Channel State Information in Multiple Antenna Systems

A Thesis
Presented to
The Academic Faculty

by

Jingnong Yang

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

School of Electrical and Computer Engineering
Georgia Institute of Technology
December 2006

Channel State Information in Multiple Antenna Systems

Approved by:

Dr. Douglas B. Williams, Chair
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Gregory D. Durgin
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. John R. Barry
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Alfred D. Andrew
School of Mathematics
Georgia Institute of Technology

Dr. Mary Ann Ingram
School of Electrical and Computer Engineering
Georgia Institute of Technology

Date Approved: 21 August 2006

This dissertation is dedicated to my wife Jijie

for her patience, love, encouragement, and support.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Prof. Douglas B. Williams for being a fantastic advisor. Prof. Williams always has great suggestions about research. I have always been greatly inspired by his research style, intellectual aspects, and insights. He has also been a good mentor concerning other aspects of life. I have really enjoyed discussing my future with him.

I also would like to express my gratitude to my thesis committee members, Prof. Mary Ann Ingram, Prof. John Barry, and Prof. Gregory Durgin for their guidance.

I would like to thank my great group-mates, Soner Özgür and Nicolas Gastaud. I wish to show my gratitude to Qing Zhao, who is one of my closest friends. I truly enjoyed every conversation I had with him about research and the future. In addition, inspiring talks with Jianxuan Du, Taewon Hwang, Guocong Song, Hua Zhang, Ning Chen, and Hua Qian, to name a few, have always been rewarding.

Finally, I would like to express my deepest gratitude to my family.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

The information-theoretic foundation of *multiple-input multiple-output* (MIMO) systems was laid out by Foschini, Gans, and Telatar [21, 74], who have showed that multiple antennas at the transmitter and the receiver provide significant capacity enhancement over single-antenna systems. Spatial diversity provided by multiple antennas enhances the throughput and reliability of wireless communications [83]. To exploit the enhanced spectral efficiency, space-time coding has been designed to achieve a specific tradeoff between diversity and multiplexing [23, 72, 73]. Most work on space-time coding deals with the case where no knowledge of the forward channel is available to the transmitter.

In a MIMO system, if the transmitter has perfect knowledge of the underlying *channel state information* (CSI), power allocation to the right singular subspace of the channel matrix can be used to achieve a higher channel capacity compared to transmission without CSI [27]. When reciprocity of the wireless channel does not hold, as in *frequency-division duplex* (FDD), perfect CSI at the transmitter requires a high-rate feedback channel, which may not be practical, particularly in fast time-varying environments. Thus, the identification and utilization of partial CSI at the transmitter are important issues.

Much work has been devoted to identifying the benefits of partial CSI at the transmitter and the design of optimal transmission schemes to exploit it. For example, when only the statistics of the channel state are available at the transmitter, an optimal transmit covariance matrix can be designed to achieve higher capacity than transmission without any CSI [76, 85, 86]. Techniques for attaining partial CSI have also been proposed [44, 53, 54].

This thesis is focused on partial CSI acquisition and utilization techniques for MIMO channels. The nature of the CSI feedback problem is a quantization of the underlying matrix channels. We propose a feedback algorithm for tracking the dominant channel subspaces for MIMO systems in a continuously time-varying environment. We exploit the correlation between channel states of adjacent time instants and quantize the variation of channel states.

Specifically, we model a subspace as one point in a Grassmann manifold, treat the variations in principal right singular subspaces of the channel matrices as a piecewise-geodesic process in the Grassmann manifold, and quantize the velocity matrix of the geodesic.

As a demonstration of optimal transmitter design given partial CSI feedback, we design a complexity-constrained MIMO OFDM system where the transmitter has knowledge of channel correlations. The transmitter is constrained to perform at most one *inverse Discrete Fourier Transform* (IDFT) per OFDM symbol on the average. We show that in the multiple input, single output case, time domain beamforming can be used to do two-dimensional eigen-beamforming. For the MIMO case, we derive design criteria for the transmitter beamforming and receiver combining weighting vectors and show some suboptimal solutions.

Most previous papers on CSI feedback did not consider uncertainties in the feedback process, such as unexpected delay or error in the feedback channel. Such uncertainties exist in reality and ignoring them results in suboptimal algorithms. We consider channel mean-feedback with an unknown delay and propose a broadcast approach that is able to adapt to the quality of the feedback.

Having considered CSI feedback problems where the receiver tries to convey its attained CSI to the transmitter, we turn to a different problem; namely, noncoherent coding design for fast fading channels, where the receiver does not have reliable CSI. Unitary space-time codes [31,82] and training based schemes [16] have been proposed historically. We propose a data-dependent superimposed training scheme to improve the performance of training based codes. The transmitter is equipped with multiple training sequences and dynamically selects a training sequence for each data sequence to minimize channel estimation error. The set of training sequences are optimized to minimize pairwise error probability between codewords.

CHAPTER I

INTRODUCTION

1.1 MIMO Wireless Communications

The biggest challenge for a reliable wireless communication system is to combat multipath fading. Spatial diversity introduced by deploying multiple antennas at the transmitter and/or the receiver is an effective technique to overcome the fading effect. In addition to reliability improvement through diversity, a MIMO system can also increase system throughput by transmitting multiple data streams. The result of improved reliability and/or throughput [83] is enhanced spectral efficiency and higher channel capacity.

The information-theoretic foundation of MIMO systems was laid out by Foschini, Gans, and Telatar [21, 74], who have shown that multiple antennas at the transmitter and the receiver provide significant capacity enhancement over single-antenna systems. To exploit MIMO channel capacity, space-time coding has been designed to achieve a specific tradeoff between diversity and multiplexing [23, 72, 73].

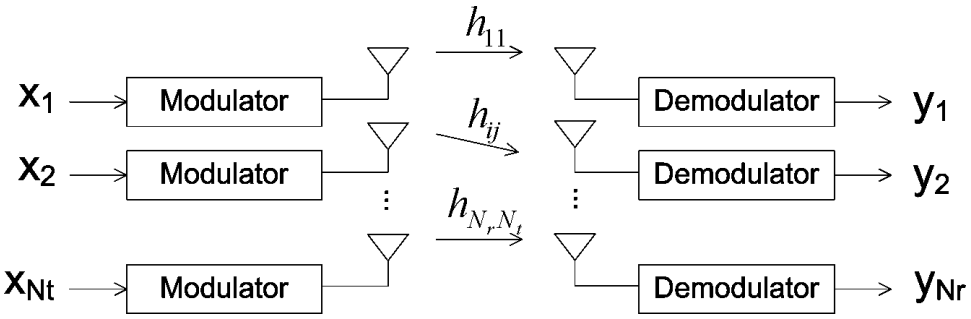


Figure 1: Diagram of a MIMO system.

Fig. 1 illustrates a MIMO channel. The complex baseband equivalent of a flat-fading MIMO channel with N_r receive antennas and N_t transmit antennas can be expressed as

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{H}$ is an $N_r \times N_t$ complex channel transfer matrix, $\boldsymbol{x}$ and $\boldsymbol{y}$ are the $N_t \times 1$ and $N_r \times 1$

transmit and receive vectors, respectively. The vector $\boldsymbol{n}$ is an $N_r \times 1$ additive noise vector, distributed as $\mathcal{CN}(\mathbf{0}, \sigma^2 \boldsymbol{I})$, i.e., complex Gaussian with zero-mean and covariance matrix $\sigma^2 \boldsymbol{I}$. The noise is independent of the transmit signal.

As shown by Telatar [74], if the receiver has perfect CSI, for fixed $\boldsymbol{H}$, among all input distributions with a given covariance matrix $\boldsymbol{Q}$, the Gaussian distribution $\boldsymbol{x} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{Q})$ maximizes the mutual information between $\boldsymbol{x}$ and $\boldsymbol{y}$, which is given by

$$\log \det \left(\boldsymbol{I} + \frac{1}{\sigma^2} \boldsymbol{H} \boldsymbol{Q} \boldsymbol{H}^H \right), \tag{2}$$

where $A \sim B$ means that A is distributed as B.

For Rayleigh fading rich scattering channels, the entries of channel matrix $\boldsymbol{H}$ are *independent and identically distributed* (i.i.d.) Gaussian random variables with zero-mean. The ergodic capacity for random channel matrices is

$$C_{\text{erg}} = E \left\{ \log \det \left(\boldsymbol{I} + \frac{1}{\sigma^2} \boldsymbol{H} \boldsymbol{Q} \boldsymbol{H}^H \right) \right\}, \tag{3}$$

where the expectation is with respect to the channel matrices. Information theoretic results [21] have shown that the ergodic capacity increases linearly as the minimum of the number of receive and transmit antennas at high signal-to-noise ratios.

