\mathbf{S}_p^*) \preceq \mathbb{I}_{RN_p}$. Combined with the inequality $\hat{\mathbf{G}}^t\overline{\hat{\mathbf{G}}} \preceq \|\hat{\mathbf{G}}\|_F^2 \mathbb{I}_R$, $\mathbf{R}_{\overrightarrow{\mathbf{W}}_{e,p}}$ can be upper bounded as

$$\begin{aligned}\mathbf{R}_{\overrightarrow{\mathbf{W}}_{e,p}} &\preceq \frac{\beta_p^2}{1+P}\mathbb{I}_{RN_p} + \frac{R\alpha_p^2}{1+P}\mathbb{I}_{RN_p} + \mathbb{I}_{RN_p} + \alpha_p^2\|\hat{\mathbf{G}}\|_F^2\mathbb{I}_{RN_p} \\ &= (k_1 + \alpha_p^2\|\hat{\mathbf{G}}\|_F^2)\mathbb{I}_{RN_p}, \end{aligned}$$

where $k_1 \triangleq 1 + \frac{\beta_p^2 + R\alpha_p^2}{1+P} = \frac{N_p+R}{R} + O\left(\frac{1}{P}\right)$. Applying the upper bound to $\mathbf{R}_{\Delta\mathbf{f}}$ given in (3.18), we have

$$\mathbf{R}_{\Delta\mathbf{f}} \preceq \text{diag}\left\{\left(1+\frac{\beta_p^2}{k_1+\alpha_p^2\|\hat{\mathbf{G}}\|_F^2}\|\hat{\mathbf{g}}_1\|^2\right)^{-1}, \cdots, \left(1+\frac{\beta_p^2}{k_1+\alpha_p^2\|\hat{\mathbf{G}}\|_F^2}\|\hat{\mathbf{g}}_R\|^2\right)^{-1}\right\}.$$

$\mathrm{MSE}(\hat{\mathbf{f}})$ can thus be upper bounded as

$$\begin{aligned}\mathrm{MSE}(\hat{\mathbf{f}}) \leq & \mathbb{E}_{\hat{\mathbf{G}}}\left[\sum_{i=1}^{R}\left(1+\frac{\beta_p^2}{k_1+\alpha_p^2\|\hat{\mathbf{G}}\|_F^2}\|\hat{\mathbf{g}}_i\|^2\right)^{-1}\right] \\ = & \mathbb{E}_{\mathbf{G}}\left\{\sum_{i=1}^{R}\left[1+\frac{\beta_p^2 P/(P+1)}{k_1+\alpha_p^2 P/(P+1)\|\mathbf{G}\|_F^2}\|\mathbf{g}_i\|^2\right]^{-1}\right\}\end{aligned}$$

This has the same form as the pairwise-error-probability (PEP) upper bound formula (30) in [25]. Following the technique in [25], we can show that it scales as $1/P$ when $P \gg 1$. Thus, we conclude that the upper bound on $\mathrm{MSE}(\hat{\mathbf{f}})$ also scales as $1/P$, which completes the proof.

The derivation of the scaling performance of $\mathrm{MSE}(\hat{\mathbf{f}})$ with perfect $\mathbf{G}$ goes the similar lines as above. Since $\mathbf{G}$ is perfectly known, the estimation error $\Delta\mathbf{G} = \mathbf{0}$, and the covariance matrix of the equivalent noise becomes $\mathbf{R}_{\overrightarrow{\mathbf{W}}_{e,p}} = (\mathbb{I}_R + \alpha_p^2 \mathbf{G}^t \overline{\mathbf{G}}) \otimes \mathbb{I}_{N_p}$. Following the derivations above, it can be shown that when $N_p = 1$, $\mathrm{MSE}(\hat{\mathbf{f}}) = R^2 \log P/P + O(1/P)$, when $N_p \geq R$, $\mathrm{MSE}(\hat{\mathbf{f}}) = O(1/P)$.

3.8.3 Proof of Theorem 3.4

Denote d_i as the ith diagonal element of $\mathbf{B}_p^*\mathbf{B}_p$. With the constraint $\mathrm{tr}(\mathbf{B}_p^*\mathbf{B}_p) = M$, there is $d_1 + \cdots + d_M = M$. Lower bound $\mathrm{tr}(\mathbf{R}_{\Delta\overrightarrow{\mathbf{H}},dir})$ as follows.

$$\begin{aligned}& \mathrm{tr}\left[\mathbb{I}_{MNR} + \frac{\beta_p^2}{1+\alpha_p^2 R}(\mathbb{I}_N \otimes \mathbf{S}_p^*\mathbf{S}_p)\right]^{-1} \\ \geq & \sum_{i=1}^{MNR}\left\{\left[\mathbb{I}_{MNR} + \frac{\beta_p^2}{1+\alpha_p^2 R}(\mathbb{I}_N \otimes \mathbf{S}_p^*\mathbf{S}_p)\right]_{ii}\right\}^{-1} \\ = & NR\sum_{j=1}^{M}\frac{1}{1+[\beta_p^2/(1+\alpha_p^2 R)]d_j} \\ \geq & \frac{MNR}{1+\beta_p^2/(1+\alpha_p^2 R)}.\end{aligned}$$

For the first inequality to become an equality, we need $\mathbb{I}_{MNR} + \frac{\beta_p^2}{1+\alpha_p^2 R}(\mathbb{I}_N \otimes \mathbf{S}_p^*\mathbf{S}_p)$ to be a diagonal matrix. This is satisfied when $\mathbf{S}_p^*\mathbf{S}_p = \mathbb{I}_{MR}$, *i.e.*, $N_p \geq MR$

and $\mathbf{S}_p$ is unitary. For the second inequality to become an equality, we need $\mathbf{B}_p^*\mathbf{B}_p = \mathbb{I}_M$, *i.e.* $\mathbf{B}_p$ is unitary. Thus, when $N_p \geq MR$, one optimal design is to have $\mathbf{S}_p^*\mathbf{S}_p = \mathbb{I}_{MR}$.

4 Training-Based Mismatched Decoding and Diversity Analysis

In this chapter, diversity, which is one of the most important indices in wireless communications, of the training-based mismatched decoding is investigated for DSTC network. In Section 4.1, the mismatched decoding DEC_1 is proposed as well as its two benchmarks. In Section 4.2, we first analyze diversity of DEC_1 considering the estimated $\mathbf{G}$ as error-free. Then, in Section 4.3, we take the estimation error of $\mathbf{G}$ into account, and analyze the performance of DEC_1 for the specific network with $M = 1$, and $R = N \geq 2$.

4.1 System Model

The notation below follow those in Section 1.4. Consider the two-step DSTC with $T_d \geq MR$ symbol intervals for each step. Denote the $T_d \times M$ information signal sent from the transmitter as $\mathbf{B}_d$, which satisfies $\mathbb{E}\ (\mathrm{tr}(\mathbf{B}_d^* \mathbf{B}_d)) = M$. Define the distributed space-time codeword as $\mathbf{S}_d \triangleq [\mathbf{A}_{1,d}\mathbf{B}_d \cdots \mathbf{A}_{R,d}\mathbf{B}_d]$. The orthogonal design of $\mathbf{S}_d$ is adopted, so we have $\mathbf{S}_d^* \mathbf{S}_d = \mathbb{I}_{MR}$ [11]. Define $\alpha_d \triangleq \sqrt{\frac{P}{R(P+1)}}$ and $\beta_d \triangleq \sqrt{\frac{P^2 T_d}{MR(P+1)}}$, where P is the transmit power of the transmitter and also the total transmit power of the relays. Following the derivation of system equation in (1.9), the signal received at the receiver can be written as

$$\mathbf{X}_d = \beta_d \mathbf{S}_d \mathbf{H} + \mathbf{W}_d, \tag{4.1}$$

where $\mathbf{W}_d$ is the equivalent noise at the receiver during the data transmission, defined as $\mathbf{W}_d \triangleq \alpha_d \left[\mathbf{A}_{1,d}\mathbf{v}_{1,d} \cdots \mathbf{A}_{R,d}\mathbf{v}_{R,d}\right] \mathbf{G} + \mathbf{W}_{r,d}$ with $\mathbf{v}_{i,d}$ the noise at the i-th relay, $i = 1, \cdots, R$, and $\mathbf{W}_{r,d}$ the noise at the receiver. Stacking the columns of $\mathbf{X}_d$ into one column vector, from (4.1), we have

$$\overrightarrow{\mathbf{X}}_d = \beta_d (\mathbf{G}^t \otimes \mathbb{I}_{T_d}) \tilde{\mathbf{S}}_d \mathbf{f} + \overrightarrow{\mathbf{W}}_d = \beta_d \mathbf{Z}_d \mathbf{f} + \overrightarrow{\mathbf{W}}_d, \tag{4.2}$$

where $\tilde{\mathbf{S}}_d \triangleq \text{diag}\{\mathbf{A}_{1,d}\mathbf{B}_d, \cdots, \mathbf{A}_{R,d}\mathbf{B}_d\}$, and $\mathbf{Z}_d \triangleq (\mathbf{G}^t \otimes \mathbb{I}_{T_d})\tilde{\mathbf{S}}_d$.

In Section 1.4, with the perfect knowledge of CSI, the coherent ML decoding and its simplified version are derived as DEC_0 and $\text{DEC}_{0,simp}$, respectively. When only the channel estimations are available at the receiver, however, the most straightforward decoding is to ignore the estimation errors and replace the channels with their estimations in DEC_0 or $\text{DEC}_{0,simp}$. We call it mismatched decoding since it does not match the training-based system equation due to the neglect of the estimation errors.

Since we adopt separate training scheme, where $\mathbf{G}$ and $\mathbf{f}$ are estimated successively, given the estimations of $\hat{\mathbf{f}}$ and $\hat{\mathbf{G}}$, we have the mismatched decoding

$$\text{DEC}_1 : \arg \min_{\mathbf{B}_d} \left(\overrightarrow{\mathbf{X}}_d - \beta_d \hat{\mathbf{Z}}_d \hat{\mathbf{f}}\right)^* \hat{\mathbf{R}}_{\overrightarrow{\mathbf{W}}_d}^{-1} \left(\overrightarrow{\mathbf{X}}_d - \beta_d \hat{\mathbf{Z}}_d \hat{\mathbf{f}}\right), \tag{4.3}$$

where $\hat{\mathbf{Z}}_d \triangleq (\hat{\mathbf{G}}^t \otimes \mathbb{I}_{T_d})\tilde{\mathbf{S}}_d$ and $\hat{\mathbf{R}}_{\overrightarrow{\mathbf{W}}_d} \triangleq \hat{\mathbf{R}}_{\mathbf{W}_d} \otimes \mathbb{I}_{T_d}$ with $\hat{\mathbf{R}}_{\mathbf{W}_d}$ the estimated covariance matrix of $\mathbf{W}_d$ defined as $\hat{\mathbf{R}}_{\mathbf{W}_d} \triangleq \mathbb{I}_N + \alpha_d^2 \overline{\hat{\mathbf{G}}^* \hat{\mathbf{G}}}$.

As a benchmark, if $\mathbf{G}$ and $\mathbf{H}$ are estimated in the training phase, where the estimation of $\mathbf{H}$ is obtained throught direct training, the mismatched decoding is derived as

$$\text{DEC}_{1,h} : \arg \min_{\mathbf{B}_d} \text{tr}\left(\left(\mathbf{X}_d - \beta_d \mathbf{S}_d \hat{\mathbf{H}}\right) \overline{\hat{\mathbf{R}}}_{\mathbf{W}_d}^{-1} \left(\mathbf{X} - \beta_d \mathbf{S}_d \hat{\mathbf{H}}\right)^*\right), \tag{4.4}$$

Note that, $\text{DEC}_{1,h}$ and DEC_1 are not equivalent, since in $\text{DEC}_{1,h}$, $\hat{\mathbf{H}}$ is obtained through direct training, not equal to $[(\hat{\mathbf{f}}_1 \hat{\mathbf{g}}_1)^t \cdots (\hat{\mathbf{f}}_R \hat{\mathbf{g}}_R)^t]^t$, where $\hat{\mathbf{f}}$ and $\hat{\mathbf{G}}$ are obtained through separate training.

Similar to the perfect CSI case, by approximating $\hat{\mathbf{G}}^t \overline{\hat{\mathbf{G}}}$ with its expectation $\frac{RP}{1+P}\mathbb{I}_N$ in $\hat{\mathbf{R}}_{\mathbf{W}_d}$, a simplified mismatched decoding $\text{DEC}_{1,simp}$ is obtained:

$$\text{DEC}_{1,simp} : \arg \max_{\mathbf{B}_d} \Re\text{tr}(\hat{\mathbf{H}} \mathbf{X}_d^* \mathbf{S}_d). \tag{4.5}$$

For the network with $M = 1$, and $R = N$, (4.5) is equivalent to

$$\mathrm{DEC}_{1,simp} : \arg \max_{s_{d,j}} \Re\, s_{d,j} \left(\sum_{i=1}^{R} \hat{\mathbf{T}}_{b,ii} \right)_j, \quad j = 1, \cdots, T_d, \tag{4.6}$$

where analogous to the definitions in perfect CSI case, $\hat{\mathbf{T}}_{b,ij}$ is the ij-th $1 \times T_d$ block element of $\hat{\mathbf{T}}$ defined as $\hat{\mathbf{T}} \triangleq \hat{\mathbf{H}}\mathbf{X}_d^*[\mathbf{A}_{1,d} \cdots \mathbf{A}_{R,d}]$.

In this chapter, we will discuss in detail about the diversity performance of mismatched decoding DEC_1. Two benchmarks are used to evaluate our proposed training scheme. The first one is DEC_0 with perfect CSI. This ideal case thus provides a lower bound on the error rate. The second one is $\mathrm{DEC}_{1,h}$, in which instead of $\mathbf{f}$, the end-to-end channel matrix $\mathbf{H}$ is estimated by direct training. As discussed in Subsection 3.5.3, by direct training, the entries of $\mathbf{H}$ are treated as if independent and the structure property of $\mathbf{H}$ is not exploited, which may result in longer training interval and performance loss.

4.2 Mismatched Decoding with Perfect G

In this section, we consider that in the mismatched decoding DEC_1, $\hat{\mathbf{G}}$ is error-free, and $\hat{\mathbf{f}}$ is given in (3.7). In Subsection 4.2.1, we show that with $N_p = N_{p,l}$, the lower bound on the training time N_p in estimating $\mathbf{f}$, full diversity cannot be always guaranteed. An upper bound on the minimum training time that guarantees full diversity is thus provided. In Subection 4.2.2, to further shorten the training time while ensure the full diversity performance with DEC_1, an adaptive training design is suggested.

4.2.1 Diversity Analysis

We first provide a lemma. Let λ_n be the minimum eigenvalue of Wishart matrix $\mathbf{W_G}$ (The background on Wishart matrix is referred to Appendix 4.5.1.).

Lemma 4.1. *When $R \leq N$, $N_p = N_{p,l}$, and $P \gg 1$, given $\lambda_n \leq 1/P$, we have $\mathbf{R}_{\Delta \mathbf{f}} \succ \mathbf{C}_{\mathbf{R}_{\Delta \mathbf{f}}}$, where $\mathbf{C}_{\mathbf{R}_{\Delta \mathbf{f}}}$ is a constant positive semi-definite matrix. The*

PEP of the network given $\lambda_n \leq 1/P$ is no smaller than a positive constant c_{PEP}.

Proof. See Appendix 4.5.2. □

Some comments on Lemma 4.1: Recall that the lower bound $N_{p,l}$ is based on the assumption that $\mathbf{G}$ is full rank. In the relay network, $\mathbf{G}$ is random. Although the probability that it is full rank is 1, $\mathbf{G}$ can be arbitrarily close to a singular matrix, i.e., $\lambda_n \leq 1/P (P \gg 1)$. In this case, some training equations in (3.5) will become dependent of others and there are not enough number of independent equations to estimate the channel coefficients. Then as Lemma 4.1 indicates, the estimation error is lower bounded by a constant regardless of the training power, and no diversity can be achieved.

Theorem 4.1. *When $R \leq N$ and $N_p = N_{p,l}$, the achievable diversity of the proposed training-based MIMO relay network is upper bounded by $N - R + 1$.*

Proof. From Lemma 4.3 in Appendix 4.5.1, when $P \gg 1$,

$$\mathbb{P}(\lambda_n \leq 1/P) = c_{\lambda_n} P^{-(N-R+1)} + o(P^{-(N-R+1)}).$$

From Lemma 4.1, we have $(PEP|\lambda_n \leq 1/P) \geq c_{PEP}$. Therefore,

$$\begin{aligned} PEP =& (PEP|\lambda_n \leq 1/P) \cdot \mathbb{P}(\lambda_n \leq 1/P) + (PEP|\lambda_n > 1/P) \cdot \mathbb{P}(\lambda_n > 1/P) \\ \geq& (PEP|\lambda_n \leq 1/P) \cdot \mathbb{P}(\lambda_n \leq 1/P) \\ \geq& c_{PEP} c_{\lambda_n} P^{-(N-R+1)} + o(P^{-(N-R+1)}). \end{aligned}$$

This means that the diversity is upper bounded by $N - R + 1$. □

Based on Theorem 4.1, a sufficient condition for diversity loss when $N_p = N_{p,l}$ is found.

Corollary 4.1. *When $N_p = N_{p,l}$, there is diversity loss if $R > \max(1, \frac{N+1}{M+1})$.*

Proof. The full diversity of the relay networks is $\min(M,N)R$ [25]. Note that $N-R+1 < NR$ is always true for $R>1$. So, $N-R+1 < \min(M,N)R$ is equivalent to $N < (M+1)R-1$ and $R>1$, which can be written as $R > \max(1, \frac{N+1}{M+1})$. □

It is noteworthy that Corollary 4.1 provides a sufficient condition on diversity loss when N_p is set to be its lower bound. Thus, the lower bound on the minimum training time, $N_{p,l}$, in some cases, is insufficient for full diversity. Some examples are the relay networks with $M=R=2$, $N=2,3$, and 4. For these cases, the diversity upper bound $N-R+1$ is shown to be achieved by simulation.

Next we provide an upper bound on the minimum training time that guarantees full diversity. The following theorem is proved.

Theorem 4.2. *When $N_p \geq MR$, the proposed training scheme achieves full diversity.*

Proof. See Appendix 4.5.3. □

Intuitively, when $N_p \geq MR$, there are always enough number of independent equations in (3.5) regardless of the quality of the relay-receiver channels. Thus full diversity can be achieved.

Corollary 4.2. *When $N=1$ or $R=1$, the minimum requirement on N_p that guarantees full diversity is MR.*

Proof. Since $N_{p,l} = \max(M, \lceil MR/N \rceil) = MR \Longleftrightarrow N=1$ or $R=1$, this can be obtained directly from Theorem 4.2. □

Intuitively, for full diversity, N_p should be large enough to render enough independent training equations for solving the unknown channels. Theorems 4.1 and 4.2 comply with this intuition. Note that for the direct training scheme, MR is a lower bound on N_p. We have shown in Theorem 4.2 that it is an upper bound for the proposed scheme. For networks with a single receive antenna

Algorithm 3 An adaptive-N_p design.

1: Choose a positive threshold ϵ.
2: Let the index i be the number of eigenvalues of $\mathbf{W_G}$ that are no less than ϵ. If $i = 0$, let $i = 1$.
3: Let $N_p = N_{p,i} \triangleq \lceil MR/i \rceil$.

or single relay, the proposed scheme does not have advantage on training time over the direct training scheme. For other network settings, however, the proposed scheme may have a lower requirement on the training length as will be shown in the simulation section.

4.2.2 Adaptive Training Time Design

From the previous subsection we know that, full diversity in data transmission cannot always be obtained with $N_{p,l}$, and to guarantee full diversity, one solution is to increase N_p to MR. Nevertheless, except for $R = 1$ or $N = 1$, MR is always larger than $N_{p,l}$. For large networks, it can be undesirable in practice to have $N_p = MR$, especially for networks with a fast fading environment. Also, it is always desirable to reduce N_p to save resource for the data transmission. In this section, we provide an adaptive-N_p design which ensures full diversity with a shorter training time than MR.

The idea of the adaptive-N_p design is as follows. When the rank of $\mathbf{G}$ is i, $i = 1, \cdots, n$, with the optimal training code design, the rank of $\mathbf{Z}_p$ is $\min(iN_p, MR)$, which is also the number of independent equations in (3.5). Thus a reliable estimation of $\mathbf{f}$ requires $N_p \geq \lceil MR/i \rceil$. Note that the rank of $\mathbf{G}$ equals the number of non-zero eigenvalues of $\mathbf{W_G}$. Thus, our strategy is when some eigenvalues of $\mathbf{W_G}$ are not large enough, we treat them as zero and increase the training time correspondingly.

The proposed adaptive training time design is presented in Algorithm 3. To perform the design, the receiver needs to feed back the index i, which is the number of eigenvalues of $\mathbf{W_G}$ that are no less than a threshold, to the transmitter and the relays. The number of bits needed to represent the

index information is $\log_2 n$. The feedback is thus low rate. For example, for networks with 10 relays and 2 receive antennas, only 1 bit feedback is needed. For networks where feedback links do not exist or are too expensive, we can resort to the fixed design and let $N_p = MR$ to guarantee full diversity. Also, when $R = 1$ or $N = 1$, from Algorithm 3, N_p is always MR, and the adaptive design reduces to the fixed training time design.

In general, N_p is a random variable whose value depends on the quality of $\mathbf{G}$. The average training time $\mathbb{E}\ (N_p)$ is calculated as follows.

Lemma 4.2. *With the adaptive-N_p training scheme in Algorithm 3,*

$$\mathbb{E}\ (N_p) = \begin{cases} MR, & n = 1 \\ F_{\lambda_2}(\epsilon)MR + (1 - F_{\lambda_2}(\epsilon))\, N_{p,2}, & n = 2\,. \\ F_{\lambda_2}(\epsilon)MR + \sum_{i=2}^{n-1} F_{i\boldsymbol{\lambda}}(\epsilon) N_{p,i} + (1 - F_{\lambda_n}(\epsilon))\, N_{p,n}, & n \geq 3 \end{cases} \quad (4.7)$$

$F_{\lambda_2}(\epsilon)$ and $F_{\lambda_n}(\epsilon)$ are given in (4.14) and (4.15) respectively, and $F_{i\boldsymbol{\lambda}}(\epsilon) = F_{\lambda_{i+1}}(\epsilon) - F_{\lambda_i}(\epsilon)$.

Proof. The proof is straightforward using the CDF of the eigenvalues of $\mathbf{W}_{\mathbf{G}}$. □

For $n \geq 2$, $\mathbb{E}\ (N_p) < MR$ for any finite ϵ. Using this adaptive-N_p design, one can control the balance between the training time interval and the estimation quality through the design of ϵ. With a higher ϵ, more reliable channel estimation can be obtained, but more time is dedicated to training and less time is dedicated to the data transmission, and vice versa. Simulation shows that with the adaptive-N_p design, full diversity is always achieved even with a training time slightly larger than $N_{p,l}$.

As straightforward results from Lemma 4.2, when $M = 1, R = N = 2$, we have $\mathbb{E}\ (N_p) = 2 - e^{-2\epsilon}$; when $M = R = 2, N = 3$, we have $\mathbb{E}\ (N_p) = e^{-2\epsilon}\epsilon^2(2 - \epsilon)$. Performance of these two networks is simulated in Section 4.2.3.

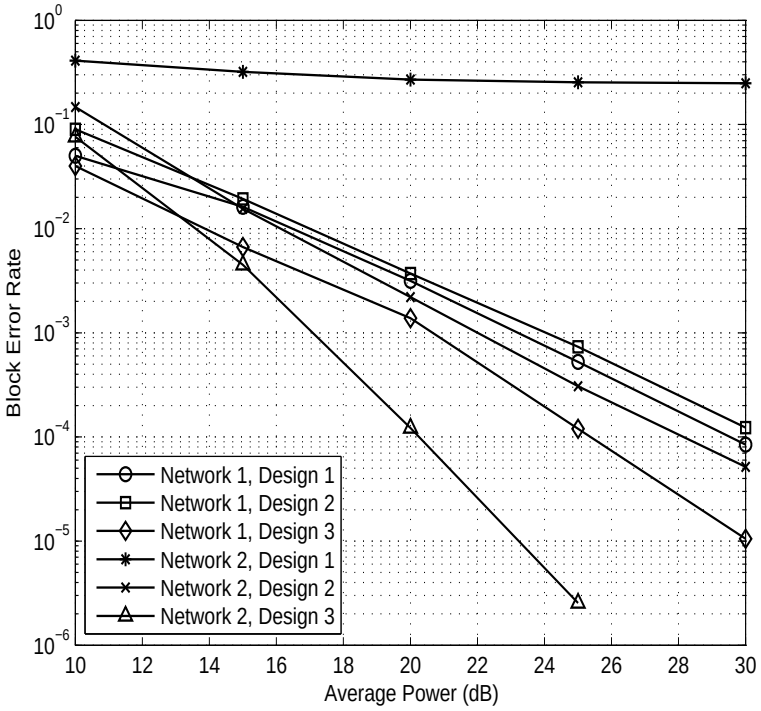

Figure 4.1: Comparison of training code designs. Network 1: $M = 1, R = N = 2$; Network 2: $M = 2, N = R = 2$.

4.2.3 Simulation Results

In this section, applying the mismatched decoding DEC_1, we show the simulated BLERs of the proposed training-based network as a function of P, the average power in the network. For all simulations, the estimated $\mathbf{G}$ is used in DEC_1 for decoding. Orthogonal distributed space-time code with BPSK modulated information symbols is applied for the data transmission [11], e.g., when $R = 2$, Alamouti code is used.

In Fig. 4.1, we verify the optimality of the proposed training code design when the mismatched decoding DEC_1 is used. Two networks are considered: Network 1 in which $M = 1, N = R = 2$; and Network 2 in which $M = N = R = 2$. N_p is set to be MR. We compare the proposed optimal pilot design, denoted as Design 3, with Design 1, in which $\mathbf{B}_p$ is designed only with the trace constraint (Particularly, for Network 1, $\mathbf{B}_p = \sqrt{2}/2[1\ 1]^t$; for

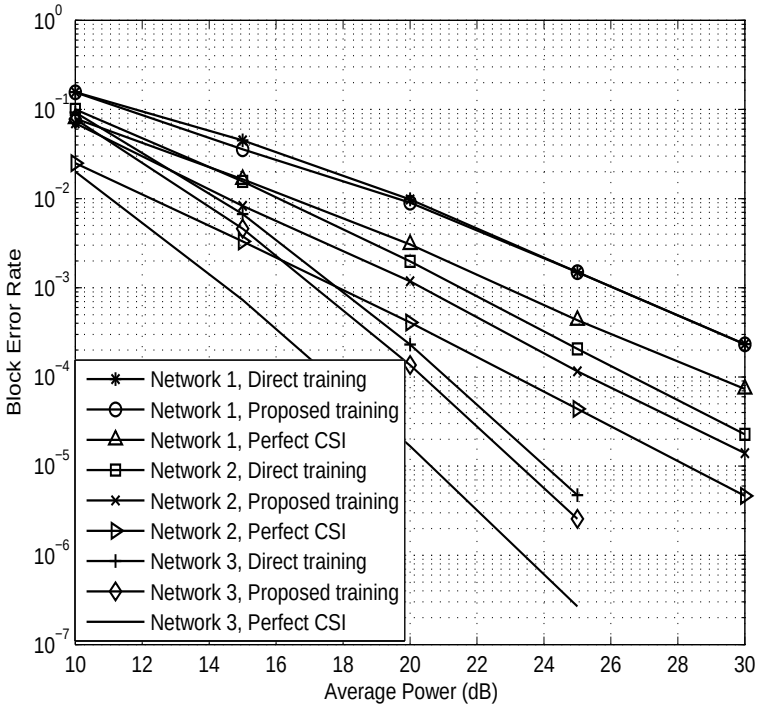

Figure 4.2: Comparison of direct training, separate training, and the perfect CSI case. Network 1: $M = N = 1, R = 2$; Network 2: $M = 1, R = N = 2$; Network 3: $M = R = N = 2$.

Network 2, $\mathbf{B}_p = 1/2 \begin{bmatrix} 1 & 1 & 1 & 1 \\ & & & \\ 1 & 1 & 1 & 1 \end{bmatrix}^t$.) and $\mathbf{A}_{i,p}$'s are randomly generated unitary matrices, and Design 2, in which $\mathbf{B}_p$ is generated optimally as $[\mathbb{I}_M \ \mathbf{0}_{M,MR-M}]^t$ and $\mathbf{A}_{i,p}$'s are random unitary matrices. Fig. 4.1 shows that in both networks, our design outperforms the other two and achieves full diversity, while both Designs 1 and 2 result in diversity loss. This shows that optimal training code design is important for full diversity.

In Fig. 4.2, we compare the performance of the proposed separate training, direct training, and the perfect CSI case. The decoding metrics employed are DEC_1, $\text{DEC}_{1,h}$, and DEC_0, respectively. For the two training schemes, we have $N_p = MR$, as it is required for the direct training. Three networks are considered: Network 1 in which $M = N = 1, R = 2$, Network 2 in which $M = 1, R = N = 2$, and Network 3 in which $M = R = N = 2$. For Network 1,

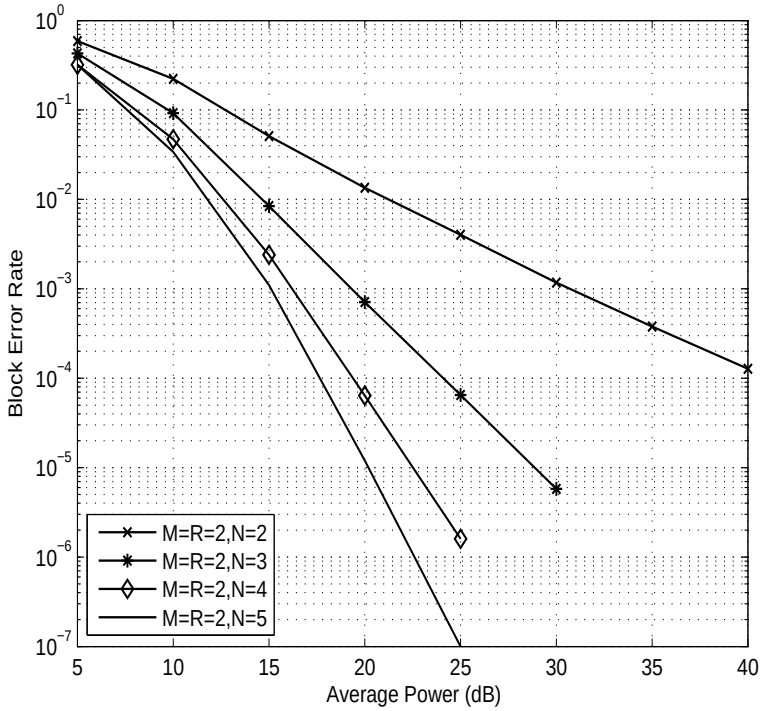

Figure 4.3: BLER performance of the training-based scheme with $R \leq N$ and $N_p = N_{p,l}$.

since $M = N = 1$, all the elements in $\mathbf{H}$ are independent. The direct training and the proposed training have about the same performance with the latter slightly better in the low to mediate SNR region. For Networks 2 and 3, since the elements in $\mathbf{H}$ are correlated, the direct training, in which entries of $\mathbf{H}$ are treated as independent, is not optimal. Fig. 4.2 shows that it is inferior to the proposed training by about 1dB. The proposed training is about 3dB, 2dB, and 2.5dB worse than the perfect CSI case in Network $1, 2$, and 3, respectively due to the estimation error.

In Fig. 4.3, we show the diversity performance of mismatched decoding DEC_1 with $N_p = N_{p,l}$. We consider networks in which $M = R = 2$ and N changes from 2 to 5. Thus, $N_{p,l} = 2$ for all the cases. Fig. 4.3 shows that the achieved diversities of the proposed scheme are about $1, 2, 3$, and 4 for $N = 2, 3, 4$, and 5, respectively. This confirms the conclusion in Theorem 4.1

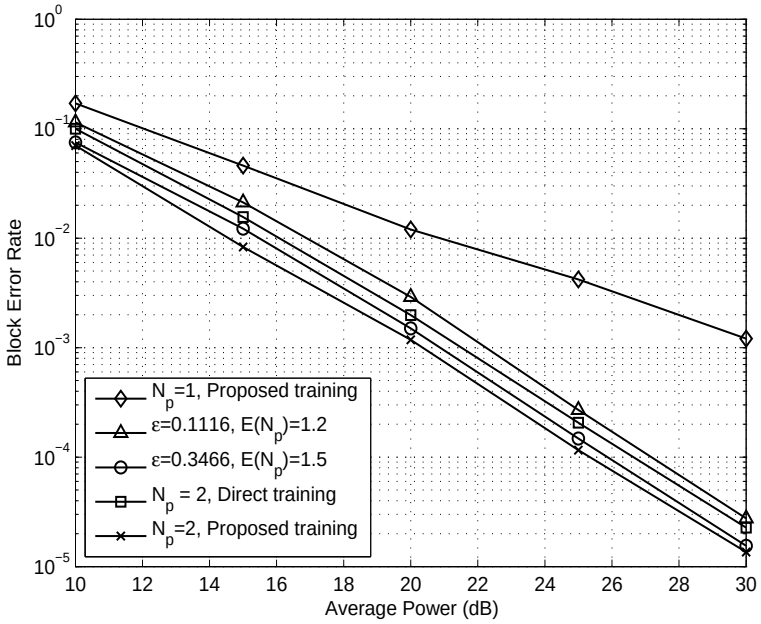

Figure 4.4: Adaptive-N_p design for network with $M = 1, N = R = 2$.

that when $R \leq N$ and $N_p = N_{p,l}$, the diversity is upper bounded by $N-R+1$. It also shows that this upper bound is tight for the simulated networks. Since the full diversity of all the networks is 4, for $N = 2, 3$, and 4, there is diversity loss. This observation is also consistent with our result in Corollary 4.1 that a sufficient condition for diversity loss is $R > \max(1, \frac{N+1}{M+1})$.

Figs. 4.4 and 4.5 show the performance of the adaptive training design. Fig. 4.4 is on the network in which $M = 1, R = N = 2$. The full diversity is 2. As shown in Fig. 4.4, when $N_p = N_{p,l} = 1$, the achieved diversity is only 1. By employing the adaptive training design, even with teh average training time $\mathbb{E}\ (N_p) = 1.2$, the diversity becomes 2, which shows the effectiveness of this design. When $\mathbb{E}\ (N_p) = 1.5$, the adaptive design is superior to the direct training design with $N_p = 2$ by 0.8dB, and close to the proposed fixed training time design with $N_p = 2$ in the high SNR region.