1.2 OFDM Systems

In a wideband wireless system, the channel shows frequency-selectivity, and the delay spread of the channel impulse response causes *intersymbol interference* (ISI), which is a challenge for the designers of single carrier communication systems. *Orthogonal frequency division multiplexing* (OFDM) [3] is a promising technology for high rate data transmission in wideband wireless systems, because it can mitigate the effects of frequency selectivity and ISI. OFDM systems achieve this mitigation by transmitting data symbols on parallel mutually orthogonal subchannels in the frequency domain. OFDM has been adopted as the basis of several wireless LAN standards such as ETSI-BRAN HIPERLAN/2, IEEE 802.11a, and IEEE 802.11n. OFDM is also being considered as a promising candidate for future generation high-speed digital communication systems such as IEEE 802.16e (Mobile WiMAX) [77].

Multiple antennas can be used with OFDM to improve system performance [38]. Traditionally, subcarrier based beamforming and combining are used [60]. Since the MIMO channels on different subcarriers are different, beamforming and combining vectors are different on each subcarrier. Consequently, each antenna requires a separate DFT or IDFT operation, resulting in great system complexity, especially when the number of deployed antennas is large.

1.3 Closed-Loop MIMO Systems

Most work on space-time coding deals with the case where no knowledge of the forward channel is available to the transmitter. But, if the transmitter has perfect CSI, power allocation to the right singular subspace of the channel matrix can be used to achieve a higher channel capacity compared to transmission without CSI [27].

To illustrate the benefit of perfect CSI, we first consider the following *multiple-input single-output* (MISO) channel,

$$y = \boldsymbol{h}\boldsymbol{x} + n, \tag{4}$$

where the received signal y and noise sample n are scalars, and the channel becomes a row vector $\boldsymbol{h}$. If the transmitter has no CSI, Telatar [74] showed that the optimal input covariance matrix is proportional to an identity matrix, i.e., $\boldsymbol{Q} = P/N_t\boldsymbol{I}$, where P is a constraint on the total transmit power. From (2), the corresponding channel capacity is given by

$$\log\det\left(1 + \frac{P}{N_t\sigma^2}\boldsymbol{h}\boldsymbol{h}^H\right) = \log\left(1 + \frac{P\|\boldsymbol{h}\|^2}{N_t\sigma^2}\right). \tag{5}$$

On the other hand, if the transmitter has perfect CSI, it is apparent that the optimal transmitter is a beamformer. The optimal transmit covariance matrix is $\boldsymbol{Q} = \frac{P}{\|\boldsymbol{h}\|^2}\boldsymbol{h}^H\boldsymbol{h}$, and the corresponding channel capacity is given by

$$\log\det\left(1 + \frac{P\|\boldsymbol{h}\|^2}{\sigma^2}\right). \tag{6}$$

Equations (5) and (6) show that a transmitter with perfect CSI achieves an N_t-fold *signal-to-noise ratio* (SNR) gain compared to a transmitter without CSI in a MISO channel.

Next we consider a MIMO channel where the transmitter has more transmit antennas than receive antennas. Let the *singular value decomposition* (SVD) of the channel matrix be

$$\boldsymbol{H} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^H, \tag{7}$$

where $\boldsymbol{\Lambda}$ has on its main diagonal the singular values λ_i; $i = 1, 2, \ldots, N_t$; of $\boldsymbol{H}$. When $N_t \geq N_r$, $\boldsymbol{H}$ has at most N_r nonzero singular values. The mechanism of this MIMO system is illustrated in Fig. 2. If the transmitter has no CSI, we can see from Fig. 2 that signal components along the directions corresponding to zero singular values are nulled by the channel and can never reach the receiver. The power of these signal components is wasted. On the other hand, when the transmitter has perfect CSI, it can transmit along the directions corresponding to nonzero singular values without suffering any signal loss.

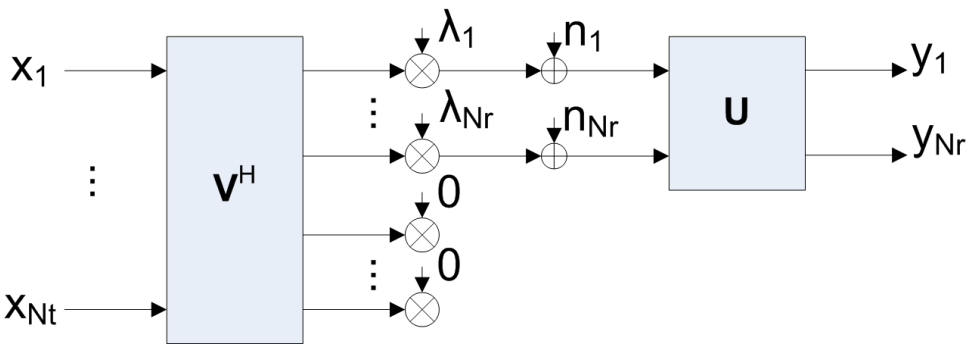


Figure 2: Illustration of a MIMO channel through SVD of the channel matrix.

Although CSI is valuable for transmitters with multiple antennas, perfect CSI is not always available at the transmitter. When reciprocity of the wireless channel does not hold, as in FDD, perfect CSI at the transmitter requires a high-rate feedback channel, which may not be practical, particularly in fast time-varying environments. These cases motivate our research for effective acquisition and utilization of partial CSI at the transmitter in practical MIMO systems.

1.4 Noncoherent Communication Systems

Reliable digital communication over fast time-varying wireless channels is becoming more and more important in view of the rapid growth of cellular and personal communications

systems. If the fading is too fast for the receiver to obtain reliable CSI, the system becomes a noncoherent one because coherent decoding at the receiver can no longer be done.

Information theoretically optimal noncoherent coding for block fading channels was considered in [31,82]. At high SNR, the optimal codewords are shown to be so-called *unitary space-time codes*, and, therefore, are unstructured. At the receiver, decoding can only be done through an exhaustive search through the codebook. The optimal signal distribution at medium ranges of SNR remains an open problem. Codebook construction for frequency-selective fading channels from a *pairwise error probability* (PEP) viewpoint was proposed in [12, 65]. Because of the complexity of codeword optimization, only *binary phase-shift keying* (BPSK) modulated binary codes were designed in [12, 65]. Such codebooks are obviously not optimal at high SNR, where higher order modulation than BPSK is required to achieve capacity.

Considering decoding complexity, training based schemes are attractive. Another name for such schemes is *pilot symbol aided modulation* (PSAM). Traditionally, training symbols are multiplexed with data symbols. For example, in flat-fading channels, pilot symbols are *time-division multiplexed* (TDM) with data symbols. In frequency-selective channels, pilot symbols are *frequency-division multiplexed* (FDM) with data symbols and modulate a selected set of subcarriers in OFDM systems. In multiuser CDMA systems, pilot symbols modulate a dedicated spreading code and are *code-division multiplexed* (CDM) with data symbols. The receivers in a training based system can first estimate the channel using training sequences and then do the decoding coherently, using the channel estimate as if it is the true channel. One such training based scheme has been shown to be superior to unitary space-time codes in certain situations [16].

As a generalization of CDM training, superimposing pilot symbols with data symbols has been proposed [20,47,84]. By adding pilot symbols and data symbols together, a higher information rate can be possibly achieved compared to TDM or FDM training, because superimposed training does not occupy separate dimensions in the time or frequency domain. However, a rigorous comparison of error rates achievable by superimposed training and TDM or FDM training was never done. As we will show in Section 6.4, an inappropriately

designed superimposed training sequence may result in very high error probability if the channel coherence time is very short. Another drawback of superimposed training is that the channel estimate is not as accurate as in TDM or FDM training, because data symbols appear as high variance noise to the channel estimator. Improvement upon superimposed training schemes is another focus point of our research.

1.5 Thesis Outline

Chapter 2 gives an extensive review of closed-loop MIMO techniques, including CSI feedback techniques and transmitter designs given partial CSI. Additionally, a brief description of the geometry of Grassmann manifold is given in Section 2.3.

The new contributions of this thesis are presented in detail in the next four chapters.

- In Chapter 3, several new partial CSI acquisition algorithms are proposed for performance improvement on the Grassmannian subspace packing scheme, including a one-bit feedback scheme and a *vector quantization* (VQ) scheme.

- In Chapter 4, we design a complexity-constrained MIMO OFDM system where the transmitter has the knowledge of channel correlations. Transmitter beamforming and receiver combining are performed in the time domain, and only one (I)DFT per OFDM block is needed at either the transmitter or the receiver.

- In Chapter 5, we consider a CSI feedback channel with an unknown delay, which makes the CSI at the transmitter arbitrarily unreliable. We propose a broadcast strategy to adapt reliably transmitted information rate to the quality of CSI.

- In Chapter 6, we propose a data-dependent superimposed training scheme to improve the performance of training based codes. The transmitter is equipped with multiple training sequences and dynamically selects a training sequence for each data sequence to minimize channel estimation error. The set of training sequences are optimized to minimize pairwise error probability between codewords.