Fig. 4.5 is on the network in which $M = R = 2, N = 3$. Similar observations can be seen. The full diversity of this network is 4. With $N_p = N_{p,l} =$

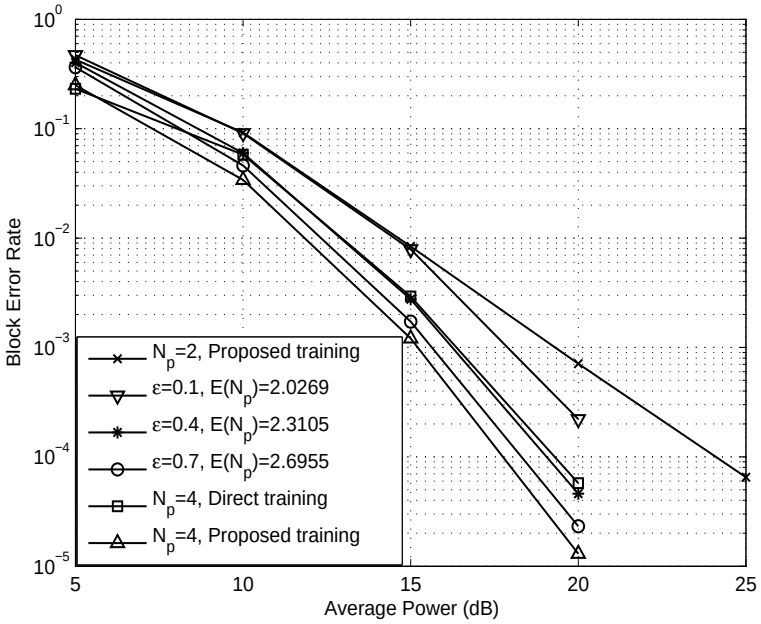

Figure 4.5: Adaptive-N_p design for network with $M = R = 2, N = 3$.

2, the achievable diversity is 2. When the adaptive training design is used, full diversity is obtained for all used ϵ values. Especially, when $\epsilon = 0.1$ and $E(N_p) = 2.0269$, full diversity is achieved at the cost of only 1% increase in training time. It is worth mentioning that with $\mathbb{E}\,(N_p) = 2.3105$, the adaptive training design outperforms the direct training design with $N_p = 4$ in the high SNR region. When $\mathbb{E}\,(N_p) = 2.6955$, the performance of the adaptive design is within 1dB of that of our proposed fixed training time design with $N_p = 4$. This improvement comes from the feedback information explained in Section 4.2.2. To use the adaptive training design, for both networks, 1 bit needs to be fed back from the receiver to both the relays and the transmitter.

4.3 Mismatched Decoding with Imperfect G

In this section, we consider the transmitter-relay channel vector $\mathbf{f}$ is estimated with the help of the estimated $\mathbf{G}$, as discussed in Section 3.4. In other words, to perform DEC_1, $\hat{\mathbf{G}}$ and $\hat{\mathbf{f}}$ given in (3.2) and (3.16) respectively are used. To

use the analytical results in Section 3.4, here, we restrict our attention to the network with $M = 1$, and $R = N \geq 2$, and only the fixed training time design is considered.

4.3.1 Diversity Analysis

We know from Theorem 4.1 that, when the estimated $\hat{\mathbf{G}}$ is error-free, with $N_p = N_{p,l} = 1$, the achieved diversity order is no larger than 1, and from Theorem 4.2 that, with $N_p = R$, full diversity can be achieved. In the following, we show that, by considering the estimation error of $\mathbf{G}$, the same diversity results can be obtained.

In analyzing the performance of decoding DEC_1, we replace $\mathbf{G}$ and $\mathbf{f}$ in the transmission equation (4.2) by $\hat{\mathbf{G}} + \Delta\mathbf{G}$ and $\hat{\mathbf{f}} + \Delta\mathbf{f}$, respectively, to obtain the following training-based transmission equation:

$$\overrightarrow{\mathbf{X}}_d = \beta_d \hat{\mathbf{Z}}_d \hat{\mathbf{f}} + \overrightarrow{\mathbf{W}}_{e,d}, \tag{4.8}$$

where $\overrightarrow{\mathbf{W}}_{e,d}$ is the noise plus estimation error term defined as

$$\begin{aligned}\overrightarrow{\mathbf{W}}_{e,d} \triangleq \beta_d \big[\hat{\mathbf{Z}}_d \Delta\mathbf{f} + (\Delta\mathbf{G}^t \otimes \mathbb{I}_{T_d}) \tilde{\mathbf{S}}_d \hat{\mathbf{f}} + (\Delta\mathbf{G}^t \otimes \mathbb{I}_{T_d}) \tilde{\mathbf{S}}_d \Delta\mathbf{f} \big] \\ + \overrightarrow{\mathbf{W}}_{g1,d} + \overrightarrow{\mathbf{W}}_{g2,d} + \overrightarrow{\mathbf{W}}_{r,d},\end{aligned} \tag{4.9}$$

with

$$\mathbf{W}_{g1,d} \triangleq \alpha_d \left[\mathbf{A}_{1,d}\mathbf{v}_{1,d} \cdots \mathbf{A}_{R,d}\mathbf{v}_{R,d} \right] \hat{\mathbf{G}} \tag{4.10}$$

and

$$\mathbf{W}_{g2,d} \triangleq \alpha_d \left[\mathbf{A}_{1,d}\mathbf{v}_{1,d} \cdots \mathbf{A}_{R,d}\mathbf{v}_{R,d} \right] \Delta\mathbf{G}. \tag{4.11}$$

From the definition of $\overrightarrow{\mathbf{W}}_{e,d}$ we can see that it is not Gaussian, which makes further analysis intractable. In what follows, we study the performance of

DEC_1 using the MSE results in Theorem 3.3. We consider the behaviour of $\overrightarrow{\mathbf{W}}_{e,d}$ for different values of N_p, corresponding to different lengths for training.

The first to consider is the case of $N_p = N_{p,l} = 1$, corresponding to the shortest training length $R+2$. Theorem 3.3 shows that when $N_p = 1$, in the high transmit power range, i.e, $P \gg 1$, $\text{MSE}(\hat{\mathbf{f}})$ scales as $\log P/P$. From Section 3.2 we know that $\text{MSE}(\hat{\mathbf{G}})$ scales as $1/P$. As $\beta_d^2 = \frac{T_d}{R}P + O(1)$ and $\alpha_d^2 = 1/R + O(1/P)$, the highest order of the average power in $\overrightarrow{\mathbf{W}}_{e,d}$ is $\log P$, which is contributed by the term $\beta_d \hat{\mathbf{Z}}_d \Delta\mathbf{f}$. The second highest order term in $\overrightarrow{\mathbf{W}}_{e,d}$ scales as 1. It is then reasonable to ignore some of the lower order terms in P and approximate $\overrightarrow{\mathbf{W}}_{e,d}$ as $\overrightarrow{\mathbf{W}}'_{e,d} \triangleq \beta_d \hat{\mathbf{Z}}_d \Delta\mathbf{f} + \overrightarrow{\mathbf{W}}_{r,d}$. An approximate system model is derived as

$$\overrightarrow{\mathbf{X}}_d = \beta_d \hat{\mathbf{Z}}_d \hat{\mathbf{f}} + \overrightarrow{\mathbf{W}}'_{e,d}. \tag{4.12}$$

It can be shown straightforwardly that $\overrightarrow{\mathbf{W}}'_{e,d}$ has zero-mean. Its covariance matrix is calculated to be

$$\mathbf{R}_{\overrightarrow{\mathbf{W}}'_{e,d}} = \beta_d^2 \hat{\mathbf{Z}}_d \mathbf{R}_{\Delta\mathbf{f}} \hat{\mathbf{Z}}_d^* + \mathbb{I}_{T_d R}. \tag{4.13}$$

By further treating $\Delta\mathbf{f}$ as if Gaussian for tractable analysis, the following diversity result is proved.

Theorem 4.3. *With the approximate system model in (4.12), when $N_p = 1$, the achieved diversity of DEC_1 is no larger than* 1.

Proof. See Appendix 4.5.4. □

Although the diversity result in Theorem 4.3 is derived with an approximate system equation, its validity is shown by simulation. Thus, with the minimum training period $R+2$, i.e., $N_p = 1$, mismatched decoding DEC_1 loses diversity. This can be explained via the behaviour of the MSE of the channel estimations. Through the derivation of the system model (4.12), we know when $N_p = 1$, the equivalent noise $\overrightarrow{\mathbf{W}}_{e,d}$ in data transmission is dominated by the term related to

the estimation error on $\mathbf{f}$, $\Delta\mathbf{f}$, which scales as $\log P$. Notice that the remaining noise power scales as 1 and the signal power scales as P. Treating $\hat{\mathbf{f}}$ as if perfect ignores the $\Delta\mathbf{f}$ term, hence induces diversity loss.

Now we consider the case of $N_p \geq R$. The total training length is thus no shorter than $3R$. In this case, from Theorem 3.3, all estimation error related terms in $\overrightarrow{\mathbf{W}}_{e,d}$ have the average power in the order of 1 or lower. Thus, ignoring the estimation error in decoding will not degrade the network diversity. Our simulation shows that in this case, full diversity can be achieved with mismatched decoding.

For the case of $1 < N_p < R$, the MSE and diversity analysis is even more challenging due to the concatenation of channel matrices and DSTC matrix. Our simulation on limited network scenarios shows that the mismatched decoding DEC_1 cannot achieve full diversity. Hence, with DEC_1, at least $3R$ symbol intervals are needed for training to ensure full diversity.

4.3.2 Simulation Results

In this subsection, simulation results are exhibited. BPSK is used for data transmission for all simulated networks. For two-relay networks, i.e., $R = 2$, Alamouti code is used, while for three-relay networks, i.e., $R = 3$, a generalized rate-1 real orthogonal code is used [1].

In Fig. 4.6, the performance of mismatched decoding DEC_1 and perfect CSI decoding DEC_0 are shown for the network with $R = 2$. We observe that when $N_p = 1$, DEC_1 achieves diversity 1 only, which conforms with our analytical result in Theorem 4.3. When $N_p = 2$, it achieves full diversity 2. Compared with the perfect CSI decoding at the BLER of 10^{-4}, when $N_p = 1$, DEC_1 is about 17dB worse; when $N_p = 2$, DEC_1 is about 2.5dB worse.

In Fig. 4.7, the performance of mismatched decoding DEC_1 and perfect CSI decoding DEC_0 are shown for the network with $R = 3$. It can be observed that when $N_p = 1$, the diversity of DEC_1 is only 1, which is consistent with Theorem 4.3. When $N_p = 2$, $\text{DEC}_{1,simp}$ achieves diversity 2. When $N_p = 3$, it

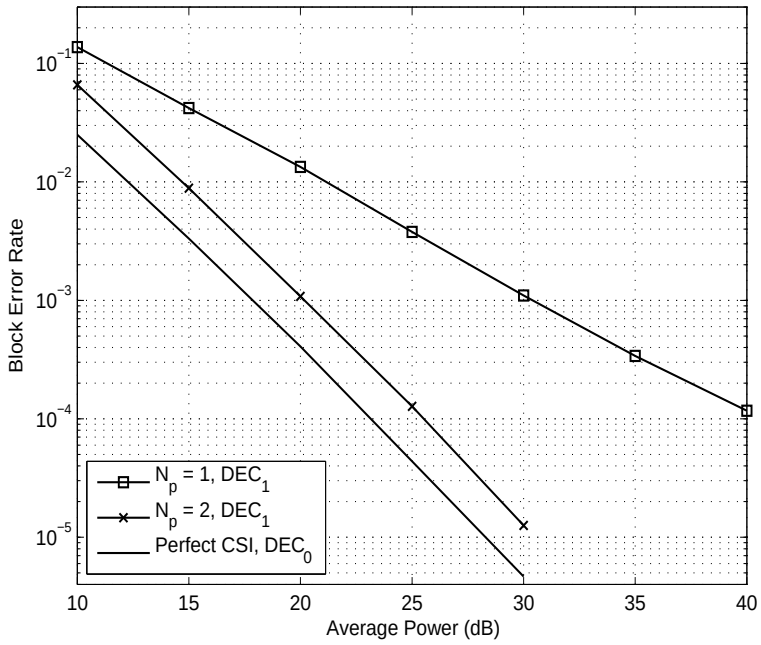

Figure 4.6: DEC_1 for the network with $M = 1, R = N = 2$.

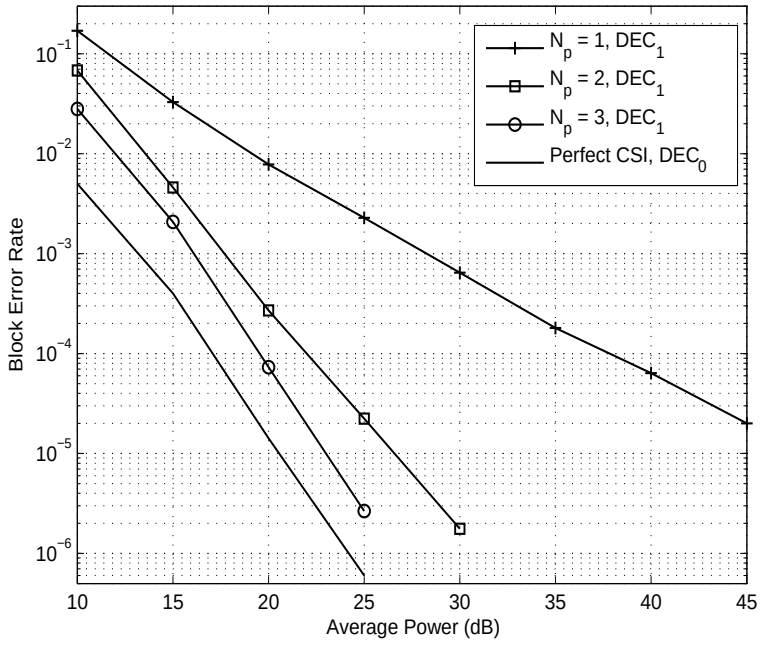

Figure 4.7: DEC_1 for the network with $M = 1, R = N = 3$.

achieves the full diversity, which is 3. Thus, for network with $R = 3$, at least 9 symbol intervals are required for full diversity if mismatched decoding DEC_1 is applied. The perfect CSI decoding DEC_0 is shown as a benchmark. Due to the imperfect training, when $N_p = 3$, the degradation of DEC_1 is about 2.5dB.

4.4 Summary

In this chapter, for the training-based DSTC network, we analyzed the diversity performance of mismatched decoding DEC_1 where the relay-receiver channel matrix $\mathbf{G}$ and the transmitter-relay channel vector $\mathbf{f}$ are estimated in the training phase with the total training time $R + 2N_p$.

In Section 4.1, the mismatched decoding DEC_1 was proposed as well as its two benchmarks. In Section 4.2, under the assumption that the estimation of $\mathbf{G}$ is error-free, the diversity performance of DEC_1 was analyzed. It was shown in Theorem 4.1 that, with $N_p = N_{p,l}$, the lower bound on N_p in estimating $\mathbf{f}$, full diversity cannot always be achieved in data transmission. A sufficient condition which can result in diversity loss with $N_p = N_{p,l}$ was provided in Corollary 4.1. To achieve full diversity, an upper bound on the minimum N_p was given in Theorem 4.2 as MR. To shorten the training time while maintain the full diversity, an adaptive training was provided whose training time length is adaptive to the relay-receiver channel matrix $\mathbf{G}$. In Section 4.3, by considering the estimation error of $\mathbf{G}$, the diversity performance of DEC_1 was analyzed for the specific network with $M = 1$ and $R = N \geq 2$. It was shown that, with the shortest training length $R + 2$ symbol intervals, the diversity order is no larger than 1, and at least $3R$ symbol intervals are required in training to achieve full diversity.

4.5 Appendices

4.5.1 Useful Results on Wishart Matrix

In this subsection, background on Wishart matrix is reviewed and a result on the smallest eigenvalue of Wishart matrix is derived, which is useful for the diversity analysis in this chapter.

Consider the $R \times N$ matrix $\mathbf{G}$ whose entries are *i.i.d.* $\mathcal{CN}(0,1)$. Let

$$\mathbf{W_G} = \begin{cases} \mathbf{G}^*\mathbf{G} \text{if } R \geq N \\ \mathbf{G}\mathbf{G}^* \text{if } R < N \end{cases}.$$

$\mathbf{W_G}$ is a Wishart matrix. Let $n \triangleq \min(R,N)$ and $m \triangleq \max(R,N)$. Order the eigenvalues of $\mathbf{W_G}$ as $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$. The cumulative distribution function (CDF) of λ_k is [53]:

$$F_{\lambda_k}(\epsilon) = K \sum_{i=1}^{k} \sum_{\boldsymbol{\mu} \in \mathcal{P}(i)} \det\left(\mathbf{F}(\boldsymbol{\mu}, i; \epsilon)\right), \tag{4.14}$$

where $K = \prod_{i=1}^{n} \frac{1}{(m-i)!(n-i)!}$, $\mathcal{P}(i)$ is the set of all permutations $(\mu_1, \cdots, \mu_n)$ of $(1, \cdots, n)$ that satisfy $\mu_1 < \cdots < \mu_{i-1}$ and $\mu_i < \cdots < \mu_n$, and $\mathbf{F}(\boldsymbol{\mu}, i; \epsilon)$ is an $n \times n$ matrix whose (u,v)th entry is defined as

$$[\mathbf{F}(\boldsymbol{\mu}, i; \epsilon)]_{u,v} \triangleq \begin{cases} \int_{\epsilon}^{\infty} e^{-\lambda} \lambda^{m-n+u+v-2} d\lambda \text{if } 1 \leq \mu_v < i \\ \int_{0}^{\epsilon} e^{-\lambda} \lambda^{m-n+u+v-2} d\lambda \text{ if } i \leq \mu_v \leq n \end{cases}.$$

The CDF of the smallest eigenvalue λ_n can be further simplified to [53]:

$$F_{\lambda_n}(\epsilon) = 1 - K \det\left(\mathbf{F}(\epsilon)\right), \tag{4.15}$$

where $\mathbf{F}(\epsilon)$ is an $n \times n$ matrix with $\int_{\epsilon}^{\infty} e^{-\lambda} \lambda^{m-n+u+v-2} d\lambda$ being its (u,v)th entry. The probability density function (PDF) of λ_n is also derived in [54] as:

$$f_{\lambda_n}(\lambda) = c_{m,n} \lambda^{m-n} e^{-\lambda n} (b_0 + \cdots + b_{(m-n)(n-1)} \lambda^{(m-n)(n-1)}), \tag{4.16}$$

where $c_{m,n}, b_0, b_1, \cdots, b_{(m-n)(n-1)}$ are constants and $b_0 > 0$. We derive in the following lemma the behaviour of the PDF of λ_n for small ϵ, which will be used in our diversity analysis.

Lemma 4.3. *For small $\epsilon > 0$, $\mathbb{P}(\lambda_n \leq \epsilon) = c_{\lambda_n}\epsilon^{m-n+1} + o\left(\epsilon^{m-n+1}\right)$, where c_{λ_n} is a positive constant independent of ϵ.*

Proof. Using (4.16), we have

$$\begin{aligned}
&\mathbb{P}(\lambda_n \leq \epsilon) = \int_0^{\epsilon} c_{m,n}\lambda^{m-n}e^{-\lambda n}(b_0 + \cdots + b_{(m-n)(n-1)}\lambda^{(m-n)(n-1)})d\lambda \\
=&c_{m,n}\left(\int_0^{\epsilon} b_0\lambda^{m-n}e^{-\lambda n}d\lambda + \cdots + \int_0^{\epsilon} b_{(m-n)(n-1)}\lambda^{m-n+(m-n)(n-1)}e^{-\lambda n}d\lambda\right) \\
=&c_{m,n}\Big(b_0 n^{-1+n-m}\gamma(1+(m-n), n\epsilon) + \ldots \\
&\qquad + b_{(m-n)(n-1)}n^{-1-mn+n^2}\gamma(1+(m-n)n, n\epsilon)\Big), \quad (4.17)
\end{aligned}$$

where $\gamma(.,.)$ is the lower incomplete gamma function defined as $\gamma(s,x) \triangleq \int_0^x t^{s-1}e^{-t}dt$. Using the fact that $\gamma(s,x) \to \frac{1}{s}x^s$ as $x \to 0$, we have $\gamma(i+m-n, n\epsilon) \to \frac{n^{i+m-n}}{i+m-n}\epsilon^{i+m-n}$ as $\epsilon \to 0$. Therefore, for small $\epsilon > 0$, the first summand of (4.17) dominates, and finally $\mathbb{P}(\lambda_n \leq \epsilon) = c_{\lambda_n}\epsilon^{m-n+1} + o(\epsilon^{m-n+1})$, where $c_{\lambda_n} = c_{m,n}b_0/(1+m-n)$. □

4.5.2 Proof of Lemma 4.1

First we prove that $\mathrm{Cov}(\Delta\mathbf{f}) \succ \mathbf{C}_{\mathrm{Cov}(\Delta\mathbf{f})}$. As $R \leq N, N_{p,l} = M \leq MR$. When $N_p = N_{p,l}$, we use Algorithm 2 for the pilot design so $\mathbf{B}_p = \mathbb{I}_M$, which leads to $\tilde{\mathbf{S}}_p\tilde{\mathbf{S}}_p^* = \tilde{\mathbf{S}}_p^*\tilde{\mathbf{S}}_p = \mathbb{I}_{MR}$. Take the singular value decomposition (SVD) of $\mathbf{G}^t$ as $\mathbf{G}^t = \mathbf{U}\mathbf{\Sigma}\mathbf{V} = \mathbf{U}\mathrm{diag}\{\lambda_1^{1/2}, \cdots, \lambda_R^{1/2}, \mathbf{0}_{N-R,R}\}\mathbf{V}$, where $\mathbf{U}$ and $\mathbf{V}$ are square unitary matrices and λ_i is the ith largest eigenvalue of $\mathbf{W}_{\mathbf{G}}$. After some algebra,

$$\begin{aligned}
\mathrm{Cov}(\Delta\mathbf{f}) =& \left[\mathbb{I}_{MR} + \beta_p^2\tilde{\mathbf{S}}_p^*(\overline{\mathbf{G}} \otimes \mathbb{I}_{N_p})\left((\mathbb{I}_N + \alpha_p^2\mathbf{G}^t\overline{\mathbf{G}})^{-1} \otimes \mathbb{I}_{N_p}\right)(\mathbf{G}^t \otimes \mathbb{I}_{N_p})\tilde{\mathbf{S}}_p\right]^{-1} \\
=& \left[\tilde{\mathbf{S}}_p^*\left(\left(\mathbb{I}_R + \beta_p^2\mathbf{V}^*\mathbf{D}\mathbf{V}\right) \otimes \mathbb{I}_{N_p}\right)\tilde{\mathbf{S}}_p\right]^{-1}
\end{aligned}$$

$$=\tilde{\mathbf{S}}_p^* \left[\left(\mathbf{V}^* \left(\mathbb{I}_R + \beta_p^2 \mathbf{D} \right)^{-1} \mathbf{V} \right) \otimes \mathbb{I}_{N_p} \right] \tilde{\mathbf{S}}_p,$$

where $\mathbf{D} \triangleq \mathbf{\Sigma}^* (\mathbb{I}_N + \alpha_p^2 \mathbf{\Sigma}\mathbf{\Sigma}^*)^{-1} \mathbf{\Sigma}$. Note that

$$(\mathbb{I}_R + \beta_p^2 \mathbf{D})^{-1} = \text{diag} \left\{ \frac{1 + \alpha_p^2 \lambda_1}{1 + (\alpha_p^2 + \beta_p^2)\lambda_1}, \cdots, \frac{1 + \alpha_p^2 \lambda_n}{1 + (\alpha_p^2 + \beta_p^2)\lambda_n} \right\}.$$

Since $\frac{1+\alpha_p^2 x}{1+(\alpha_p^2+\beta_p^2)x}$ is a decreasing function of x, given $\lambda_n \leq 1/P$, we have

$$\frac{1 + \alpha_p^2 \lambda_n}{1 + (\alpha_p^2 + \beta_p^2)\lambda_n} \geq \frac{1 + \alpha_p^2 \frac{1}{P}}{1 + (\alpha_p^2 + \beta_p^2)\frac{1}{P}} = \frac{RP + 1 + R}{RP + R + 1 + P} = \frac{R}{1+R} + O(\frac{1}{P}).$$

Therefore, when $P \gg 1$, $(\mathbb{I}_R + \beta_p^2 \mathbf{D})^{-1} \succ \text{diag}\{0, \cdots, 0, \frac{R}{1+R}\}$, which means that $\text{Cov}(\Delta \mathbf{f})$ is lower bounded by

$$\mathbf{C}_{\text{Cov}(\Delta \mathbf{f})} \triangleq \tilde{\mathbf{S}}_p^* \left[(\mathbf{V}^* \text{diag}\{0, \cdots, 0, \frac{R}{1+R}\} \mathbf{V}) \otimes \mathbb{I}_{N_p} \right] \tilde{\mathbf{S}}_p,$$

a positive semi-definite matrix.

With the above conclusion, we now prove that $PEP \geq c_{PEP}$. Assume the estimated $\hat{\mathbf{G}}$ as error-free, i.e., $\hat{\mathbf{G}} = \mathbf{G}$. Let $\mathbf{B}_{1,d}$ and $\mathbf{B}_{2,d}$ be two information matrices. Define $\mathbf{Z}_{1,d} \triangleq (\mathbf{G}^t \otimes \mathbb{I}_{T_d}) \tilde{\mathbf{S}}_{1,d}$, and $\mathbf{Z}_{2,d} \triangleq (\mathbf{G}^t \otimes \mathbb{I}_{T_d}) \tilde{\mathbf{S}}_{2,d}$. Let $\mathbf{b} \triangleq (\mathbf{Z}_{1,d} - \mathbf{Z}_{2,d}) \hat{\mathbf{f}} \triangleq \Delta \mathbf{Z}_d \hat{\mathbf{f}}$ and $\mathbf{a} \triangleq \beta_d \mathbf{Z}_{1,d} \Delta \mathbf{f} + \overrightarrow{\mathbf{W}}_d$. Since $\Delta \mathbf{f}$ and $\overrightarrow{\mathbf{W}}_d$ are uncorrelated, $\mathbf{a}$ is a Gaussian random variable with the distribution $\mathcal{CN}(0, \mathbf{R_a})$, where $\mathbf{R_a}$ is the associated covariance matrix. Using $\text{Cov}(\Delta \mathbf{f}) \succ \mathbf{C}_{\text{Cov}(\Delta \mathbf{f})}$, $\mathbf{R_a}$ can be lower bounded as

$$\begin{aligned}
\mathbf{R_a} =& \mathbb{E}\left(\mathbf{a}\mathbf{a}^*\right) \\
=& \beta_d^2 \mathbf{Z}_{1,d} \text{Cov}(\Delta \mathbf{f}) \mathbf{Z}_{1,d}^* + \mathbf{R}_{\overrightarrow{\mathbf{W}}_d} && (4.18) \\
\succ& \beta_d^2 \mathbf{Z}_{1,d} \mathbf{C}_{\text{Cov}(\Delta \mathbf{f})} \mathbf{Z}_{1,d}^*. && (4.19)
\end{aligned}$$

With the mismatched decoding DEC_1, we have

$$
\begin{aligned}
&PEP(\mathbf{B}_{1,d} \to \mathbf{B}_{2,d}|\mathbf{G}, \hat{\mathbf{f}}) \\
=&\mathbb{P}\left(-\beta_d^2 \mathbf{b}^* \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1} \mathbf{b} \geq \beta_d \mathbf{b}^* \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1} \mathbf{a} + \beta_d \mathbf{a}^* \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1} \mathbf{b} | \mathbf{G}, \hat{\mathbf{f}}\right) \qquad (4.20) \\
=&\mathbb{P}\Bigg(- \beta_d^2 \mathbf{b}^* \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1} \mathbf{b} \geq \mathrm{tr}\left(\beta_d (\mathbf{R}_{\mathbf{a}}^{-1/2}\mathbf{a})^* \mathbf{R}_{\mathbf{a}}^{1/2} \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1} \mathbf{b} + \beta_d \mathbf{R}_{\mathbf{a}}^{-1/2}\mathbf{a} (\mathbf{R}_{\mathbf{a}}^{1/2} \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1} \mathbf{b})^*\right) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \Big| \mathbf{G}, \hat{\mathbf{f}}\Bigg) \\
=&Q\left(\frac{\beta_d^2 \mathbf{b}^* \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1} \mathbf{b}}{\sqrt{2}\|\beta_d \mathbf{R}_{\mathbf{a}}^{1/2} \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1} \mathbf{b}\|_F} \Big| \mathbf{G}, \hat{\mathbf{f}}\right) \qquad (4.21) \\
\geq&Q\left(\frac{\hat{\mathbf{f}}^* \Delta \mathbf{Z}_d^* \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1} \Delta \mathbf{Z}_d \hat{\mathbf{f}}}{\sqrt{2\mathrm{tr}\left(\mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1} \Delta \mathbf{Z}_d \hat{\mathbf{f}} \hat{\mathbf{f}}^* \Delta \mathbf{Z}_d^* \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1} \mathbf{Z}_{1,d} \mathbf{C}_{\mathrm{Cov}(\Delta \mathbf{f})} \mathbf{Z}_{1,d}^*\right)}} \Big| \mathbf{G}, \hat{\mathbf{f}}\right), \qquad (4.22)
\end{aligned}
$$

where in (4.21) $Q(\cdot)$ is the Q-function defined as $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt$, and the equality comes from the fact that $\mathbf{R}_{\mathbf{a}}^{-1/2}\mathbf{a} \sim \mathcal{CN}(0, \mathbb{I}_{NT_d})$ and the property that if $\mathbf{N} \sim \mathcal{CN}(0, \mathbb{I})$, $\mathrm{tr}(\mathbf{C}\mathbf{N}^* + \mathbf{C}^*\mathbf{N}) \sim \mathcal{N}(0, 2\|\mathbf{C}\|_F^2)$ [1]. (4.22) follows from (4.19) and the fact that $Q(.)$ is a decreasing function. Since $\alpha_d = \frac{1}{\sqrt{R}} + O(\frac{1}{P})$, when $P \gg 1$, we can treat α_d and hence $\mathbf{R}_{\overrightarrow{\mathbf{W}}_d}$ as constants. Therefore, the right hand side of (4.22) is a positive constant independent of the power, which means that the PEP given $\mathbf{G}$ and $\hat{\mathbf{f}}$ is lower bounded by a positive constant. Note that the distributions of $\mathbf{G}$ and $\hat{\mathbf{f}}$ are irrelative of P, so after averaging over $\mathbf{G}$ and $\hat{\mathbf{f}}$, the PEP lower bound is still a positive constant, denoted as c_{PEP}, which completes the proof.

4.5.3 Proof of Theorem 4.2

Under the pre-designed power allocation, let $\alpha_p = \alpha_d \triangleq \alpha$. From the PEP in (4.20), using the Chernoff bound, we have for $\lambda > 0$

$$\mathbb{P}(\mathbf{B}_{1,d} \to \mathbf{B}_{2,d}|\mathbf{G}, \hat{\mathbf{f}})$$

$$
\begin{aligned}
&= \mathbb{P}\left(-\beta_d^2 \mathbf{b}^* \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1} \mathbf{b} - \beta_d \mathbf{b}^* \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1} \mathbf{a} - \beta_d \mathbf{a}^* \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1} \mathbf{b} > 0 | \mathbf{G}, \hat{\mathbf{f}}\right) \\
&\leq \mathbb{E}_{\mathbf{a}|\mathbf{G},\hat{\mathbf{f}}} \left(e^{-\lambda\left(\beta_d^2 \mathbf{b}^* \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1} \mathbf{b} + \beta_d \mathbf{b}^* \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1} \mathbf{a} + \beta_d \mathbf{a}^* \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1} \mathbf{b}\right)} \right) \\
&= \int e^{-\lambda\left(\beta_d^2 \mathbf{b}^* \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1} \mathbf{b} + \beta_d \mathbf{b}^* \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1} \mathbf{a} + \beta_d \mathbf{a}^* \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1} \mathbf{b}\right)} \frac{e^{-\mathbf{a}^* \mathbf{R}_{\mathbf{a}}^{-1} \mathbf{a}}}{\pi^{NT_d} \det \mathbf{R}_{\mathbf{a}}} d\mathbf{a} \\
&= \left(\int \frac{1}{\pi^{NT_d} \det \mathbf{R}_{\mathbf{a}}} e^{-\left(\mathbf{a} + \lambda \beta_d \mathbf{R}_{\mathbf{a}} \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1} \mathbf{b}\right)^* \mathbf{R}_{\mathbf{a}}^{-1} \left(\mathbf{a} + \lambda \beta_d \mathbf{R}_{\mathbf{a}} \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1} \mathbf{b}\right)} d\mathbf{a} \right) \\
&\qquad\qquad e^{-\lambda \beta_d^2 \mathbf{b}^* \left(\mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1} - \lambda \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1} \mathbf{R}_{\mathbf{a}} \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1}\right) \mathbf{b}} \\
&= e^{-\lambda \beta_d^2 \mathbf{b}^* \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1/2} \left(\mathbb{I}_{NT_d} - \lambda \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1/2} \mathbf{R}_{\mathbf{a}} \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1/2}\right) \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1/2} \mathbf{b}} \\
&< e^{-\lambda \frac{1 - \lambda(T_d/N_p + 1)(N + \alpha^2 \|\mathbf{G}\|_F^2)}{N + \alpha^2 \|\mathbf{G}\|_F^2} \beta_d^2 \mathbf{b}^* \mathbf{b}},
\end{aligned} \tag{4.23}
$$

where the last inequality is obtained using the following claim whose proof is provided later:

$$
\begin{aligned}
&\mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1/2} \left(\mathbb{I}_{NT_d} - \lambda \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1/2} \mathbf{R}_{\mathbf{a}} \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1/2}\right) \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1/2} \\
&\succ \frac{1 - \lambda(T_d/N_p + 1)(N + \alpha^2 \|\mathbf{G}\|_F^2)}{N + \alpha^2 \|\mathbf{G}\|_F^2} \mathbb{I}_{NT_d}.
\end{aligned} \tag{4.24}
$$

Let $\lambda = \frac{1}{2(T_d/N_p+1)(N+\alpha^2\|\mathbf{G}\|_F^2)}$. We have from (4.23),

$$
\mathbb{P}(\mathbf{B}_{1,d} \to \mathbf{B}_{2,d} | \mathbf{G}, \hat{\mathbf{f}}) < e^{-\frac{1}{4(T_d/N_p+1)(N+\alpha^2\|\mathbf{G}\|_F^2)^2} \beta_d^2 \mathbf{b}^* \mathbf{b}}. \tag{4.25}
$$