A complete list of acronyms and mathematical notations can be found in Appendix C.

CHAPTER II

CLOSED-LOOP MIMO SYSTEMS

In a MIMO channel where the reciprocity of uplink and downlink does not hold, the transmitter cannot derive CSI for the downlink from the received uplink signal. In this case, for the system to achieve a higher capacity than that achievable by transmission without CSI, the downlink receiver has to feedback CSI to the transmitter. Perfect CSI feedback usually requires a high-rate feedback channel, which may not be practical, particularly in fast time-varying environments. Thus, the feedback and utilization of partial CSI become important issues.

The data transmission in the downlink and CSI feedback through the uplink constitute a closed-loop system, as depicted in Fig. 3. Vector $\boldsymbol{f}$ is a representation of the partial CSI. The feedback channel may introduce delays or errors to the partial CSI, making $\boldsymbol{f}'$, a possibly unreliable copy of partial CSI, available to the transmitter. Lots of interesting problems arise from this closed-looped system model, including

1. What partial CSI should be fed back?
2. What is the most efficient way to represent the partial CSI?
3. How to design transmitters given partial CSI?
4. How to deal with unreliable CSI caused by delays or errors in the feedback channel?

Chapter 3 is related to questions 1 and 2 above, Chapter 4 is related to question 3, and Chapter 5 tries to answer question 4.

In this chapter, we review closed-loop MIMO systems, including existing CSI feedback techniques and optimal transmitter designs given partial CSI. Additionally, a brief description of the geometry of Grassmann manifold is given in Section 2.3, which is a necessary mathematical preparation for Chapter 3.

2.1 Transmitter Design with Partial CSI

As shown in Fig. 3, the downlink transmitter receives partial CSI feedback $\boldsymbol{f}'$ from the receiver. The true channel matrix, which the transmitter does not fully know, can be modeled as a Gaussian random matrix (or vector) whose mean and covariance is given in the feedback. This section reviews the design of the optimal transmitter in a MISO channel, described in (4), where the transmitter knows that the channel vector is distributed as $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

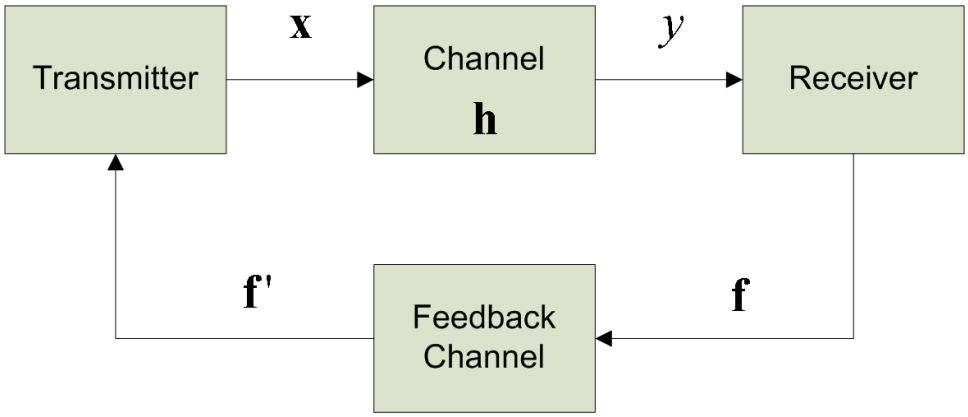


Figure 3: Feedback system model.

The problem is, for $\boldsymbol{h} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, to find the input distribution $p(\boldsymbol{x})$ that maximizes the mutual information between $\boldsymbol{x}$ and y, subject to a power constraint $E\|\boldsymbol{x}\|^2 \leq P$. For the case of a fixed channel, the optimal input distribution is zero-mean Gaussian. The optimization problem is now one of finding the optimal choice of covariance matrix $\boldsymbol{Q}$ of $\boldsymbol{x}$ maximizing the mutual information between $\boldsymbol{x}$ and y averaged over $\boldsymbol{h}$ with power constraint P:

$$\max_{\boldsymbol{Q}, \mathrm{trace}(\boldsymbol{Q})=P} E_{\boldsymbol{h}} \log\left(1 + \frac{\boldsymbol{h}\boldsymbol{Q}\boldsymbol{h}^H}{\sigma^2}\right), \tag{8}$$

where $E_{\boldsymbol{h}}$ denotes expectation with regard to the distribution of $\boldsymbol{h}$.

The transmission strategy is completely characterized by the covariance matrix $\boldsymbol{Q}$. The strategy consists of transmitting independent complex Gaussian symbols along the directions specified by the eigenvectors of $\boldsymbol{Q}$, with the corresponding eigenvalues specifying the powers allocated to each direction.

Presently, the general solution to the optimization problem in (8) for the general form of $\boldsymbol{h} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is not known. Visotsky and Madhow [76] gave solutions for the following two cases.

- Mean feedback. In this case the channel is modeled at the transmitter as $\boldsymbol{h} \sim \mathcal{CN}(\boldsymbol{\mu}, \alpha \boldsymbol{I})$, where the mean $\boldsymbol{\mu}$ can be interpreted as an estimate of the channel given the feedback, and α can be interpreted as the variance of the estimate. Mean feedback also models the time-division duplex case, where the transmitter estimates the uplink channel and uses the estimate as the mean of the downlink channel. The variance arises from the delay between the channel estimation time and the time when such CSI is used for transmission.

- Covariance feedback. In this case, the channel known to the transmitter is $\boldsymbol{h} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{\Sigma})$. This models a situation in which the channel may be varying too rapidly for the feedback to give an accurate estimate of the current channel value. The covariance matrix models the geometry of the propagation path, which changes more slowly.

Next we list the solutions of the above two cases. Let $\boldsymbol{Q}^0$ be the optimal covariance matrix of $\boldsymbol{x}$, and let $\boldsymbol{Q}^0 = \boldsymbol{U}^0 \boldsymbol{\Lambda}^0 \boldsymbol{U}^{0H}$ be the SVD of $\boldsymbol{Q}^0$. The optimum solution of mean feedback is to use $\boldsymbol{\mu}/\|\boldsymbol{\mu}\|$ as one of the singular vectors of $\boldsymbol{Q}^0$, and arbitrary orthonormal vectors as the remaining singular vectors. The power allocation among the directions is given by water-filling. For covariance feedback, the optimal $\boldsymbol{U}^0$ is the same as the singular-vector matrix of $\boldsymbol{\Sigma}$, and the singular values are given by water-filling.

As mentioned before, when the transmitter has no CSI, the optimal transmit covariance matrix is proportional to the identity matrix. Space-time coding has been designed for this case to spread transmit power across multiple independently fading transmit/receive antenna paths [23, 72, 73]. Given the feedback of partial CSI in the forms of either mean feedback or covariance feedback, the optimal transmit covariance matrix is no longer white, but space-time coding can still be used. As shown by [36, 85, 86], each eigen-direction characterized by one singular vector of the optimal transmit covariance matrix can be

treated as one transmit antenna, and existing space-time coding can be applied over these directions.

2.2 Acquisition of Partial CSI at the Transmitter

When the wireless channel is not reciprocal, i.e., when the uplink and downlink channels cannot be inferred from each other, as is the case in FDD, the receiver needs to feedback the downlink channel to the transmitter. Because the feedback channel is a scarce resource, the receiver needs to find an efficient way to represent the channel. Basically, this is a quantization problem. The difference from traditional quantization problems is that the distortion measure here is not mean-square error but the capacity penalty resulting from using non-perfect CSI instead of perfect CSI at the transmitter.

We review the channel quantization schemes in the literature. Without loss of generality, we consider i.i.d. MIMO Rayleigh flat-fading channels.

2.2.1 Unstructured Quantization

The early methods for quantizing channels were heuristic and coarse, and mostly focused on one-dimensional beamforming. The design criterion was similar for both MIMO and MISO, and we use MISO as an example. Following channel model (4), the channel model here is given by

$$y = \boldsymbol{h}\boldsymbol{w}d + n, \tag{9}$$

where the unit-norm $\boldsymbol{w}$ is the beamforming vector to be designed, and d is the data symbol. When the transmitter has perfect CSI, the optimal (in terms of maximizing both mutual information and received SNR) beamforming vector is simply $\boldsymbol{h}^H/\|\boldsymbol{h}\|$, where $(\cdot)^H$ denotes Hermitian transpose. This solution is called maximum-ratio transmission, similar to the maximum-ratio combining in diversity receivers.