Next we take the expectation over $\hat{\mathbf{f}}$ to get the conditional PEP on $\mathbf{G}$. Note that $\hat{\mathbf{f}} \sim \mathcal{CN}(0, \mathbf{R}_{\hat{\mathbf{f}}})$, where $\mathbf{R}_{\hat{\mathbf{f}}} = \mathbb{I}_{MR} - \mathrm{Cov}(\Delta \mathbf{f})$. Recall that $\mathbf{b} = \Delta \mathbf{Z}_d \hat{\mathbf{f}}$ and let $q = \frac{1}{4(T_d/N_p+1)(N+\alpha^2\|\mathbf{G}\|_F^2)^2}$. From (4.25),

$$
\begin{aligned}
&\mathbb{P}(\mathbf{B}_{1,d} \to \mathbf{B}_{2,d} | \mathbf{G}) \\
&< \int \frac{1}{\pi^{MR} \det \mathbf{R}_{\hat{\mathbf{f}}}} e^{-\hat{\mathbf{f}}^* \mathbf{R}_{\hat{\mathbf{f}}}^{-1} \hat{\mathbf{f}}} e^{-q \beta_d^2 \hat{\mathbf{f}}^* \Delta \mathbf{Z}_d^* \Delta \mathbf{Z}_d \hat{\mathbf{f}}} d\hat{\mathbf{f}} \\
&= \det\left(\mathbf{R}_{\hat{\mathbf{f}}}^{-1} + q \beta_d^2 \Delta \mathbf{Z}_d^* \Delta \mathbf{Z}_d\right)^{-1} \det \mathbf{R}_{\hat{\mathbf{f}}}^{-1} \\
&= \det\left(\mathbb{I}_{NT_d} + q \beta_d^2 \Delta \mathbf{Z}_d \mathbf{R}_{\hat{\mathbf{f}}} \Delta \mathbf{Z}_d^*\right)^{-1}.
\end{aligned} \tag{4.26}
$$

Rewrite $\mathbf{Z}_d$ as $\mathbf{Z}_d = (\mathbb{I}_N \otimes \mathbf{S}_d)\left([\mathbf{G}_1^t, \cdots, \mathbf{G}_N^t]^t \otimes \mathbb{I}_M\right)$, where

$$\mathbf{G}_i \triangleq \mathrm{diag}\{g_{1i}, \cdots, g_{Ri}\}.$$

Thus $\Delta\mathbf{Z}_d = (\mathbb{I}_N \otimes \Delta\mathbf{S}_d)\left([\mathbf{G}_1^t, \cdots, \mathbf{G}_N^t]^t \otimes \mathbb{I}_M\right)$, where $\Delta\mathbf{S}_d \triangleq \mathbf{S}_{1,d} - \mathbf{S}_{2,d}$, and the diagonal matrix

$$\begin{aligned}\Delta\mathbf{Z}_d^*\Delta\mathbf{Z}_d=&\left(\overline{[\mathbf{G}_1^t, \cdots, \mathbf{G}_N^t]} \otimes \mathbb{I}_M\right)(\mathbb{I}_N \otimes \Delta\mathbf{S}_d^*)(\mathbb{I}_N \otimes \Delta\mathbf{S}_d)\left([\mathbf{G}_1^t, \cdots, \mathbf{G}_N^t]^t \otimes \mathbb{I}_M\right)\\ \succeq&\sigma_{\min}^2\left(\sum_{i=1}^{N}\mathbf{G}_i^*\mathbf{G}_i\right)\otimes\mathbb{I}_M, \end{aligned} \tag{4.27}$$

where $\sigma_{\min}^2$ is the smallest eigenvalue of $\Delta\mathbf{S}_d^*\Delta\mathbf{S}_d$. Since fully diverse orthogonal deign is used in the data transmission, $\sigma_{\min}^2 > 0$. We now provide an upper bound on $\mathrm{Cov}(\Delta\mathbf{f})$, which is proved later, for further derivation:

$$\mathrm{Cov}(\Delta\mathbf{f}) \preceq \mathrm{diag}\left\{\left(1+\frac{\beta_p^2\|\mathbf{g}_1\|_F^2}{N+\alpha^2\|\mathbf{G}\|_F^2}\right)^{-1}, \cdots, \left(1+\frac{\beta_p^2\|\mathbf{g}_R\|_F^2}{N+\alpha^2\|\mathbf{G}\|_F^2}\right)^{-1}\right\}\otimes\mathbb{I}_M. \tag{4.28}$$

From (4.28),

$$\begin{aligned}\mathbf{R}_{\hat{\mathbf{f}}}=&\mathbb{I}_{MR} - \mathrm{Cov}(\Delta\mathbf{f})\\ \succeq&\mathrm{diag}\left\{\frac{\beta_p^2\|\mathbf{g}_1\|_F^2}{N+\alpha^2\|\mathbf{G}\|_F^2+\beta_p^2\|\mathbf{g}_1\|_F^2}, \cdots, \frac{\beta_p^2\|\mathbf{g}_R\|_F^2}{N+\alpha^2\|\mathbf{G}\|_F^2+\beta_p^2\|\mathbf{g}_R\|_F^2}\right\}\otimes\mathbb{I}_M.\end{aligned} \tag{4.29}$$

From (4.26), (4.27) and (4.29), we can show that

$$\begin{aligned}&\mathbb{P}(\mathbf{B}_{1,d}\to\mathbf{B}_{2,d}|\mathbf{G})\\ <&\det\Bigg(\mathbb{I}_{MR}\\ &+q\beta_d^2\sigma_{\min}^2\mathrm{diag}\left\{\frac{\beta_p^2\|\mathbf{g}_1\|_F^2\|\mathbf{g}_1\|_F^2}{N+\alpha^2\|\mathbf{G}\|_F^2+\beta_p^2\|\mathbf{g}_1\|_F^2}, \cdots, \frac{\beta_p^2\|\mathbf{g}_R\|_F^2\|\mathbf{g}_R\|_F^2}{N+\alpha^2\|\mathbf{G}\|_F^2+\beta_p^2\|\mathbf{g}_R\|_F^2}\right\}\otimes\mathbb{I}_M\Bigg)^{-1}\end{aligned}$$

$$=\prod_{i=1}^{R}\left(1+\frac{\beta_d^2\|\mathbf{g}_i\|_F^2\sigma_{\min}^2}{4(T_d/N_p+1)(N+\alpha^2\|\mathbf{G}\|_F^2)^2}\frac{\beta_p^2\|\mathbf{g}_i\|_F^2}{N+\alpha^2\|\mathbf{G}\|_F^2+\beta_p^2\|\mathbf{g}_i\|_F^2}\right)^{-M}.$$

Since

$$\frac{\beta_p^2\|\mathbf{g}_i\|_F^2}{N+\alpha^2\|\mathbf{G}\|_F^2+\beta_p^2\|\mathbf{g}_i\|_F^2}=\left(1+\frac{N+\alpha^2\|\mathbf{G}\|_F^2}{\beta_p^2\|\mathbf{g}_i\|_F^2}\right)^{-1}>1-\frac{N+\alpha^2\|\mathbf{G}\|_F^2}{\beta_p^2\|\mathbf{g}_i\|_F^2}$$

and $\beta_d^2/\beta_p^2=T_d/N_p$, we have

$$\begin{aligned}
&\mathbb{P}(\mathbf{B}_{1,d}\to\mathbf{B}_{2,d}|\mathbf{G})\\
<&\prod_{i=1}^{R}\left(1+\frac{\beta_d^2\|\mathbf{g}_i\|_F^2\sigma_{\min}^2}{4(T_d/N_p+1)(N+\alpha^2\|\mathbf{G}\|_F^2)^2}-\frac{\sigma_{\min}^2T_d}{4(T_d+N_p)(N+\alpha^2\|\mathbf{G}\|_F^2)}\right)^{-M}\\
<&\prod_{i=1}^{R}\left[\left(1-\frac{\sigma_{\min}^2T_d}{4N(T_d+N_p)}\right)+\frac{\beta_d^2\|\mathbf{g}_i\|_F^2\sigma_{\min}^2}{4(T_d/N_p+1)(N+\alpha^2\|\mathbf{G}\|_F^2)^2}\right]^{-M}\\
=&\gamma^{-MR}\prod_{i=1}^{R}\left(1+\frac{\beta_d^2\|\mathbf{g}_i\|_F^2\sigma_{\min}^2}{4\gamma(T_d/N_p+1)(N+\alpha^2\|\mathbf{G}\|_F^2)^2}\right)^{-M},
\end{aligned}\tag{4.30}$$

where $\gamma\triangleq 1-\frac{\sigma_{\min}^2T_d}{4N(T_d+N_p)}$. Notice that $\gamma>0$ since $\sigma_{\min}^2\le 4$ for any code design. (4.30) is the same as the PEP upper bound formula (30) in [25] except for the square in the denominator. Notice that when $P\gg 1$, $\alpha\sim 1$ and $\beta_d^2\sim P$. Following the techniques in [25], we can show that the square does not affect the diversity result and full diversity can be obtained.

Now we prove (4.28). With perfect knowledge of $\mathbf{G}$, we have $\mathbf{R}_{\overrightarrow{\mathbf{W}}_p}=(\mathbb{I}_N+\alpha_p^2\overline{\mathbf{G}^*\mathbf{G}})\otimes\mathbb{I}_{Np}$ and $\mathbf{R}_{\overrightarrow{\mathbf{W}}_d}=(\mathbb{I}_N+\alpha_p^2\overline{\mathbf{G}^*\mathbf{G}})\otimes\mathbb{I}_{T_d}$. They can be bounded as

$$\begin{aligned}
\mathbb{I}_{NN_p}&\preceq\mathbf{R}_{\overrightarrow{\mathbf{W}}_p}\preceq(N+\alpha^2\|\mathbf{G}\|_F^2)\mathbb{I}_{NN_p},\\
\mathbb{I}_{NT_d}&\preceq\mathbf{R}_{\overrightarrow{\mathbf{W}}_d}\preceq(N+\alpha^2\|\mathbf{G}\|_F^2)\mathbb{I}_{NT_d}.
\end{aligned}\tag{4.31}$$

Using the upper bound of $\mathbf{R}_{\overrightarrow{\mathbf{W}}_p}$,

$$\begin{aligned}
\mathrm{Cov}(\Delta\mathbf{f})\preceq&\left(\mathbb{I}_{MR}+\beta_p^2(N+\alpha^2\|\mathbf{G}\|_F^2)^{-1}\mathbf{Z}_p^*\mathbf{Z}_p\right)^{-1}\\
=&\Big(\mathbb{I}_{MR}+\beta_p^2(N+\alpha^2\|\mathbf{G}\|_F^2)^{-1}(\overline{[\mathbf{G}_1^t,\cdots,\mathbf{G}_N^t]}\otimes\mathbb{I}_M)
\end{aligned}$$

$$(\mathbb{I}_N \otimes \mathbf{S}_p^* \mathbf{S}_p) \left([\mathbf{G}_1^t, \cdots, \mathbf{G}_N^t]^t \otimes \mathbb{I}_M \right) \Big)^{-1},$$

which can be simplified as the right hand side of (4.28). The second equality holds since $(\mathbb{I}_N \otimes \mathbf{S}_p^* \mathbf{S}_p) = \mathbb{I}_{NMR}$, which is derived from $\mathbf{S}_p^* \mathbf{S}_p = \mathbb{I}_{MR}$ when $N_p \geq MR$.

Finally we prove (4.24). Using (4.28),

$$\begin{aligned}
&\beta_d^2 \mathbf{Z}_{1,d} \mathrm{Cov}(\Delta \mathbf{f}) \mathbf{Z}_{1,d}^* \\
\preceq &(\mathbb{I}_N \otimes \mathbf{S}_{1,d}) \\
&\overbrace{\left([\mathbf{G}_1^t \ \cdots \mathbf{G}_N^t]^t \underbrace{\mathrm{diag}\left\{ \frac{\beta_d^2}{1 + \frac{\beta_p^2 \|\mathbf{g}_1\|_F^2}{N + \alpha^2 \|\mathbf{G}\|_F^2}}, \cdots, \frac{\beta_d^2}{1 + \frac{\beta_p^2 \|\mathbf{g}_R\|_F^2}{N + \alpha^2 \|\mathbf{G}\|_F^2}} \right\}}_{\mathbf{\Lambda}} ([\mathbf{G}_1^t \ \cdots \mathbf{G}_N^t]^t)^* \right)}^{\mathbf{X}} \otimes \mathbb{I}_M \\
&(\mathbb{I}_N \otimes \mathbf{S}_{1,d})^* \\
\preceq &(\mathbb{I}_N \otimes \mathbf{S}_{1,d}) \underbrace{\mathrm{diag}\left\{ \sum_{i=1}^N \mathbf{G}_i \mathbf{G}_i^* \mathbf{\Lambda}, \cdots, \sum_{i=1}^N \mathbf{G}_i \mathbf{G}_i^* \mathbf{\Lambda} \right\}}_{\mathbf{Y}} \otimes \mathbb{I}_M (\mathbb{I}_N \otimes \mathbf{S}_{1,d})^* \qquad (4.32) \\
= &(\mathbb{I}_N \otimes \mathbf{S}_{1,d}) \left(\mathbb{I}_N \otimes \mathrm{diag}\left\{ \frac{\beta_d^2 \|\mathbf{g}_1\|_F^2}{1 + \frac{\beta_p^2 \|\mathbf{g}_1\|_F^2}{N + \alpha^2 \|\mathbf{G}\|_F^2}}, \cdots, \frac{\beta_d^2 \|\mathbf{g}_R\|_F^2}{1 + \frac{\beta_p^2 \|\mathbf{g}_R\|_F^2}{N + \alpha^2 \|\mathbf{G}\|_F^2}} \right\} \otimes \mathbb{I}_M \right) (\mathbb{I}_N \otimes \mathbf{S}_{1,d})^* \\
\prec &\left((N + \alpha^2 \|\mathbf{G}\|_F^2) T_d / N_p \right) (\mathbb{I}_N \otimes \mathbf{S}_{1,d}) (\mathbb{I}_N \otimes \mathbf{S}_{1,d})^* \qquad (4.33) \\
= &\left((N + \alpha^2 \|\mathbf{G}\|_F^2) T_d / N_p \right) (\mathbb{I}_N \otimes \mathbf{S}_{1,d} \mathbf{S}_{1,d}^*) \\
\prec &\left((N + \alpha^2 \|\mathbf{G}\|_F^2) T_d / N_p \right) \mathbb{I}_{NT_d}. \qquad (4.34)
\end{aligned}$$

(4.32) holds since it can be shown that $\mathbf{X} \preceq \mathbf{Y}$. To derive the inequality in (4.33), the fact $\beta_d^2 / \beta_p^2 = T_d / N_p$ is used. (4.34) holds since $\mathbf{S}_{1,d} \mathbf{S}_{1,d}^* \preceq \mathbb{I}_{T_d}$ can be obtained from $\mathbf{S}_{1,d}^* \mathbf{S}_{1,d} = \mathbb{I}_{MR}$ when $T_d \geq MR$. Recall the expression of $\mathbf{R_a}$ in (4.18). Using the upper bound of $\beta_d^2 \mathbf{Z}_{1,d} \mathrm{Cov}(\Delta \mathbf{f}) \mathbf{Z}_{1,d}^*$ in (4.34) together with the bounds of $\mathbf{R}_{\overrightarrow{\mathbf{W}}_d}$ in (4.31), $\mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1/2} \mathbf{R_a} \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1/2} \prec (T_d/N_p + 1)(N + \alpha^2 \|\mathbf{G}\|_F^2) \mathbb{I}_{NT_d}$, hence $\mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1/2} \left(\mathbb{I}_{NT_d} - \lambda \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1/2} \mathbf{R_a} \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1/2} \right) \mathbf{R}_{\overrightarrow{\mathbf{W}}_d}^{-1/2} \succ \frac{1 - \lambda (T_d/N_p + 1)(N + \alpha^2 \|\mathbf{G}\|_F^2)}{N + \alpha^2 \|\mathbf{G}\|_F^2} \mathbb{I}_{NT_d}$.

4.5.4 Proof of Theorem 4.3

Take the SVD of $\hat{\mathbf{G}}^t$ as $\hat{\mathbf{G}}^t = \mathbf{U}\hat{\mathbf{\Sigma}}\mathbf{V}$ where $\mathbf{U}$ and $\mathbf{V}$ are $R\times R$ unitary matrices, and $\hat{\mathbf{\Sigma}} = \text{diag}\{\hat{\lambda}_1^{1/2}, \cdots, \hat{\lambda}_R^{1/2}\}$ with $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_R$. We first prove two lemmas to help the diversity analysis.

Lemma 4.4. *When $P \gg 1$, given that $\hat{\lambda}_R \leq 1/P$, we have $\mathbf{R}_{\Delta\mathbf{f}} \succ \mathbf{C}_{\mathbf{R}_{\Delta\mathbf{f}}}$, where*

$$\mathbf{C}_{\mathbf{R}_{\Delta\mathbf{f}}} = \mathbf{V}^* \text{diag}\left\{0, \cdots, 0, \frac{2R}{2R+1}\right\}\mathbf{V}.$$

Proof. Replace $\hat{\mathbf{G}}^t$ with its SVD in (3.38),

$$\mathbf{R}_{\Delta\mathbf{f}} = \mathbf{V}^* \text{diag}\left\{\frac{k+\alpha_p^2\hat{\lambda}_1}{k+(\alpha_p^2+\beta_p^2)\hat{\lambda}_1}, \cdots, \frac{k+\alpha_p^2\hat{\lambda}_R}{k+(\alpha_p^2+\beta_p^2)\hat{\lambda}_R}\right\}\mathbf{V}. \tag{4.35}$$

When $\hat{\lambda}_R \leq 1/P$, we have $\frac{k+\alpha_p^2\hat{\lambda}_R}{k+(\alpha_p^2+\beta_p^2)\hat{\lambda}_R} \geq \frac{k+\alpha_p^2/P}{k+(\alpha_p^2+\beta_p^2)/P} = \frac{2R}{2R+1}+O\left(\frac{1}{P}\right)$, where the first inequality follows from the fact that $\frac{k+\alpha_p^2 x}{k+(\alpha_p^2+\beta_p^2)x}$ is a decreasing function. Thus, we have $\mathbf{R}_{\Delta\mathbf{f}} \succ \mathbf{C}_{\mathbf{R}_{\Delta\mathbf{f}}}$. □

Lemma 4.5. *The PEP of the network given $\hat{\lambda}_R \leq 1/P$ is no smaller than a positive constant c_{PEP}.*

Proof. Now we calculate the PEP of mistaking the codeword $\mathbf{S}_{1,d}$ with $\mathbf{S}_{2,d}$. Under the pre-designed power allocation, let $\alpha_p = \alpha_d \triangleq \alpha$. The two codewords correspond to the information vectors $\mathbf{s}_{1,d}$ and $\mathbf{s}_{2,d}$ respectively. Given that $\mathbf{S}_{1,d}$ is sent, from the approximate training-based system equation (4.12), we have

$$\overrightarrow{\mathbf{X}}_{1,d} = \beta_d\hat{\mathbf{Z}}_{1,d}\hat{\mathbf{f}} + \overrightarrow{\mathbf{W}}'_{e1,d},$$

where $\hat{\mathbf{Z}}_{1,d} \triangleq (\hat{\mathbf{G}}^t \otimes \mathbb{I}_{T_d})\tilde{\mathbf{S}}_{1,d}$, and the equivalent noise term $\overrightarrow{\mathbf{W}}'_{e1,d} \triangleq \beta_d\hat{\mathbf{Z}}_{1,d}\Delta\mathbf{f} + \overrightarrow{\mathbf{W}}_{r,d}$ which follows $\mathcal{CN}(\mathbf{0}, \mathbf{R}_{\overrightarrow{\mathbf{W}}'_{e1,d}})$. Using the result in Lemma 4.4 that $\mathbf{R}_{\Delta\mathbf{f}} \succ \mathbf{C}_{\mathbf{R}_{\Delta\mathbf{f}}}$, the covariance matrix of $\overrightarrow{\mathbf{W}}'_{e1,d}$ can be bounded as

$$\mathbf{R}_{\overrightarrow{\mathbf{W}}'_{e1,d}} = \beta_d^2\hat{\mathbf{Z}}_{1,d}\mathbf{R}_{\Delta\mathbf{f}}\hat{\mathbf{Z}}_{1,d}^* + \mathbb{I}_{T_dR} \succ \beta_d^2\hat{\mathbf{Z}}_{1,d}\mathbf{C}_{\mathbf{R}_{\Delta\mathbf{f}}}\hat{\mathbf{Z}}_{1,d}^*. \tag{4.36}$$

Let $\mathbf{b} \triangleq (\hat{\mathbf{Z}}_{1,d} - \hat{\mathbf{Z}}_{2,d})\hat{\mathbf{f}} \triangleq \Delta\hat{\mathbf{Z}}_d\hat{\mathbf{f}}$, where $\hat{\mathbf{Z}}_{2,d} \triangleq (\hat{\mathbf{G}}^t \otimes \mathbb{I}_{T_d})\tilde{\mathbf{S}}_{2,d}$. Using the mismatched decoding DEC_1, the conditional PEP can be calculated as follows.

$$
\begin{aligned}
&\mathrm{PEP}(\mathbf{S}_{1,d} \to \mathbf{S}_{2,d}|\hat{\mathbf{G}}, \hat{\mathbf{f}}) \\
=&\mathbb{P}\Bigg(\left(\overrightarrow{\mathbf{X}}_{1,d} - \beta_d\hat{\mathbf{Z}}_{1,d}\hat{\mathbf{f}}\right)^* \hat{\mathbf{R}}^{-1}_{\overrightarrow{\mathbf{W}}_d} \left(\overrightarrow{\mathbf{X}}_{1,d} - \beta_d\hat{\mathbf{Z}}_{1,d}\hat{\mathbf{f}}\right) \\
&\qquad\qquad \geq \left(\overrightarrow{\mathbf{X}}_{1,d} - \beta_d\hat{\mathbf{Z}}_{2,d}\hat{\mathbf{f}}\right)^* \hat{\mathbf{R}}^{-1}_{\overrightarrow{\mathbf{W}}_d} \left(\overrightarrow{\mathbf{X}}_{1,d} - \beta_d\hat{\mathbf{Z}}_{2,d}\hat{\mathbf{f}}\right) |\hat{\mathbf{G}}, \hat{\mathbf{f}} \Bigg) \\
=&\mathbb{P}\Bigg(-\beta_d^2\mathbf{b}^*\hat{\mathbf{R}}^{-1}_{\overrightarrow{\mathbf{W}}_d}\mathbf{b} \geq \mathrm{tr}\Big(\beta_d(\mathbf{R}^{-1/2}_{\overrightarrow{\mathbf{W}}'_{e1,d}} \overrightarrow{\mathbf{W}}'_{e1,d})^*\mathbf{R}^{1/2}_{\overrightarrow{\mathbf{W}}'_{e1,d}} \hat{\mathbf{R}}^{-1}_{\overrightarrow{\mathbf{W}}_d}\mathbf{b} + \\
&\qquad\qquad \beta_d\mathbf{R}^{-1/2}_{\overrightarrow{\mathbf{W}}'_{e1,d}} \overrightarrow{\mathbf{W}}'_{e1,d}(\mathbf{R}^{1/2}_{\overrightarrow{\mathbf{W}}'_{e1,d}} \hat{\mathbf{R}}^{-1}_{\overrightarrow{\mathbf{W}}_d}\mathbf{b})^*\Big)\Big|\hat{\mathbf{G}}, \hat{\mathbf{f}} \Bigg) \\
=&Q\left(\frac{\beta_d^2\mathbf{b}^*\hat{\mathbf{R}}^{-1}_{\overrightarrow{\mathbf{W}}_d}\mathbf{b}}{\sqrt{2}\|\beta_d\mathbf{R}^{1/2}_{\overrightarrow{\mathbf{W}}'_{e1,d}} \hat{\mathbf{R}}^{-1}_{\overrightarrow{\mathbf{W}}_d}\mathbf{b}\|_F} \Big| \hat{\mathbf{G}}, \hat{\mathbf{f}} \right) \qquad (4.37) \\
\geq&Q\left(\frac{\hat{\mathbf{f}}^*\Delta\hat{\mathbf{Z}}_d^*\hat{\mathbf{R}}^{-1}_{\overrightarrow{\mathbf{W}}_d}\Delta\hat{\mathbf{Z}}_d\hat{\mathbf{f}}}{\sqrt{2\mathrm{tr}\left(\hat{\mathbf{Z}}^*_{1,d}\hat{\mathbf{R}}^{-1}_{\overrightarrow{\mathbf{W}}_d}\Delta\hat{\mathbf{Z}}_d\hat{\mathbf{f}}\hat{\mathbf{f}}^*\Delta\hat{\mathbf{Z}}_d^*\hat{\mathbf{R}}^{-1}_{\overrightarrow{\mathbf{W}}_d}\hat{\mathbf{Z}}_{1,d}\mathbf{C}_{\mathbf{R}_{\Delta\mathbf{f}}}\right)}} \Big| \hat{\mathbf{G}}, \hat{\mathbf{f}} \right), \qquad (4.38)
\end{aligned}
$$

where in (4.37) $Q(\cdot)$ is the Q-function defined as $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-t^2/2}dt$, and the equality comes from the fact that $\mathbf{R}^{-1/2}_{\overrightarrow{\mathbf{W}}'_{e1,d}} \overrightarrow{\mathbf{W}}'_{e1,d} \sim \mathcal{CN}(\mathbf{0}, \mathbb{I}_{T_dR})$ and the property that if $\mathbf{N} \sim \mathcal{CN}(\mathbf{0}, \mathbb{I})$, $\mathrm{tr}(\mathbf{CN}^* + \mathbf{C}^*\mathbf{N}) \sim \mathcal{N}(0, 2\|\mathbf{C}\|_F^2)$ [1]. (4.38) follows from (4.36) and the fact that Q-function is a decreasing function. For the tractability of analysis, we make another approximation $\hat{\mathbf{G}}\hat{\mathbf{G}}^* \approx \alpha^2R^2\mathbb{I}_R$ for large P. That is, we replace $\hat{\mathbf{G}}\hat{\mathbf{G}}^*$ with its average. This approximation is expected to be tight especially for large R, since $\hat{\mathbf{G}}\hat{\mathbf{G}}^* \to \alpha^2R^2\mathbb{I}_R$ as $R \to \infty$. With this approximation, we have $\hat{\mathbf{R}}_{\overrightarrow{\mathbf{W}}_d} \approx (1 + \alpha^4R^2)\mathbb{I}_{T_dR}$, $\Delta\hat{\mathbf{Z}}_d^*\Delta\hat{\mathbf{Z}}_d \approx \alpha^2R^2\|\mathbf{s}_{1,d} - \mathbf{s}_{2,d}\|_F^2$, and

$$
\Delta\hat{\mathbf{Z}}_d^*\hat{\mathbf{Z}}_{1,d} \approx \alpha^2R^2\left(\tilde{\mathbf{S}}_{1,d} - \tilde{\mathbf{S}}_{2,d}\right)^*\tilde{\mathbf{S}}_{1,d} = \alpha^2R^2(\mathbf{s}_{1,d} - \mathbf{s}_{2,d})^*\mathbf{s}_{1,d}\mathbb{I}_R.
$$

Thus, from (4.38), given $\hat{\lambda}_R \leq 1/P$, we have

$$\mathrm{PEP}(\mathbf{S}_{1,d} \to \mathbf{S}_{2,d}|\hat{\mathbf{G}}, \hat{\mathbf{f}}) \geq Q\left(\frac{\|\mathbf{s}_{1,d} - \mathbf{s}_{2,d}\|_F^2 \hat{\mathbf{f}}^* \hat{\mathbf{f}}}{\sqrt{2}|(\mathbf{s}_{1,d} - \mathbf{s}_{2,d})^* \mathbf{s}_{1,d}| \sqrt{\mathrm{tr}\left(\hat{\mathbf{f}}\hat{\mathbf{f}}^* \mathbf{C}_{\mathbf{R}_{\Delta \mathbf{f}}}\right)}}\right). \quad (4.39)$$

The argument of the Q-function in (4.39) is irrelative of P and the denominator of the argument is non-zero. The value of Q-function in (4.39) is a positive number independent of P. Note that it is also independent of $\hat{\mathbf{G}}$ other than the condition $\hat{\lambda}_R \leq 1/P$. Thus, after performing the expectation over $\hat{\mathbf{f}}$, the PEP given $\hat{\lambda}_R \leq 1/P$ is no less than a positive number, which we call c_{PEP}. □

With the result in Lemma 4.5, we can lower bound the overall PEP. When $P \gg 1$, $\mathbb{P}(\hat{\lambda}_R \leq 1/P) = 1 - e^{-R(1+P)/P^2} \approx RP^{-1} + O(P^{-2})$. Together with Lemma 4.5, we have

$$\begin{aligned}\mathrm{PEP} &= (\mathrm{PEP}|\hat{\lambda}_R \leq 1/P)\mathbb{P}(\hat{\lambda}_R \leq 1/P) + (\mathrm{PEP}|\hat{\lambda}_R > 1/P)\mathbb{P}(\hat{\lambda}_R > 1/P) \\ &> (\mathrm{PEP}|\hat{\lambda}_R \leq 1/P)\mathbb{P}(\hat{\lambda}_R \leq 1/P) > c_{PEP} RP^{-1} + O(P^{-2}),\end{aligned}$$

which means that the diversity order is upper bounded by 1.

5 Training-Based Matched Decoding and Diversity Analysis

In the previous chapter, for the training-based DSTC network, we discussed mismatched decoding where the estimation error is ignored for simplicity of implementation. In this chapter, we study a decoding strategy where the estimation error is taken into account and treated as one part of the noises. Since it matches the training-based transmission equation, we call it matched decoding. For the tractability of the analysis, all investigations in this chapter are based on the network with $M = 1$, and $R = N \geq 2$. The channel estimations $\hat{\mathbf{G}}$ and $\hat{\mathbf{f}}$ given in (3.2) and (3.16) are employed, respectively.

In Section 5.1, the matched decoding DEC_2 is derived and its diversity performance is analyzed. We will see that, in contrast to mismatched decoding, matched decoding DEC_2 can achieve full diversity with $N_p = N_{p,l}$. However, its implementation is computationally prohibitive, which could make it impractical in reality. Thus in Section 5.2, a modified matched decoding, adaptive decoding, is introduced to balance the performance and complexity. For comparison, the complexity analysis of mismatched decoding, matched decoding, and adaptive decoding is provided in Section 5.3. Simulation results are exhibited in Section 5.4.

5.1 Matched Decoding

Having shown in Section 4.3 that mismatched decoding DEC_1 loses diversity when $N_p = N_{p,l} = 1$ due to the neglect of the estimation error in decoding, our particular interest is: Can full diversity be achieved when $N_p = 1$, i.e., the shortest training time is used? For this consideration, we further investigate the approximate training-based system equation in (4.12).

To derive the ML decoding metric matching with (4.12), we need the conditional PDF of $\overrightarrow{\mathbf{X}}_d$ given $\hat{\mathbf{G}}$ and $\hat{\mathbf{f}}$, denoted as $\mathbb{P}(\overrightarrow{\mathbf{X}}_d|\hat{\mathbf{G}}, \hat{\mathbf{f}})$. However, since $\overrightarrow{\mathbf{W}}'_{e,d}$

is not Gaussian for $\Delta\mathbf{f}$ is not Gaussian, it's complicated to derive such conditional PDF. For the tractability of the analysis, we treat $\Delta\mathbf{f}$ as if Gaussian, as we did in Section 4.3. Thus, there is

$$\mathbb{P}(\overrightarrow{\mathbf{X}}_d|\hat{\mathbf{G}},\hat{\mathbf{f}}) = \frac{1}{\pi^{T_d R}\det \mathbf{R}_{\overrightarrow{\mathbf{W}}'_{e,d}}} e^{-\left(\overrightarrow{\mathbf{X}}_d-\beta_d\hat{\mathbf{Z}}_d\hat{\mathbf{f}}\right)^* \mathbf{R}^{-1}_{\overrightarrow{\mathbf{W}}'_{e,d}} \left(\overrightarrow{\mathbf{X}}_d-\beta_d\hat{\mathbf{Z}}_d\hat{\mathbf{f}}\right)}. \tag{5.1}$$

By maximizing this PDF, the following decoding rule is obtained:

$$\text{DEC}_2 : \arg\min_{\mathbf{s}_d} \left(\ln\det(\mathbf{R}_{\overrightarrow{\mathbf{W}}'_{e,d}}) + \left(\overrightarrow{\mathbf{X}}_d - \beta_d\hat{\mathbf{Z}}_d\hat{\mathbf{f}}\right)^* \mathbf{R}^{-1}_{\overrightarrow{\mathbf{W}}'_{e,d}} \left(\overrightarrow{\mathbf{X}}_d - \beta_d\hat{\mathbf{Z}}_d\hat{\mathbf{f}}\right) \right). \tag{5.2}$$

Compared with DEC_1 in (4.3), in DEC_2, the covariance matrix of the estimation error term $\beta_d\hat{\mathbf{Z}}_d\Delta\mathbf{f}$ is incorporated through $\mathbf{R}_{\overrightarrow{\mathbf{W}}'_{e,d}}$. We call it matched decoding, because as opposed to DEC_1, it takes into account the estimation error and matches the approximate training-based transmission equation. Note that, since the true $\Delta\mathbf{f}$ is non-Gaussian, this decoding is not the optimal ML decoding of the approximate training-based transmission equation, but a suboptimal decoding. Simulation shows that DEC_2 achieves full diversity even when $N_p = 1$. Thus, with the proposed matched decoding DEC_2, the shortest training time $R+2$ symbol intervals are enough for the training phase to achieve full diversity in data transmission, while with mismatched decoding DEC_1, the minimum training length for full diversity is $3R$ symbol intervals.

However, this reduction in training time or improvement in diversity comes with a price on computational complexity. It is noteworthy that $\mathbf{R}_{\overrightarrow{\mathbf{W}}'_{e,d}}$ in (4.13) depends on the transmitted signal through $\hat{\mathbf{Z}}_d$. DEC_2 thus cannot be reduced to decoupled symbol-by-symbol decoding. The update of $\mathbf{R}^{-1}_{\overrightarrow{\mathbf{W}}'_{e,d}}$ during decoding further aggravates the computational load.

Algorithm 4 Adaptive decoding (A-DEC)

1: Choose the positive threshold ϵ.
2: Let $\hat{\lambda}_R$ be the smallest eigenvalue of $\hat{\mathbf{G}}\hat{\mathbf{G}}^*$. If $\hat{\lambda}_R < \epsilon$, employ DEC_2; otherwise, employ $\mathrm{DEC}_{1,simp}$.

5.2 Adaptive Decoding

In a training-based communication system, it is desirable to achieve reliable communication with both a short training period and low decoding complexity. From Section 4.1, we know that $\mathrm{DEC}_{1,simp}$ can be performed symbol-wise; in addition, from simulation, $\mathrm{DEC}_{1,simp}$ is observed to have almost the same performance as DEC_1. Therefore, although $\mathrm{DEC}_{1,simp}$ possesses a low complexity, it cannot achieve full diversity with the minimum training length, $R+2$ symbol intervals. We also know that, DEC_2 can achieve full diversity with the minimum training length, but it requires joint decoding of all information symbols. In this subsection, we propose an adaptive decoding (A-DEC) by switching between $\mathrm{DEC}_{1,simp}$ and DEC_2 based on the quality of $\hat{\mathbf{G}}$ to achieve a balance between performance and complexity with the shortest training length.