Now the problem becomes that of how to quantize the beamforming vector $\boldsymbol{w}$, or, equivalently, the channel vector $\boldsymbol{h}$. One example of the early techniques is transmit antenna selection [7, 29, 48], where the receiver informs the transmitter to use a subset of antennas that have larger gains than the remaining. A white covariance matrix is applied to the

selected subset of transmit antennas. Another example is quantized equal-gain transmission [42], which uniformly quantizes the phases of the channel vector without considering the amplitudes. Better quantization performance was achieved by Narula et al [54], using the quantized maximum ratio transmission, where the optimal one-dimensional beamforming vector was quantized. The beamforming codebook was obtained using the Lloyd algorithm and a specific codebook design methodology was not developed.

2.2.2 Packing in the Grassmann Manifold

A systematic design of quantized beamforming vectors was given by [44] and [53]. The problem solved here is to quantize the unit-norm beamforming vectors. Let the codebook of unit-norm beamforming vectors be $\widetilde{\boldsymbol{W}} = \{\tilde{\boldsymbol{w}}_1, \tilde{\boldsymbol{w}}_2, \ldots, \tilde{\boldsymbol{w}}_N\}$. The quantizer at the receiver $Q_{\widetilde{\boldsymbol{W}}} : \mathbb{C}^{1\times N_t} \rightarrow \{\tilde{\boldsymbol{w}}_1, \tilde{\boldsymbol{w}}_2, \ldots, \tilde{\boldsymbol{w}}_N\}$ is a function that selects the element of the codebook that maximizes the equivalent channel gain. Thus,

$$Q_{\widetilde{\boldsymbol{W}}}(\boldsymbol{h}) = \arg \max_{1\leq i\leq N} |\boldsymbol{h}\tilde{\boldsymbol{w}}_i|^2. \tag{10}$$

Since the distortion defined by the capacity penalty is hardly tractable, the measure of distortion in [44] was defined as

$$D(\widetilde{\boldsymbol{W}}) = E_{\boldsymbol{h}}\{\|\boldsymbol{h}\|^2 - |\boldsymbol{h}Q_{\widetilde{\boldsymbol{W}}}(\boldsymbol{h})|^2\}. \tag{11}$$

By minimizing an upper bound of $D(\widetilde{\boldsymbol{W}})$, the criterion of designing the codebook $\widetilde{\boldsymbol{W}}$ became maximization of

$$\delta(\widetilde{\boldsymbol{W}}) = \min_{1\leq k<l\leq N} \sqrt{1 - |\tilde{\boldsymbol{w}}_k^H \tilde{\boldsymbol{w}}_l|^2}. \tag{12}$$

The expression $\sqrt{1 - |\tilde{\boldsymbol{w}}_i^H \tilde{\boldsymbol{w}}_j|^2}$ is the chordal distance [11] between two one-dimensional subspaces represented by the vectors $\tilde{\boldsymbol{w}}_i$ and $\tilde{\boldsymbol{w}}_j$. The problem becomes one of finding the set, or packing, of N lines in $\mathbb{C}^{N_t}$ that has maximum minimum distance between any pair of lines. The mathematical name for such a problem is *Grassmannian line packing* [11]. The complex Grassmann manifold $\mathcal{G}_{m,n}$ is the set of all n-dimensional subspaces of the space $\mathbb{C}^m$. Therefore, the Grassmannian line packing problem is finding a set of N points in $\mathcal{G}_{N_t,1}$ that have maximum minimum distance.

Once the beamforming quantization problem is formulated as a Grassmannian packing problem, the good packings that have already been found can be used as beamforming quantization codebooks. The geometrical properties of the Grassmann manifold can be used to analyze the performance of quantized beamforming [50, 51].

The one-dimensional Grassmannian line packing has been extended to multi-dimensional beamforming for MIMO channels [43], where the channel model becomes

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n} = \boldsymbol{H}\boldsymbol{W}\boldsymbol{d} + \boldsymbol{n}. \tag{13}$$

In the above equation, $\boldsymbol{d}$ is a vector of coded message symbols, and $\boldsymbol{W}$ is the beamforming matrix that maps the data vector $\boldsymbol{d}$ to the transmitted signals $\boldsymbol{x}$. We assume that the covariance matrix of $\boldsymbol{d}$ is proportional to an identity matrix whose trace equals to the power constraint P. Let the SVD of $\boldsymbol{H}$ be $\boldsymbol{H} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^H$. Clearly, if the transmitter has perfect CSI, columns of $\boldsymbol{W}$ should be proportional to columns of $\boldsymbol{V}$ and possibly power-loaded. If the system designer prefers to use a beamforming matrix with rank N_s such that $N_s < N_t$, $\boldsymbol{W}$ should contain right singular vectors corresponding to the N_s largest singular values of $\boldsymbol{H}$. Therefore, the beamforming matrix $\boldsymbol{W}$ acts as valid CSI to quantize and feedback. This is exactly the problem considered in [43].

Let $\boldsymbol{W} = \begin{pmatrix} \boldsymbol{w}_1 & \boldsymbol{w}_2 & \dots & \boldsymbol{w}_{N_s} \end{pmatrix}$ be a quantized beamforming matrix for channel $\boldsymbol{H}$. The effective received power in each direction specified by a column $\boldsymbol{w}_k, k = 1, \dots, N_s$, is given by $\|\boldsymbol{H}\boldsymbol{w}_k\|^2$. Since $\boldsymbol{W}$ is quantized and not very accurate, $\|\boldsymbol{H}\boldsymbol{w}_k\|^2, k = 1, \dots, N_s$ might not be very different from each other, or their difference might not be as significant as the difference between singular values of $\boldsymbol{H}$. Therefore, the power loading on the column vectors of $\boldsymbol{W}$ can be ignored. This way, a valid codebook for quantizing $\boldsymbol{W}$ could be a set of $N_t \times N_s$ matrices with orthonormal columns. Also, the order of columns inside $\boldsymbol{W}$ is not important. In other words, each codeword represents a subspace spanned by its column vectors. It is the subspaces that we want to quantize, not the specific orthonormal basis (comprised of column vectors of a codeword) spanning these subspaces. According to [43], codebook design becomes the finding of a set of N points in the Grassmann manifold $\mathcal{G}_{N_t,N_s}$ such that the minimum chordal distance between any two codewords is minimized. This problem is called

Grassmannian subspace packing. Again, good constructions of Grassmannian subspace packing can be used as codebooks for quantizing the beamforming matrices.

2.2.3 Stochastic Gradient Methods

The Grassmannian line/subspace packing schemes quantize the channel without considering its history. However, the channel states of adjacent time instants are almost always correlated. Therefore, the knowledge of previous channel states at the transmitter provides some prior information for the current channel state. If the transmitter knows the previous channel states, then only the variation of the channel states needs to be fed back. Banister and Zeidler proposed a gradient sign feedback algorithm to track the variation of the channel's dominant right singular subspaces [4, 5]. Their algorithm is summarized in the following. Since the algorithm deals with channel variation with time, we introduce a discrete-time variable n in the channel model (13):

$$\boldsymbol{y}[n] = \boldsymbol{H}[n]\boldsymbol{x}[n] + \boldsymbol{n}[n] = \boldsymbol{H}[n]\boldsymbol{W}[n]\boldsymbol{d}[n] + \boldsymbol{n}[n], \tag{14}$$

where $\boldsymbol{W}[n]$ is the beamforming matrix used by the transmitter at time n.

The orthonormal precoding matrix $\boldsymbol{W}[n]$ that resides in the principal right singular subspace of $\boldsymbol{H}[n]$ can be obtained [4] by maximizing the cost function

$$J[n] = \|\boldsymbol{H}[n]\boldsymbol{W}[n]\|_F^2, \tag{15}$$

which is the received power at discrete time n, subject to the constraint that $\boldsymbol{W}^H[n]\boldsymbol{W}[n] = \boldsymbol{I}_{N_s}$. In this way, the feedback problem was posed as an optimization problem in [4].

The cost function maximization can be carried out adaptively. One of the most important methods used in the optimization literature is known as gradient descent, which involves adaptive convergence to a point in a vector space corresponding to the global minimum of a cost function defined on the space. At each iteration of the adaptation, an estimate of the gradient of the cost function is formed, and the estimate of the optimal vector is revised to move in the direction of the gradient vector. The most widely used gradient search technique is the stochastic gradient search. In the case under examination,

the adaptation has to take place at the transmitter. However, feedback of the entire gradient vector (or matrix) over the reverse link is not practical. The perturbation algorithm in [4] is a method of providing an approximation to the gradient of the cost function to the transmitter with minimal feedback. The approach is similar to simultaneous perturbation stochastic approximation approaches [68]. At each iteration, the effect of a stochastic perturbation on the cost function is studied, and, based on this effect, the estimate of the optimal vector is moved toward or away from the direction of the perturbation vector.