The idea of A-DEC is to use the more complicated but reliable matched decoding DEC_2 only when necessary. Since we target at the shortest training, only the $N_p = 1$ case is considered in this part. As discussed before, when $\hat{\mathbf{G}}$ is close to singular, the equations in the training model (3.19) are close to dependent, thus large estimation error occurs. Let $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_R$ be the ordered eigenvalues of $\hat{\mathbf{G}}\hat{\mathbf{G}}^*$. We thus use $\hat{\lambda}_R$, the smallest eigenvalue of $\hat{\mathbf{G}}\hat{\mathbf{G}}^*$, to indicate the quality of $\hat{\mathbf{G}}$. When $\hat{\lambda}_R$ is less than a pre-designed threshold ϵ, we consider $\hat{\mathbf{G}}$ as ill-conditioned (close to singular) and adopt DEC_2; otherwise, the low-complexity decoding $\mathrm{DEC}_{1,simp}$ is used. A-DEC is also described in Algorithm 4.

With A-DEC, the balance between performance and complexity can be controlled by adjusting the value of ϵ. When $\epsilon = 0$, A-DEC reduces to $\mathrm{DEC}_{1,simp}$; when ϵ approaches infinity, it becomes DEC_2. With a smaller ϵ, A-DEC has lower complexity but achieves less reliability, and vice versa. Simulation shows

that for any positive ϵ, full diversity can be obtained when $N_p = 1$.

5.3 Complexity Analysis

We now study the complexity of the proposed decoding schemes, including the mismatched decodings (DEC_1 and $\text{DEC}_{1,simp}$), the matched decoding (DEC_2), and the adaptive decoding (A-DEC).

To measure complexity, the unit of flop is employed, which is defined as the amount of the calculation associated with one elementary operation (addition or multiplication) [66,67]. The followings are the rules we adopt in calculating the number of flops. To calculate $\mathbf{AB} + \mathbf{C}$ where $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are $m \times n$, $n \times p$, and $m \times p$ matrices respectively, $2mnp$ flops are needed. To save flops, the calculation of the inverse of an $n \times n$ non-singular matrix is performed by solving linear equations, where $2n^3/3$ flops are required. To calculate the determinant of an $n \times n$ matrix, $2n^3/3$ flops are required.

Let $\mathcal{C}$ be the modulation for the information symbols $s_{d,i}$'s and $|\mathcal{C}|$ the cardinality of $\mathcal{C}$. Assume that a rate-1 (with respect to each step) orthogonal DSTC is used in data transmission, e.g., Alamouti code. Orthogonal DSTC with other rates can be treated similarly. By the aforementioned rules, the number of flops needed for performing DEC_1, $\text{DEC}_{1,simp}$, and DEC_2 can be calculated to be:

$$\begin{aligned}
&\text{Comp}(\text{DEC}_1) \\
&= R^3 \left[\left(\frac{2}{3} T_d + 2 \right) T_d^2 |\mathcal{C}|^T + 2 \right] + 2R^2 T_d |\mathcal{C}|^T + 2RT_d |\mathcal{C}|^T, \qquad (5.3) \\
&\text{Comp}(\text{DEC}_{1,simp}) \\
&= R^2(2T_d^2 + 2T_d + 1) + RT_d + T_d(|\mathcal{C}| - 1), \qquad (5.4) \\
&\text{Comp}(\text{DEC}_2) \\
&= R^3 \left[\left(\frac{4}{3} T_d^2 + 4T_d + 2 \right) T_d |\mathcal{C}|^T + 10 \right] + 2R^2 T_d |\mathcal{C}|^T + 2RT_d |\mathcal{C}|^T + |\mathcal{C}|^T, \qquad (5.5)
\end{aligned}$$

respectively, where we have used $\text{Comp}(\cdot)$ to indicate the complexity of a de-

coding. It is plain to see that, for DEC_1 and DEC_2, the associated complexities are cubical in the network size R and exponential in T_d due to the joint symbol decoding; for $\mathrm{DEC}_{1,simp}$, the associated complexity is quadratical in R and T_d due to symbol-wise decoding.

For A-DEC, the average number of flops can be derived approximately as follows:

$$\begin{aligned} &\mathbb{E}\left[\mathrm{Comp}(\text{A-DEC})\right] \\ &= \mathbb{P}(\hat{\lambda}_R < \epsilon)\mathrm{Comp}(\mathrm{DEC}_2) + \mathbb{P}(\hat{\lambda}_R \geq \epsilon)\mathrm{Comp}(\mathrm{DEC}_{1,simp}). \end{aligned} \tag{5.6}$$

Since $\hat{\mathbf{G}}\hat{\mathbf{G}}^*$ is a Wishart matrix [54], it can be shown that $\hat{\lambda}_R$ is exponentially distributed with mean $P/\left(R(P+1)\right)$. Thus $\mathbb{P}(\hat{\lambda}_R < \epsilon) = 1 - e^{-R\epsilon(1+P)/P} \approx R\epsilon$ for small ϵ, which means that DEC_2 will be adopted with an approximate probability of $R\epsilon$.

In Table 5.1 and 5.2, we list the required numbers of flops for DEC_1, $\mathrm{DEC}_{1,simp}$, DEC_2, and A-DEC with several common constellations when $R = 2$ and 3 respectively. T_d, the length of each step in data transmission, is set to be 2 when $R = 2$ and 4 when $R = 3$.

Table 5.1: Numbers of flops in coherent decodings with $R = 2$

	DEC_1	$\text{DEC}_{1,simp}$	DEC_2	A-DEC $\epsilon = 0.1$	A-DEC $\epsilon = 0.01$	A-DEC $\epsilon = 0.001$
BPSK	539	58	1162	279	81	61
4QAM	2107	62	4406	931	149	71
16QAM	33467	86	69285	13926	1470	225

Table 5.2: Numbers of flops in coherent decodings with $R = 3$

	DEC_1	$\text{DEC}_{1,simp}$	DEC_2	A-DEC $\epsilon = 0.1$	A-DEC $\epsilon = 0.01$	A-DEC $\epsilon = 0.001$
BPSK	3.38×10^4	385	6.98×10^4	2.12×10^4	2.47×10^3	594
4QAM	5.41×10^5	393	1.11×10^6	3.34×10^5	3.38×10^4	3.73×10^3
16QAM	1.38×10^8	441	2.84×10^8	8.54×10^7	8.54×10^6	8.55×10^5

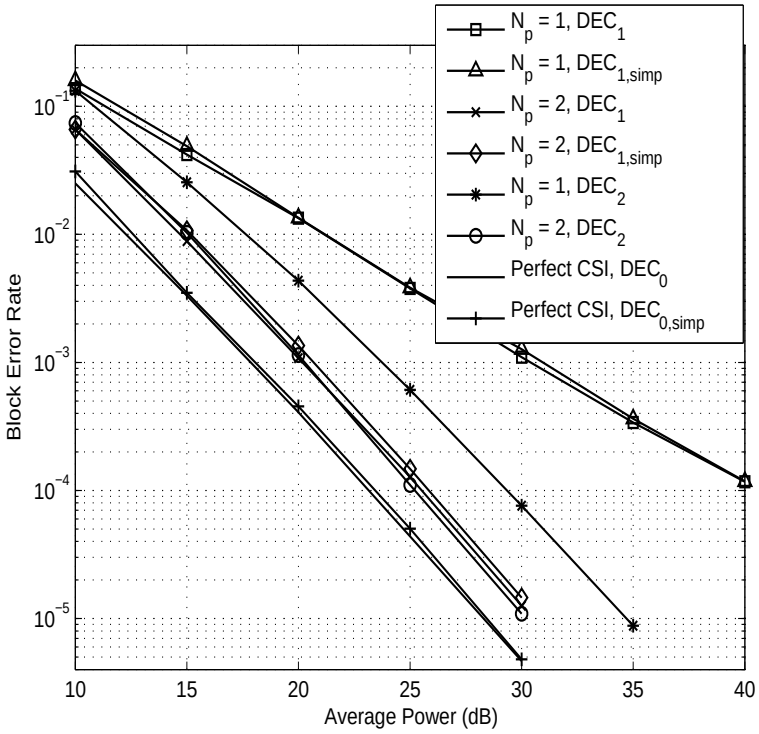

Figure 5.1: Mismatched decoding, matched decoding, and perfect CSI decoding for the network with $M = 1, R = N = 2$.

5.4 Simulation Results

In this section, simulation results are exhibited. BPSK is used for data transmission for all simulated networks. For two-relay networks, i.e., $R = 2$, Alamouti code is used, while for three-relay networks, i.e., $R = 3$, a generalized rate-1 real orthogonal code is used [1].

In Fig. 5.1, the performance of mismatched decodings ($\text{DEC}_1$and $\text{DEC}_{1,simp}$), matched decoding (DEC_2), and perfect CSI decodings (DEC_0 and $\text{DEC}_{0,simp}$) are shown for the network with $R = 2$. First, we observe that the differences between DEC_1 and $\text{DEC}_{1,simp}$, and between DEC_0 and $\text{DEC}_{0,simp}$ are negligible. But note that, from Table 5.1, the complexity of the simplified decodings in terms of the number of flops is only 1/10 that of the original ones. Second, for DEC_2, full diversity is achieved even for $N_p = 1$, and the performance is about 11dB better than that of DEC_1 at the BLER of 10^{-4}. When $N_p = 2$, DEC_2 is only slightly better than DEC_1. But the complexity of DEC_2 is about

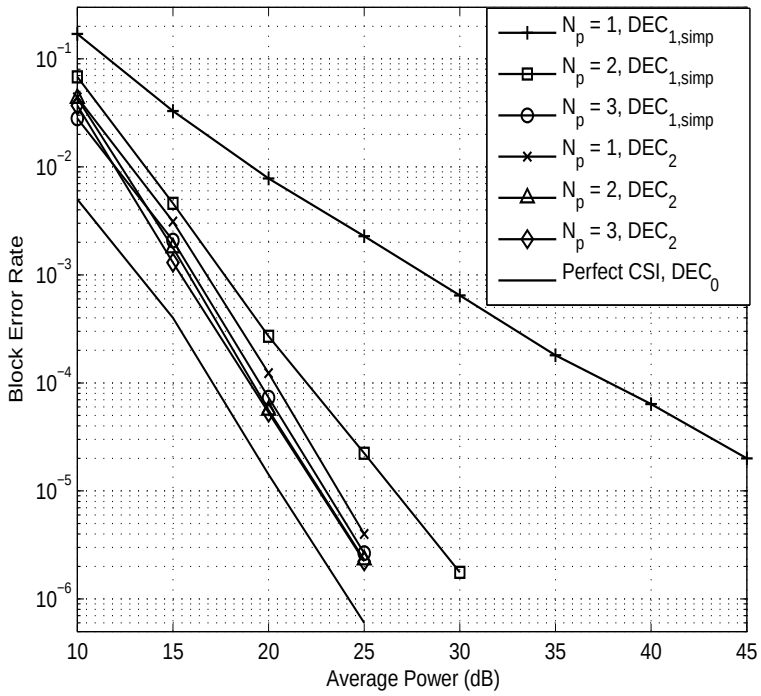

Figure 5.2: Mismatched decoding, matched decoding, and perfect CSI decoding for the network with $M = 1, R = N = 3$.

twice that of DEC_1 and 20 times that of $\text{DEC}_{1,simp}$. Compared with the perfect CSI decoding, when $N_p = 1$, DEC_2 is about 6dB worse; when $N_p = 2$, it is about 2.5dB worse.

In Fig. 5.2, the performance of the mismatched decoding $\text{DEC}_{1,simp}$, matched decoding DEC_2, and perfect CSI decoding DEC_0 are shown for the network with $R = 3$. For the matched decoding DEC_2, full diversity is achieved when $N_p = 1$. When N_p increases from 1 to 2, the performance of DEC_2 is improved by about 0.5dB. Very little improvement can be obtained with further increase in N_p. For $N_p = 1$, the advantage of DEC_2 over $\text{DEC}_{1,simp}$ is about 22.5dB at the BLER level of 2×10^{-5}, while the difference reduces to 3dB when $N_p = 2$, and is negligible when $N_p = 3$. From Table 5.2, the complexity of DEC_2 is 181 times that of $\text{DEC}_{1,simp}$. The perfect CSI decoding DEC_0 is shown as a benchmark. Due to imperfect training, for DEC_2, when $N_p = 1$ (5 symbol intervals for training), the degradation is about 3dB; when $N_p = 2$ and 3, the

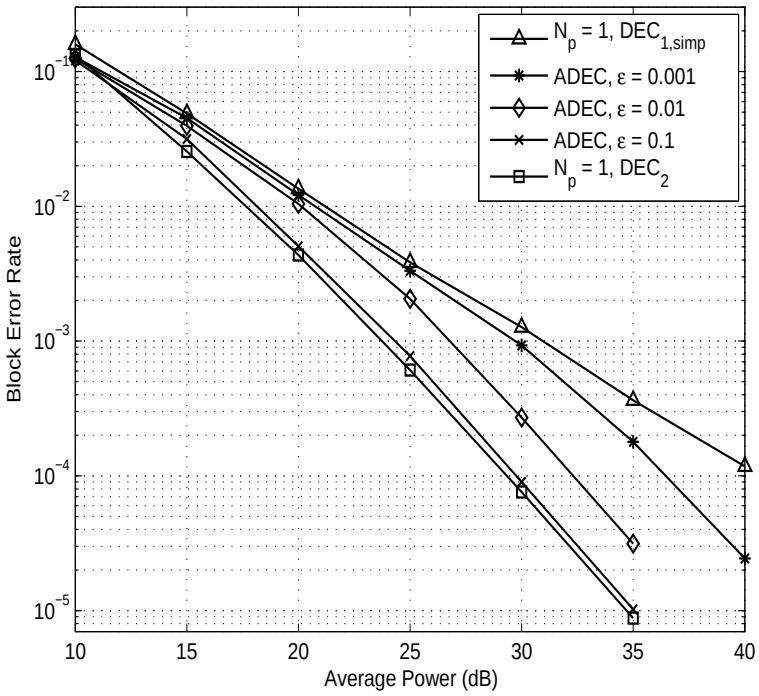

Figure 5.3: Adaptive decoding for the network with $R = 2$.

degradation is about 2dB.

In Fig. 5.3, the performance of adaptive decoding is shown for the network with $R = 2$. N_p is set to be 1, thus the minimum training length, 4 symbol intervals, is applied. ϵ is chosen as 0.001, 0.01, and 0.1. It is observed that all three A-DECs achieve full diversity. As ϵ increases, BLER decreases. BLERs of the mismatched and matched decodings are also shown for comparison. A-DEC has about the same performance as the mismatched decoding $\text{DEC}_{1,simp}$ in the low power region. But as the power increases, its BLER decreases much faster due to its full diversity. It is about 3dB better at the BLER of 10^{-4} even when ϵ is as small as 0.001. From Table 5.1, we can see that this improvement is gained with only 3 extra flops in each block decoding. When $\epsilon = 0.01$, A-DEC is about 2.5dB worse than the matched decoding DEC_2 with a complexity only 1/14 of DEC_2. When $\epsilon = 0.1$, A-DEC performs almost the same as DEC_2, but its complexity is less than 1/4 of DEC_2. Therefore, the proposed adaptive decoding is an efficient decoding scheme. By choosing a

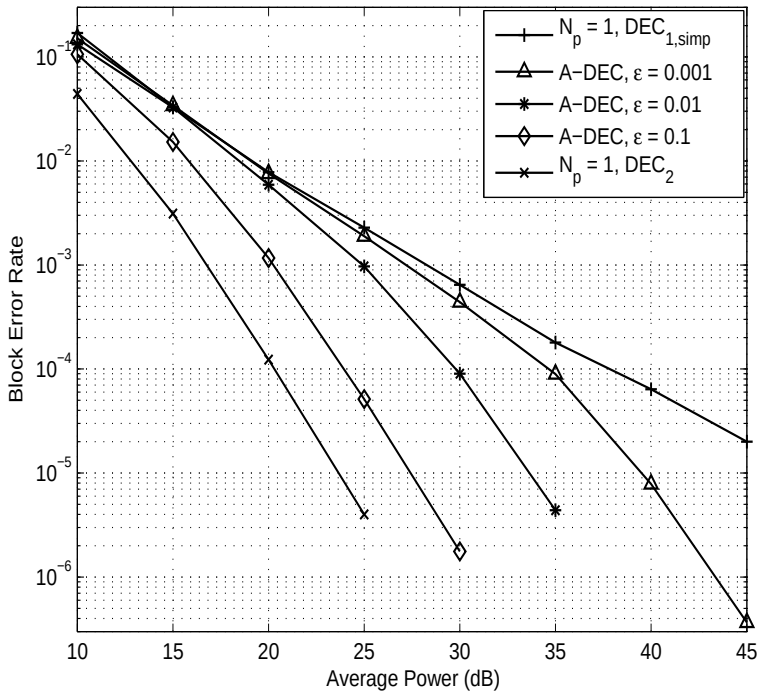

Figure 5.4: Adaptive decoding for the network with $R = 3$.

proper ϵ, we can balance reliability and complexity.