In [4], the current weight matrix $\boldsymbol{W}[n]$ is perturbed as

$$\begin{aligned} \boldsymbol{W}_+[n] &= \boldsymbol{W}[n] + \beta\Delta\boldsymbol{W}, \\ \boldsymbol{W}_-[n] &= \boldsymbol{W}[n] - \beta\Delta\boldsymbol{W}, \end{aligned} \tag{16}$$

where β is a step size, and $\Delta\boldsymbol{W}$ contains i.i.d. Gaussian entries, which, for example, can be generated by synchronized pseudo-random number generators at the transmitter and the receiver. The effect of perturbation is estimated as

$$s[n+1] = \|\boldsymbol{H}[n+1]\boldsymbol{W}_+[n]\|^2 - \|\boldsymbol{H}[n+1]\boldsymbol{W}_-[n]\|^2. \tag{17}$$

It can be shown that

$$E\{s[n+1]\Delta\boldsymbol{W}\} \propto \boldsymbol{H}^H[n+1]\boldsymbol{H}[n+1]\boldsymbol{W}[n+1], \tag{18}$$

where the right hand side is the gradient of the cost function (15). Furthermore, it is sufficient to have only the quantity $\text{sign}(s[n])$ instead of $s[n]$ in the left hand side of (18), and $\text{sign}(s[n])$ can be deemed as a highly quantized form of $s[n]$. The adaptation of the weight matrix is given by the iteration

$$\boldsymbol{W}[n+1] = G(\boldsymbol{W}[n] + s[n+1]\beta\Delta\boldsymbol{W}), \tag{19}$$

where $G(\cdot)$ is the Gram-Schmidt column orthonormalization of the input matrix.

Contrary to the stochastic perturbation in [4], a deterministic perturbation was proposed in [59]. This approach uses a set of predetermined vectors for the perturbation and cycles through them. Any orthonormal set of vectors that form a basis for $\mathbb{C}^{N_t}$ can be used to form a perturbation set. The expectation of the product of the sign-feedback and the perturbation

vectors was shown to also be proportional to the gradient vector. The simulation result in [59] showed that the deterministic approach actually has better initial convergence, but more variance after it reaches its final state. Also, the deterministic approach has the possibility of converging to a local minimum. Since the performance of stochastic and deterministic perturbation is similar, the stochastic perturbation in [4] serves as the major reference in the following chapter.

2.3 Grassmann Manifolds

In this section we summarize some properties of the geometry of Grassmann manifolds used in this thesis [19]. For details concerning Grassmann manifolds, the readers are referred to standard texts such as [9]. A complex Grassmann manifold $\mathcal{G}_{N_t,N_s}$ contains all N_s-dimensional subspaces of $\mathbb{C}^{N_t}$. As mentioned previously, since the transmit beamforming matrix can be any orthonormal $N_t \times N_s$ matrix that forms an orthonormal basis for the principal right singular subspace of the MIMO channel, the Grassmann manifold is a natural description of the domain of the transmit beamforming matrices.

We introduce all necessary notations using time instant 0 as an example. The $N_t \times N_s$ orthonormal transmit weight matrix at time 0 is $\boldsymbol{W}[0]$. Denoted as $\langle \boldsymbol{W}[0] \rangle$, a point in the Grassmann manifold is an equivalent class

$$\langle \boldsymbol{W}[0] \rangle = \{ \boldsymbol{W}[0]\boldsymbol{Q}_{N_s} : \boldsymbol{Q}_{N_s} \text{ is any } N_s \times N_s \text{ unitary matrix} \}, \tag{20}$$

i.e., a point in the Grassmann manifold is the set of all $N_t \times N_s$ orthonormal matrices whose columns span the same subspace as spanned by the columns of $\boldsymbol{W}[0]$. $\boldsymbol{W}[0]$ is an arbitrary basis of $\langle \boldsymbol{W}[0] \rangle$. We sometimes refer to $\boldsymbol{W}[0]$ as a subspace or a point of the Grassmann manifold, in the sense that $\boldsymbol{W}[0]$ is a basis of such a point. Let $\boldsymbol{Q}[0] = (\boldsymbol{W}[0] \,|\, \boldsymbol{Z}[0])$, where $\boldsymbol{Z}[0]$ contains as columns an orthonormal basis of the orthogonal complement of $\langle \boldsymbol{W}[0] \rangle$. In other words, $\boldsymbol{Q}[0]$ is a square unitary matrix.

A geodesic is the curve of shortest length between two points on a manifold. We will model the stochastic process in the Grassmann manifold starting at $\langle \boldsymbol{W}[n-1] \rangle$ and ending at $\langle \boldsymbol{W}[n] \rangle$ induced by channel variation as along a geodesic. Therefore, we need an

explicit description of the geodesics in $\mathcal{G}_{N_t,N_s}$. Geodesics in $\mathcal{G}_{N_t,N_s}$ starting from $\boldsymbol{W}[0]$ are parameterized by [19]

$$\boldsymbol{W}(t) = \boldsymbol{Q}[0] \exp(t\boldsymbol{B}[0])\boldsymbol{I}_{N_t,N_s}, \tag{21}$$

where $t \in \mathbb{R}$ is the time variable. It can be verified that $\boldsymbol{W}(0) = \boldsymbol{W}[0]$. As t varies along the real axis, (21) parameterizes the basis of subspaces along a geodesic curve in the Grassmann manifold that passes $\langle \boldsymbol{W}[0] \rangle$ at time $t = 0$. The skew Hermitian matrix $\boldsymbol{B}[0]$ is further restricted to be of the form

$$\boldsymbol{B}[0] = \begin{pmatrix} \mathbf{0} & -\boldsymbol{A}^H[0] \\ \boldsymbol{A}[0] & \mathbf{0} \end{pmatrix}, \boldsymbol{A}[0] \in \mathbb{C}^{(N_t-N_s)\times N_s}. \tag{22}$$

We denote the point reached by the geodesic at time $t = 1$ as $\boldsymbol{W}[1]$, therefore,

$$\boldsymbol{W}[1] = \boldsymbol{W}(1) = \boldsymbol{Q}[0] \exp(\boldsymbol{B}[0])\boldsymbol{I}_{N_t,N_s}. \tag{23}$$

The complex dimensionality of the Grassmann manifold $\mathcal{G}_{N_t,N_s}$ is $(N_t - N_s)N_s$ [9]. Given $\langle \boldsymbol{W}[0] \rangle$, the degrees of freedom of $\langle \boldsymbol{W}[1] \rangle$ are embodied by the matrix $\boldsymbol{A}[0]$ in the lower-left corner of $\boldsymbol{B}[0]$. The matrix $\boldsymbol{A}[0]$ determines the point $\boldsymbol{W}[1]$ uniquely given $\boldsymbol{W}[0]$. Recall that for a point x on the unit circle in Euclidean space $\mathbb{C}^1$, xe^{jwt} is the parametrization of a piece of arc of the unit circle, and w is called the angular velocity that brings x to $y = xe^{jw}$ in unit time. Recognizing the similarity of this parametrization to (21) and noticing the redundant structure of $\boldsymbol{B}[0]$ in (22), we define $\boldsymbol{A}[0]$ as the velocity that takes $\boldsymbol{W}[0]$ to $\boldsymbol{W}[1]$ in unit time. In the MIMO subspace tracking context, $\langle \boldsymbol{W}[0] \rangle$ is the outdated knowledge of the transmit subspace at the transmitter, and $\langle \boldsymbol{W}[1] \rangle$ is the current transmit subspace that we want to be as close as possible to the principal right singular subspace of the current channel matrix. The essence of the proposed algorithms in the next chapter is to quantize $\boldsymbol{A}[0]$ instead of $\boldsymbol{W}[1]$ itself.

Here, we summarize the interrelationship among $\boldsymbol{W}[0]$, $\boldsymbol{W}[1]$, $\boldsymbol{A}[0]$, and $\boldsymbol{B}[0]$ [22]. Let $0 \leq \theta_1 \leq \ldots \leq \theta_{N_s} \leq \pi/2$ be the principal angles [28] between the subspaces $\langle \boldsymbol{W}[0] \rangle$ and $\langle \boldsymbol{W}[1] \rangle$. Let $\boldsymbol{U}_1\boldsymbol{C}\boldsymbol{V}_1^H$ be an SVD of the $N_s \times N_s$ matrix $\boldsymbol{W}[0]^H\boldsymbol{W}[1]$, where $\boldsymbol{C}$ is a

diagonal matrix with elements $\cos\theta_k$, $1 \le k \le N_s$, on the diagonal. The following Cosine-Sine Decomposition [28] defines a key relationship:

$$\begin{pmatrix} \boldsymbol{W}^H[0]\boldsymbol{W}[1] \\ \boldsymbol{Z}^H[0]\boldsymbol{W}[1] \end{pmatrix} = \underbrace{\begin{pmatrix} \boldsymbol{U}_1 & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{U}_2 & \boldsymbol{U}_3 \end{pmatrix}}_{\boldsymbol{U}} \begin{pmatrix} \boldsymbol{C} \\ \boldsymbol{S} \\ \boldsymbol{0} \end{pmatrix} \boldsymbol{V}_1^H, \tag{24}$$

where $\boldsymbol{S}$ is a diagonal matrix with elements $\sin\theta_k$ on the diagonal, $\widetilde{\boldsymbol{U}}_2 = [\boldsymbol{U}_2 \,|\, \boldsymbol{U}_3]$ is an $(N_t - N_s) \times (N_t - N_s)$ unitary matrix, and $\boldsymbol{U}_2$ has N_s columns. It turns out that one possible choice for the velocity matrix is

$$\boldsymbol{A}[0] = \boldsymbol{U}_2 \boldsymbol{\Theta} \boldsymbol{U}_1^H, \tag{25}$$

where $\boldsymbol{\Theta} = \mathrm{diag}(\theta_1, \theta_2, \ldots, \theta_{N_s})$, and (25) happens to be an SVD of $\boldsymbol{A}[0]$. Denoting $\boldsymbol{U} = \mathrm{diag}(\boldsymbol{U}_1, \widetilde{\boldsymbol{U}}_2)$, it can be shown that

$$\exp(\boldsymbol{B}[0]) = \boldsymbol{U}\boldsymbol{R}\boldsymbol{U}^H, \tag{26}$$

where

$$\boldsymbol{R} = \begin{pmatrix} \boldsymbol{C} & -\boldsymbol{S} & \boldsymbol{0} \\ \boldsymbol{S} & \boldsymbol{C} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{N_t - 2N_s} \end{pmatrix}. \tag{27}$$

Given $\boldsymbol{W}[0]$ and $\boldsymbol{A}[0]$, an SVD of $\boldsymbol{A}[0]$ can be performed, and then $\boldsymbol{W}[1]$ is easily computed using (26) without performing matrix exponentials.

Having reviewed the state-of-the-art of partial CSI feedback techniques and optimal transmitter designs given channel statistics, we are now ready to present our contributions in this area, starting in the next chapter. We will propose a new partial CSI feedback scheme in Chapter 3 that utilizes both time-domain correlation of channel states and intrinsic properties of variation between subspaces, as discussed in this section.

CHAPTER III

TRANSMISSION SUBSPACE TRACKING FOR MIMO SYSTEMS WITH LOW RATE FEEDBACK

3.1 Introduction

Consider a MIMO system with N_t transmit antennas and N_r receive antennas where the receiver tries to quantize an N_s-dimensional principal right singular subspace and feed it back to the transmitter. When $N_t \geq N_s \geq N_r$, Banister and Zeidler [4] showed that transmission within N_s-dimensional principal subspace of the channel matrix has a power gain of precisely N_t/N_s over transmission without CSI. They also showed that the performance penalty of equal power allocation among dimensions within the principal subspace relative to water filling is minor, especially at high *signal-to-noise ratios* (SNRs). Therefore, we focus on quantization of the principal subspace and equal power allocation in this chapter.

The Grassmannian subspace packing scheme in [43] quantizes the channel without considering its history. However, the channel states of adjacent time instants are almost always correlated. Therefore, the knowledge of previous channel states at the transmitter provides some prior information for the current channel state. If the transmitter knows the previous channel states, then only the variation of the channel states needs to be fed back. Quantization schemes proposed in [4, 49, 61] fit into this context. In [49], Mondal et al. showed an adaptive quantization scheme that adapts to a time-varying distribution of the channel. In [61], the orthonormal precoding matrix is parameterized using a series of Givens rotations, and scalar quantization based on adaptive Delta modulation is used for each parameter. A gradient sign feedback algorithm for tracking the dominant transmit subspaces in MIMO systems was proposed in [4]. The scheme in [4] utilizes the correlation between adjacent channel states and tracks the channel variation. However, it does not exploit any intrinsic property of subspace variation.

Here, a new partial CSI acquisition algorithm is proposed. Similar to [43], we consider the transmit subspaces as points in a complex Grassmann manifold. The difference from [43] is that we quantize the variation between subspaces. Specifically, based on a recent model proposed for conventional subspace tracking [69], we model the variations between the dominant subspaces of channels at adjacent time instants to be along geodesics in the Grassmann manifold. Instead of quantizing the subspaces themselves, we propose to quantize the geodesic trajectory connecting two subspaces. More specifically, we quantize a key entity that characterizes a geodesic arc: the velocity matrix, which can be considered to be a generalization of angular speed in a one-dimensional complex space. A one-bit feedback scheme, essentially an approximate stochastic gradient algorithm, is utilized to indicate the preferred sign of a randomly generated velocity matrix of the geodesic. The number of entries in the velocity matrix is smaller than the transmit precoding matrix itself, resulting in reduced variance in the gradient estimate and faster convergence compared to [4]. The expected direction of this geodesic at the starting point is approximately proportional to the gradient of the same cost function as in [4] but defined on the Grassmann manifold. A VQ scheme that quantizes the velocity matrix using a Gaussian codebook is also proposed. The performance of the Gaussian VQ is better than that of Grassmannian subspace packing for a fairly large range of Doppler frequencies. Compared to [43] and [4], our algorithms utilize time-domain channel correlation and intrinsic properties of the variation of subspaces. Compared to [61], our schemes have a VQ nature and quantize a smaller number of parameters ($(N_t - N_s)N_s$ vs $(2N_t - 1 - N_s)N_s$).

The difference between this work and [49] is more subtle. In the language of vector quantization literature, the schemes proposed in this chapter fall into the category of predictive VQ. Here the current CSI serves as a prediction of CSI for the next time instant, and the difference is quantized out of a fixed codebook. A vector quantizer is *adaptive* if the codebook or the encoding rule is changed in time in order to match observed local statistics of the input sequence [24], which is the case for the switched codebook VQ in [49]. The word "adaptive" is usually associated with systems where the codebook changes slowly with respect to the vector rate rather than changing substantially with each successive vector.

Therefore, predictive VQ is not categorized in the research literature as being adaptive [24]. Nonetheless, we do expect similar performance between this work and [49] in certain circumstances.

We emphasize that the problem at hand has little to do with the conventional subspace tracking literature. The purpose of conventional subspace tracking is to obtain good estimates of principal subspaces from noisy observations. Our purpose is to approximate a MIMO channel with its principal right singular subspace and to reduce the rate of information that is fed back by quantizing the variation between subspaces. Here, we assume that the channel estimation, and hence the subspace estimation, at the receiver is almost perfect. Tracking here is used to convey the variation of a subspace stochastic process to the transmitter. In other words, the nature of the subspace tracking problem in [69] is estimation, while the nature of our problem is quantization. All that we take from the subspace tracking literature is a model for the subspace stochastic process recently adopted in [69].

The outline of this chapter is as follows. The problem setting is given in Section 3.2. Assumptions taken in this chapter are listed in Section 3.3. The one-bit feedback algorithm is demonstrated in Section 3.4 [80], and its convergence analysis is provided in Section 3.5. The VQ scheme is described in Section 3.6. Numerical examples are shown in Section 3.7, and Section 3.8 concludes this chapter.

Notation: Bold upper (lower) letters denote matrices (column vectors).

3.2 Problem Setting

3.2.1 Partial CSI

We consider a flat-fading MIMO channel with N_r receive antennas and N_t transmit antennas characterized by the discrete-time input-output relationship

$$\boldsymbol{y}[n] = \boldsymbol{H}[n]\boldsymbol{x}[n] + \boldsymbol{n}[n], \tag{28}$$

where $\boldsymbol{H}[n]$ is an $N_r \times N_t$ complex channel transfer matrix and $\boldsymbol{n}[n]$ is an $N_r \times 1$ zero-mean complex Gaussian noise vector with covariance matrix $N_0\boldsymbol{I}$. The SVD of $\boldsymbol{H}[n]$ is defined as

$$\boldsymbol{H}[n] = \boldsymbol{U}[n]\boldsymbol{\Lambda}[n]\boldsymbol{V}^H[n], \tag{29}$$

where $\boldsymbol{\Lambda}[n]$ has on its main diagonal the singular values of $\boldsymbol{H}[n]$ in descending order. The N_s-dimensional principal right singular subspace of $\boldsymbol{H}[n]$ is spanned by the first N_s columns of $\boldsymbol{V}[n]$, which we denote as $\widetilde{\boldsymbol{V}}[n]$. Note that $\widetilde{\boldsymbol{V}}[n]$ is an orthonormal matrix, i.e., $\widetilde{\boldsymbol{V}}^H[n]\widetilde{\boldsymbol{V}}[n] = \boldsymbol{I}_{N_s}$.