In Fig. 5.4, the performance of A-DEC is shown for the network with $R = 3$. N_p is set to be 1, i.e., the minimum training length, 5 symbol intervals, is applied. Similar observations to those in Fig. 5.3 can be seen. All A-DECs achieve the full diversity, which is 3. In contrast with $\text{DEC}_{1,simp}$ at the BLER of 2×10^{-5}, A-DEC with $\epsilon = 0.001$ has about 7dB advantage and 1.5 times complexity; A-DEC with $\epsilon = 0.01$ has about 12.5dB advantage and 6.4 times complexity; A-DEC with $\epsilon = 0.1$ has about 18.5dB advantage and 55 times complexity. In contrast with DEC_2 at the BLER of 10^{-5}, A-DEC with $\epsilon = 0.001$ has about 15.5dB disadvantage and 1/118 complexity; A-DEC with $\epsilon = 0.01$ has about 10dB disadvantage and 1/28 complexity; A-DEC with $\epsilon = 0.1$ has about 4dB disadvantage and 1/3 complexity.

5.5 Summary

For the network with $M = 1$, $R = N \geq 2$, we considered the matched decoding based on the approximated training-based system equation. It was shown that as opposed to the mismatched decoding, which only achieves diversity one when $N_p = N_{p,l} = 1$, matched decoding can achieve full diversity when $N_p = 1$. However, its complexity is prohibitively high. A modified matched decoding, adaptive decoding, was hence introduced by switching between the simplified mismatched decoding $\mathrm{DEC}_{1,simp}$ and matched decoding DEC_2 to balance the performance and complexity. The complexity analysis with respect to mismatched decoding, matched decoding, and adaptive decoding was also given. Simulation results were exhibited to demonstrate our conclusions.

6 Summary and Future Work

Cooperative relay network is very popular these days since it can exploit possible cooperation among the relay nodes in the networks to provide spatial diversity. A lot of cooperative schemes have been proposed, such as AF relaying, DF relaying, and DSTC etc. For most of the proposed cooperative schemes, accurate CSI is required at the receiver to achieve the desired performance. In practice, to get such CSI information, channel training is employed where known pilot signal is sent to track the channels.

In this thesis project, we considered channel training (Chapter 3) and decoding (Chapter 4 and Chapter 5) for MIMO relay network with multiple single-antenna relays, and multiple transmit and receive antennas. Compared with the previous work on channel training and performance study with single relay or transmit/receive antenna, design and analysis for MIMO relay network is more challenging. For one thing, the existence of multiple transmit and receive antennas can largely complicate the training design and the corresponding performance analysis; for another, due to the multiple communication stages, channel training for one stage could be coupled with channel trainings for other stages, which leads to a non-Gaussian estimation model making analysis even more involved.

In Chapter 3, channel training was investigated for MIMO relay networks. The objective of channel training is to estimate the transmitter-relay channel vector $\mathbf{f}$, the relay-receiver channel matrix $\mathbf{G}$, and the end-to-end channel matrix $\mathbf{H}$ at the receiver, where the LMMSE estimator is empoyed. Note that, with our training scheme, the relays are not required to be equipped with estimators. The associated training design, which includes training scheme design, training code design, training time design, and power allocation, were discussed respectively. We showed that, since the relay-receiver link is a virtual multiple-antenna system, the training design results in [2] are used. For the training of $\mathbf{f}$, DSTC scheme was employed. Since the equivalent pilot signal

in estimating $\mathbf{f}$ needs the information of $\mathbf{G}$, the training of $\mathbf{f}$ was first discussed for the MIMO relay network by assuming the estimated $\mathbf{G}$ as perfect. Then, considering the estimation error of $\mathbf{G}$, the training of $\mathbf{f}$ was revisited for the specific network with $M = 1, R = N \geq 2$. To estimate $\mathbf{H}$, two training schemes, called separate training and direct training, were provided and compared.

Employing the channel estimations provided in Chapter 3, we investigated decoding strategies for the training-based DSTC network. Two coherent decodings were considered: mismatched decoding (Chapter 4) in which channel estimations are treated as if perfect, and matched decoding (Chapter 5) in which estimation error is taken into account.

For mismatched decoding DEC_1, diversity performance was first discussed for the MIMO relay network neglecting the estimation error of $\mathbf{G}$. It was shown that, with $N_p = N_{p,l}$, the shortest training time in estimating $\mathbf{f}$, full diversity cannot always be achieved in data transmission. To achieve full diversity, an upper bound on the minimum N_p was given as MR. To shorten the training time while maintain the full diversity, an adaptive training was provided whose training time length is adaptive to the relay-receiver channel matrix $\mathbf{G}$. Then, by considering the estimation error of $\mathbf{G}$, diversity of DEC_1 was analyzed for the specific network with $M = 1$ and $R = N \geq 2$. It was shown that, with the shortest training length $R+2$ symbol intervals, the diversity order is no larger than 1, and at least $3R$ symbol intervals are required in training to achieve full diversity.

The matched decoding DEC_2 was considered for the specific network with $M = 1$, $R = N \geq 2$. We showed that, matched decoding can achieve full diversity when $N_p = N_{p,l} = 1$. However, its complexity is prohibitively high. A modified matched decoding, adaptive decoding, was hence introduced by switching between the simplified mismatched decoding $\mathrm{DEC}_{1,simp}$ and matched decoding DEC_2 to balance the performance and complexity.

In the following, we list and summarize some future works related to this

thesis project.

- Extension of channel training to correlated channel model: In this work, we only considered the i.i.d. channel model. As discussed in Subsection 3.6.1, when channel correlation is considered, our training scheme cannot be fully applied even when the correlation matrices are known at the receiver. Further investigation on training design such as training code design and training time design is needed. When the channel correlation matrices are unknown at the receiver, the channel estimation problem should also include the estimation of correlation matrices, where the associated training design is needed.

- Extension of training of $\mathbf{f}$ with estimated $\mathbf{G}$ to general network: When the estimation error of $\mathbf{G}$ is considered in estimating $\mathbf{f}$, our analysis is focused on for the specific network with $M = 1$, $R = N \geq 2$. In Subsection 3.4.4, we discussed the extension to the general network, where the main difficulty lies in the analysis of training properties such as $\mathrm{MSE}(\hat{\mathbf{f}})$. To perform further analysis, more involved computation is needed.

References

[1] H. Jafarkhani, *Space-Time Coding: Theory and Practice*, Cambridge University Press, 2005.

[2] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inf. Theory.*, vol. 49, pp. 951–963, Apr. 2003.

[3] J. G. Proakis, *Digital Communication*, 5th edition, McGraw-Hill, 2007.

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

[5] J. Mietzner, R. Schober, L. Lampe, W. H. Gerstacker, and P. A. Hoeher, "Multiple-antenna techniques for wireless communications a comprehensive literature survey," *IEEE Commun. Surveys Tuts.*, vol. 11, pp. 87–105, second quarter, 2009.

[6] P. Wolniansky, J. Foschini, G. Golden, and R. Valenzuela, "V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel," *in Proc. URSI ISSSE*, pp. 295–300, Sep., 1998.

[7] S. Loyka, and F. Gagnon, "Performance analysis of the V-BLAST algorithm: an analytical approach," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 768–783, Jul., 2004.

[8] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. Jourl.*, vol. 1, pp. 41–59, 1996.

[9] L. C. Godara, "Application of antenna arrays to mobile communications Part I: performance improvement, feasibility, and system considerations; Part II: beam-forming and direction-of-arrival considerations," *Proc. IEEE*, vol. 85, pp. 1031–1060, Jul. 1997.

[10] L. C. Godara, "Application of antenna arrays to mobile communications Part II: beam-forming and direction-of-arrival considerations," *Proc. IEEE*, vol. 85, pp. 1195–1245, Aug. 1997.

[11] Y. Jing and H. Jafarkhani, "Using orthogonal and quasi-orthogonal designs in wireless relay networks," *IEEE Trans. Inf. Theory*, vol. 53, pp. 4106–4118, Nov. 2007.

[12] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.

[13] M. K. Simon, and M. S. Alouini, *Digital Communication over Fading Channels*, 2th edition, Wiley-IEEE Press, 2004.

[14] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity-part I: System description," *IEEE Trans. Commun.*, vol. 51, pp. 1927–1938, Nov. 2003.

[15] A. Nosratinia, T. E. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Commun. Mag.*, pp. 74–80, Oct. 2004.

[16] A. Scaglione, D. L. Goeckel, and J. N. Laneman, "Cooperative communications in mobile Ad Hoc networks," *IEEE Signal Process. Mag.*, vol. 23, pp. 18–29, Sep. 2006.

[17] R. U. Nabar, H. Bolcskei, and F. W. Kneubuhler, "Fading relay channels: Performance limits and space-time signal design," *IEEE Jourl. Selec. Areas on Commu.*, vol. 22, pp. 1099–1109, Aug. 2004.

[18] A. Bletsas, D. P. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," *IEEE Jourl. Selec. Areas on Commu.*, vol. 24, pp. 659–672, Mar. 2006.

[19] R. Madan, N. B. Mehta, A. F. Molisch, and J. Zhang, "Energy-efficient cooperative relaying over fading channels with simple relay selection," in *IEEE Trans. Wireless Commun.*, vol. 7, pp. 3013–3025, Aug. 2008.

[20] Y. Jing, and H. Jafarkhani, "Single and multiple relay selection schemes and their achievable diversity orders," in *IEEE Trans. Wireless Commun.*, vol. 8, pp. 1414–1423, Mar. 2009.

[21] Y. Jing and H. Jafarkhani, "Network beamforming using relays with perfect channel information," *IEEE Trans. Inf. Theory*, vol. 55, pp. 2499–2517, Jun. 2009.

[22] V. Havary-Nassab, S. Shahbazpanahi, A. Grami, and Z. Luo, "Distributed Beamforming for relay networks based on second-order statistics of the channel state information," in *IEEE Trans. Signal Process.*, vol. 56, pp. 4306–4316, Sept. 2008.

[23] H. El. Gamal and D. Aktas, "Distributed space-time filtering for cooperative wireless networks," in *Proc. of IEEE GLOBECOM,* pp. 1826–1830, Dec. 2003.

[24] Y. Jing and B. Hassibi, "Distributed space-time coding in wireless relay networks," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 3524–3536, Dec. 2006.

[25] Y. Jing and B. Hassibi, "Diversity analysis of distributed space-time codes in relay networks with multiple transmit/receive antennas," *EURASIP J-ASP*, 2008.

[26] J. N. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless network," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2415–2425, Oct. 2003.

[27] F. Oggier and B. Hassibi, "An algebraic family of distributed space-time codes for wireless relay networks," in *Proc. of IEEE ISIT*, July 2006.

[28] T. Kiran and B. S. Rajan, "Distributed space-time codes with reduced decoding complexity," in *Proc. of IEEE ISIT*, July 9-14 2006.

[29] J. N. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless network," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2415–2425, Oct. 2003.

[30] F. Oggier and B. Hassibi, "A coding strategy for wireless networks with no channel information," in *Proc. of Allerton Conference*, Sep. 2006.

[31] S. G. Rajan and B. S. Rajan, "Noncoherent low-decoding-complexity space-time codes for wireless relay networks," in *Proc. of IEEE ISIT,* pp. 1521–1525, June 24-29 2007.

[32] Y. Jing and H. Jafarkhani, "Distributed differential space-time coding in wireless relay networks," *IEEE Trans. Commun.*, vol. 56, pp. 1092–1100, July 2008.

[33] T. Kiran and B. S. Rajan, "Partial-coherent distributed space-time codes with differential encoder and decoder," in *Proc. of IEEE ISIT*, July 2006.

[34] P. V. Kumar, K. Vinodh, M. Anand, and P. Elia, "Diversity-multiplexing gain tradeoff and DMT-optimal distributed space-time codes for certain cooperative communication protocols: overview and recent results," in *Proc. of ITA Wokshop*, Jan. 2007.

[35] Y. Li and X. G. Xia, "A family of distributed space-time trellis codes with asynchronous cooperative diversity," *IEEE Trans. Commun.*, vol. 55, pp. 790–800, Apr. 2007.

[36] Y. Li., W. Zhang, and X. G. Xia, "Distributive high-rate space-frequency codes achieving full cooperative and multipath diversities for asynchronous cooperative communications," in *Proc. of IEEE ISIT*, pp. 2612–2616, July 2006.

[37] X. Guo and X. G. Xia, "A distributed space-time coding in asynchronous wireless relay networks," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 1812–1816, May 2008.

[38] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmissions," *IEEE Trans. Signal Process.*, vol. 21, pp. 12–25, Nov. 2004

[39] M. Médard, "The effect upon channel capacity in wireless communications of perfect and Imperfect knowledge of the channel," *IEEE Trans. Inf. Theory.*, vol. 46, pp. 933–946, May 2000.

[40] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inf. Theory.*, vol. 52, pp. 2203–2214, May 2006.

[41] A. Lapidoth and S. Shamai, "Fading channels: How perfect need "perfect side information" be? " *IEEE Trans. Inf. Theory.*, vol. 48, pp. 1118–1134, May 2002.

[42] V. Tarokh, A. Naguib, N. Seshadri, and A. R. Calderbank, "Space-time codes for high-data-rate wireless communication: Performance criteria in the presence of channel estimation errors, mobility, and multiple paths," *IEEE Trans. Commun.*, vol. 47, pp. 199–207, Feb.1999

[43] G. Taricco and E. Biglieri, "Space-time decoding with imperfect channel estimation," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 1874–1888, Jul. 2005.

[44] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, pp. 884–893, Mar. 2006.

[45] A. Vosoughi, and A. Scaglione, "Everything you always wanted to know about training: Guidelines derived using the affine precoding framework

and the CRB," *IEEE Trans. Signal Process.*, vol. 54, pp. 940–954, Mar. 2006.

[46] A. Vosoughi, and A. Scaglione, "On the effect of receiver estimation error upon channel mutual information," *IEEE Trans. Signal Process.*, vol. 54, pp. 459–472, Feb. 2006.

[47] F. Mazzenga, "Channel estimation and equalization for M-QAM transmission with a hidden pilot sequence," *IEEE Trans. Broadcasting*, vol. 46, pp. 170–176, Jun. 2000.

[48] R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, 1968.

[49] A. Viterbi and J. Omura, *Principles of Digital Communication and Coding*, McGraw-Hill, 1979.

[50] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai (Shitz), "On information rates for mismatched decoders," *IEEE Trans. Inf. Theory*, vol. 40, pp. 1953–1967, Nov. 1994.

[51] E. G. Larsson, "Diversity and channel estimation errors," *IEEE Trans. Comm.*, vol. 52, pp. 205–208, Feb. 2004.

[52] S. Kay, "Fundamentals of Statistical Signal Processing: Estimation Theory," *Englewood Cliffs, NJ: Prentice-Hall*, 1993.

[53] L. G. Ordóñez, D. P. Palomar and J. R. Fonollosa, "Ordered eigenvalues of a general class of hermitian random matrices with application to the performance analysis of MIMO Systems," *IEEE Trans. Signal Process.*, vol. 57, pp. 672–688, 2009.

[54] A. Edelman, *Eigenvalues and condition numbers of random matrices*, Ph.D. thesis, Massachusetts Institute of Technology, Department of Mathematics, 1989.

[55] H. Mheidat and M. Uysal, "Non-coherent and mismatched-coherent receivers for distributed STBCs with amplify-and-forward relaying," *IEEE Trans. Wireless Comm.*, vol. 6, pp. 4060–4070, Nov. 2007.

[56] B. Gedik and M. Uysal, "Impact of imperfect channel estimation on the performance of amplify-and-forward relaying," *IEEE Trans. Wireless Comm.*, vol. 8, pp. 1468–1479, Mar. 2009.

[57] O. Amin, B. Gedik, and M. Uysal, "Channel estimation for amplify-and-forward relaying: Cascaded against disintegrated estimators," *IET Commun*, vol. 4, pp. 1207–1216, 2010.

[58] F. Gao, T. Cui, and A. Nallanathan, "On channel estimation and optimal training design for amplify and forward relay networks," *IEEE Trans. Wireless Comm.*, vol. 7, pp. 1907–1916, May 2008.

[59] F. Gao, T. Cui, and A. Nallanathan, "Optimal training design for channel estimation in decode-and-forward relay networks with individual and total power constraints," *IEEE Trans. Signal Process.*, vol. 56, pp. 5937–5949, Dec. 2008.

[60] F. Gao, R. Zhang, and Y. Liang, "On channel estimation for amplify-and-forward two-way relay networks," *Proc. of IEEE GLOBECOM*, Nov. 2008.

[61] C. S. Patel, and G. L. Stüber, "Channel estimation for amplify and forward relay based cooperation diversity systems," *IEEE Trans. Wireless Comm.*, vol. 6, pp. 2348–2356, Jun. 2007.

[62] G. S. Rajan and B. S. Rajan, "Leveraging coherent distributed space-time codes for noncoherent communication in relay networks via training," *IEEE Trans. Wireless Comm.*, vol. 8, pp. 683–688, Feb. 2009,

[63] S. Sun and Y. Jing, "Channel training and estimation in distributed space-time coded relay networks with multiple transmit/receive antennas," *Proc. of IEEE WCNC*, Apr. 2010.

[64] J. P. Kermoal, L. Schumacher, K. I. Pedersen, and P. E. Mogensen, "A stochastic MIMO radio channel model with experimental validation," *IEEE Jourl. Selec. Areas on Commu.*, pp. 1211–1226, Aug. 2002.

[65] E. Telatar, "Capacity of multi-antenna Gaussian channels," *AT & T-Bell Labs Internal Tech. Memo.*, Jun. 1995

[66] N. Higham, *Accuracy and stability of numerical algorithms*, SIAM, 2nd edition, 2002.

[67] G. Golub and C. Van Loan, *Matrix computations*, Hopkins Fulfillment Service, 3rd edition, 1996.

[68] R. A. Horn, and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1986.