Given perfect knowledge of $\boldsymbol{H}[n]$ at the transmitter, the capacity maximizing transmission is along the directions specified by the column vectors of the matrix $\boldsymbol{V}[n]$. However, perfect knowledge of $\boldsymbol{H}[n]$ is usually costly to obtain. It has been discussed in [4, 43, 61] that feeding back the principal subspace spanned by the first several columns of $\boldsymbol{V}[n]$ is a good tradeoff between performance and feedback bandwidth requirement. We also take this approach. When there exists considerable correlation between channel matrices at adjacent time instants, it is conceivable that the principal right singular subspaces of these channel matrices also have time-domain correlation. Conditioned on the principal channel subspaces of previous time instants, the subspace of the current time instant has a non-uniform probability mass. This correlation can be utilized to reduce the amount of feedback. The optimal quantization codebook will no longer be uniform, and codewords will be denser in regions where the probability mass of the subspaces is high.

The objective of the proposed algorithm is to track an $N_t \times N_s$ orthonormal complex weight matrix $\boldsymbol{W}[n]$ that maps an $N_s \times 1$ complex vector $\boldsymbol{d}$ of coded message symbols to the transmitted signals

$$\boldsymbol{x}[n] = \boldsymbol{W}[n]\boldsymbol{d}[n]. \tag{30}$$

The tracking attempts to extract the principal right singular subspace, giving

$$\boldsymbol{W}[n]\boldsymbol{W}^H[n] = \widetilde{\boldsymbol{V}}[n]\widetilde{\boldsymbol{V}}^H[n]. \tag{31}$$

As discussed in [4], when $N_t \geq N_r = N_s$, an orthonormal precoding matrix $\boldsymbol{W}[n]$ satisfying (31) is asymptotically capacity maximizing at high SNR. It has also been shown by Scaglione

et al. [62] that the precoding matrix $\boldsymbol{W}[n]$ as in (31) minimizes the mean-squared error of the symbol estimate for a linear receiver subject to a maximum singular value constraint on the precoder. For both of the above two optimality criteria, an optimal $\boldsymbol{W}[n]$ remains optimal if it is replaced by $\boldsymbol{W}[n]\boldsymbol{U}$ where $\boldsymbol{U}$ is an arbitrary $N_s \times N_s$ unitary matrix. In other words, the optimal linear precoding matrix could be any orthonormal matrix whose columns span the same subspace as the columns of $\widetilde{\boldsymbol{V}}[n]$ do. Clearly, the subspace spanned by the column vectors of $\widetilde{\boldsymbol{V}}[n]$ should be fed back, instead of the orthonormal matrix $\widetilde{\boldsymbol{V}}[n]$ itself. The mathematical modeling of subspaces can be expressed in terms of the Grassmann manifold. Therefore, we are now faced with a problem of quantization of a stochastic process in the Grassmann manifold.

Note that if $N_s > N_t/2$, we only need to track the orthogonal complement of $\widetilde{\boldsymbol{V}}[n]$. Therefore, without loss of generality, we assume that $N_s \leq N_t/2$.

3.3 Assumptions

Before elaboration of the algorithm, we list our assumptions. First of all, we make assumptions on the joint distribution of the entries of the velocity matrix $\boldsymbol{A}[n]$. Exact modeling of the statistical properties of $\boldsymbol{A}[n]$ is difficult, because $\boldsymbol{A}[n]$, the velocity matrix that takes the last quantized subspace to the new channel principal subspace in unit time, depends not only on the stochastic process formed by $\boldsymbol{H}[n]$, but also on quantization errors on subspaces of previous time instants. In this chapter, we assume that the entries of the velocity matrix $\boldsymbol{A}[n]$ are i.i.d. $\mathcal{CN}(0, a^2)$, where a is a parameter depending on the Doppler frequency. Conceptually, this corresponds to a first-order Markov model on the process formed by the principal right singular subspaces $\boldsymbol{W}[n]$ [75]. Other models are possible depending on the application. For example, in the context of subspace tracking, Srivastava and Klassen [69] assume a first-order Markov model on the velocity process and, therefore, a second-order Markov model on the subspace process.

In situations where the feedback channel is of a very low rate, the batch of bits describing the forward channel takes some time to get through the feedback channel. Consequently, the receivers usually need channel prediction, and it is the predicted channel that is quantized.

We assume that the receiver channel estimation error manifests itself as additive white Gaussian noise added to the true channel matrix. Denoting the channel estimate at time n as $\widetilde{\boldsymbol{H}}[n]$, we have

$$\widetilde{\boldsymbol{H}}[n] = \boldsymbol{H}[n] + \boldsymbol{n}_1[n], \tag{32}$$

where $\boldsymbol{n}_1[n]$ denotes the estimation error with i.i.d. $\mathcal{CN}(0, N_1)$ entries.

Future channel values can be predicted with a linear predictor. If the fading processes for the entries of $\boldsymbol{H}[n]$ are i.i.d., prediction for a future channel value between a transmitter-receiver antenna pair can be done without regard to other antenna pairs. Let N_p be the length of the predictor filter, and $\boldsymbol{p}^{(ij)} = [p_0^{(ij)}, p_1^{(ij)}, \ldots, p_{N_p-1}^{(ij)}]^T$ be the coefficients of the predictor for the channel between transmitter antenna j and receiver antenna i. Denote the predicted channel matrix as $\widehat{\boldsymbol{H}}[n]$. Denote the entries on the ith row and jth column of matrices $\boldsymbol{H}[n]$, $\widetilde{\boldsymbol{H}}[n]$, and $\widehat{\boldsymbol{H}}[n]$ as $h_{ij}[n]$, $\tilde{h}_{ij}[n]$, and $\hat{h}_{ij}[n]$, respectively. Suppose that at time n the receiver predicts the MIMO channel values Q time instants ahead. With the notation $\tilde{\boldsymbol{h}}_{ij}[n] = [\tilde{h}_{ij}[n], \tilde{h}_{ij}[n-1], \ldots, \tilde{h}_{ij}[n-N_p+1]]^T$, the channel coefficient at time $n+Q$ can be predicted as

$$\hat{h}_{ij}[n+Q] = \sum_{p=0}^{N_p-1} p_p^{(ij)*} \tilde{h}_{ij}[n-p] = \boldsymbol{p}^{(ij)H} \tilde{\boldsymbol{h}}_{ij}[n]. \tag{33}$$

The channel predictor that minimizes the mean-squared error $E\{|\hat{h}_{ij}[n+Q] - h_{ij}[n+Q]|^2\}$ is given by [37]

$$\begin{aligned} \boldsymbol{p}^{(ij)} &= \left(E\left\{\tilde{\boldsymbol{h}}_{ij}[n]\tilde{\boldsymbol{h}}_{ij}^H[n]\right\}\right)^{-1} E\left\{h_{ij}^*[n+Q]\tilde{\boldsymbol{h}}_{ij}[n]\right\} \\ &= \boldsymbol{R}^{-1}\boldsymbol{r}, \end{aligned} \tag{34}$$

where $\boldsymbol{R} = E\left\{\tilde{\boldsymbol{h}}_{ij}[n]\tilde{\boldsymbol{h}}_{ij}^H[n]\right\}$ is the covariance matrix of $\tilde{\boldsymbol{h}}_{ij}[n]$ and $\boldsymbol{r} = E\left\{h_{ij}^*[n+Q]\tilde{\boldsymbol{h}}_{ij}[n]\right\}$ is the correlation between $h_{ij}^*[n+Q]$ and $\tilde{\boldsymbol{h}}_{ij}[n]$. Assuming the channel is time-varying according to Jakes' model with Doppler frequency F_D, it is well-known that $E\{h_{ij}^*[n+m]h_{ij}[n]\} = J_0(2\pi F_D|m|T_s)$, where $J_0(\cdot)$ is the zeroth order Bessel function of the first kind and T_s is the symbol period. Thus, we have

$$[\boldsymbol{R}]_{p,q} = J_0(2\pi|p-q|F_D T_s) + N_1\delta(p-q), \tag{35}$$

$$[\boldsymbol{r}]_p = J_0(2\pi|Q+p|F_D Ts), \tag{36}$$

where $\delta(\cdot)$ is the Kronecker delta function and the subscripts, p and q, start from 0. Finally, we can see that the predictor coefficients $\boldsymbol{p}^{(ij)}$ are the same for any i and j.

For the proposed one-bit feedback algorithm, we assume that channel estimation at the receiver, feedback, and computation of the transmit weight matrix at the transmitter do not consume time and occur instantly. The frequency of these operations is the same as the feedback bit rate. Similarly, for the Grassmannian subspace packing scheme or Gaussian VQ scheme with a codebook of size 2^N, we assume that the channel at time $n + \frac{3}{2}(N-1)$ (which is the midpoint of the interval in which this codeword will be used as the transmit beamforming matrix) is predicted by the receiver at time n, and the N-bit codeword is fed back at times n, $n+1$, $\ldots$, $n+N-1$. The dequantized beamforming matrix for time $n-1$ is held constant at the transmitter for times $n-1$, n, $\ldots$, $n+N-2$. For the purposes of quantization and precoding, we assume that channel prediction at the receiver is perfect, and no optimization of the quantization and precoding schemes is done with regard to estimation or prediction errors.

3.4 One-Bit Feedback Algorithm

For each feedback period, the random velocity $\boldsymbol{A}[n]$ is generated with i.i.d. $\mathcal{CN}(0, a^2)$ entries. We assume that the transmitter and the receiver have common knowledge of the value of $\boldsymbol{A}[n]$ at any time instant. $\boldsymbol{A}[n]$ can be conveyed from the transmitter to the receiver multiplexed with data symbols [4], or synchronously generated by identical pseudo-random number generators at the transmitter and the receiver.

With each instance of the random matrix $\boldsymbol{A}[n]$, we approximate the new transmit subspace as one point in the Grassmann manifold reached by a geodesic in unit time, starting from the current transmit subspace and using $\boldsymbol{A}[n]$ as the velocity matrix. Using the parameterized geodesic in (21), with one bit of ambiguity, the new transmit subspace $\boldsymbol{W}[n+1]$ can be either $\boldsymbol{Q}[n]\exp(\boldsymbol{B}[n])\boldsymbol{I}_{N_t,N_s}$ or $\boldsymbol{Q}[n]\exp(-\boldsymbol{B}[n])\boldsymbol{I}_{N_t,N_s}$. The binary feedback s is determined as

$$s[n+1] = \text{sign}(\|\boldsymbol{H}[n+1]\boldsymbol{Q}[n]\exp(\boldsymbol{B}[n])\boldsymbol{I}_{N_t,N_s}\|_F^2 - \|\boldsymbol{H}[n+1]\boldsymbol{Q}[n]\exp(-\boldsymbol{B}[n])\boldsymbol{I}_{N_t,N_s}\|_F^2). \tag{37}$$

The decision defined by (37) is binary encoded and provided as feedback from the receiver to the transmitter. This way, the feedback decision selects which sign-direction is preferable in terms of maximizing received power, as the result of an advancement along the geodesic.

The weight matrix update at the transmitter is given by

$$\boldsymbol{W}[n+1] = \boldsymbol{Q}[n] \exp(s[n+1]\boldsymbol{B}[n])\boldsymbol{I}_{N_t,N_s}. \tag{38}$$

Since $\boldsymbol{W}[n+1]$ is orthonormal, it serves as the new transmit matrix directly and no further orthonormalization is necessary, as opposed to the algorithm in [4], where a Gram-Schmidt QR factorization is required. The tracking algorithm is summarized in Table 1.

The parameter a controls the average length of the geodesic arc. To see this, first introduce a matrix $\boldsymbol{A}_w[0]$ with i.i.d. $\mathcal{CN}(0,1)$ entries such that $\boldsymbol{A}[0] = a\boldsymbol{A}_w[0]$, for $a > 0$. The arc length distance between $\boldsymbol{W}[0]$ and $\boldsymbol{W}[1]$ is derived from the intrinsic geometry of the Grassmann manifold and defined in terms of the principal angles between $\langle \boldsymbol{W}[0] \rangle$ and $\langle \boldsymbol{W}[1] \rangle$ as in [19]

$$d(\boldsymbol{W}[0], \boldsymbol{W}[1]) = \left(\sum_{i=1}^{N_s} \theta_i^2 \right)^{1/2}. \tag{39}$$

Note that

$$\|\boldsymbol{A}[0]\|_F^2 = \mathrm{tr}(\boldsymbol{A}[0]\boldsymbol{A}^H[0]) = \mathrm{tr}(\boldsymbol{U}_2\boldsymbol{\Theta}\boldsymbol{U}_1^H\boldsymbol{U}_1\boldsymbol{\Theta}^H\boldsymbol{U}_2^H) = \mathrm{tr}(\boldsymbol{\Theta}\boldsymbol{\Theta}^H) = \sum_{i=1}^{N_s} \theta_i^2. \tag{40}$$

Therefore, the arc length $d(\boldsymbol{W}[0], \boldsymbol{W}[1]) = \|\boldsymbol{A}[0]\|_F = a\|\boldsymbol{A}_w[0]\|_F$, i.e., the parameter a is proportional to the average arc length of the geodesic. Intuitively, the arc length of the geodesic traversed in unit time for high Doppler frequencies is larger than for low Doppler frequencies. Therefore, the parameter a should be chosen monotonically with Doppler frequency.

3.5 Gradient Extraction

In this section, we analyze the convergence behavior of this algorithm assuming that $\boldsymbol{H}$ is static and non-random, and, thus, we eliminate the discrete-time index from the notation of $\boldsymbol{H}$. We will show that the expectation of the direction of the geodesic is approximately proportional to the gradient of the cost function J defined in (15) when the norm of $\boldsymbol{A}[0]$ is small, which is the case when $\langle \boldsymbol{W}[0] \rangle$ is not far from $\langle \widetilde{\boldsymbol{V}} \rangle$.

Table 1: Tracking algorithm summary

Initialize:	
$\boldsymbol{W}[0] = \begin{pmatrix} \boldsymbol{I}_{N_s} \\ \boldsymbol{0} \end{pmatrix}$	(transmitter and receiver)
for $n = 0 : \infty$	
$\boldsymbol{A}[n]$=common realization of Gaussian random matrix	(transmitter and receiver)
Compute $\boldsymbol{Z}[n]$ such that $\boldsymbol{Q}[n] = (\boldsymbol{W}[n]\|\boldsymbol{Z}[n])$ is unitary	(transmitter and receiver)
Compute $s[n+1]$ using (37)	(receiver)
$\boldsymbol{W}[n+1] = \boldsymbol{Q}[n] \exp(s[n+1]\boldsymbol{B}[n])\boldsymbol{I}_{N_t,N_s}$	(transmitter and receiver)

We first simplify the expression of s in (37). When the norm of $\boldsymbol{A}[0]$ is small, the values of the θ_ks are small, too. In this case, a first-order approximation gives

$$\boldsymbol{C} \approx \boldsymbol{I} \text{ and } \boldsymbol{S} \approx \boldsymbol{\Theta}. \tag{41}$$

With this approximation, it is shown in Appendix A.1 that

$$s[1] \approx \operatorname{sign} \operatorname{tr} \Re(\boldsymbol{W}^H[0]\boldsymbol{H}^H\boldsymbol{H}\boldsymbol{Z}[0]\boldsymbol{A}[0]). \tag{42}$$

We now show that the expectation of the direction of this random geodesic is the same as the gradient of the cost function J. Differentiation of $\boldsymbol{W}(t)$ in (21), deeming t as the only variable, and with the sign feedback s explicitly written, gives

$$\dot{\boldsymbol{W}}(t) = s[1]\boldsymbol{Q}[0] \exp(t\boldsymbol{B}[0])\boldsymbol{B}[0]\boldsymbol{I}_{N_t,N_s}. \tag{43}$$

The chosen direction for one bit feedback at time $t = 0$ is equal to

$$\dot{\boldsymbol{W}}(0) = s[1]\boldsymbol{Q}[0]\boldsymbol{B}[0]\boldsymbol{I}_{N_t,N_s} = s[1]\boldsymbol{Z}[0]\boldsymbol{A}[0]. \tag{44}$$

The expectation of this direction is

$$E\{\dot{\boldsymbol{W}}(0)\} = \boldsymbol{Z}[0]E\{s[1]\boldsymbol{A}[0]\}. \tag{45}$$

It can be shown that

$$E\{\dot{\boldsymbol{W}}(0)\} = a\frac{1}{\sqrt{\pi}}\frac{\boldsymbol{Z}[0]\boldsymbol{Z}^H[0]\boldsymbol{H}^H\boldsymbol{H}\boldsymbol{W}[0]}{\|\boldsymbol{Z}^H[0]\boldsymbol{H}^H\boldsymbol{H}\boldsymbol{W}[0]\|_F}, \tag{46}$$

as proved in Appendix A.2. At the point $\boldsymbol{W}[0]$, the gradient of the cost function J defined on the Grassmann manifold is given by [19]

$$(\boldsymbol{I} - \boldsymbol{W}[0]\boldsymbol{W}^H[0])J_{\boldsymbol{W}[0]}, \tag{47}$$

where $\left[J_{\boldsymbol{W}[0]}\right]_{ij} = \frac{\partial J[0]}{\partial[\boldsymbol{W}[0]]_{ij}}$. With $J[0]$ defined in (15), $J_{\boldsymbol{W}[0]} = 2\boldsymbol{H}^H\boldsymbol{H}\boldsymbol{W}[0]$. Noticing that $\boldsymbol{I} - \boldsymbol{W}[0]\boldsymbol{W}^H[0] = \boldsymbol{Z}[0]\boldsymbol{Z}^H[0]$, we have that the gradient of J at point $\boldsymbol{W}[0]$ is equal to

$$2\boldsymbol{Z}[0]\boldsymbol{Z}^H[0]\boldsymbol{H}^H\boldsymbol{H}\boldsymbol{W}[0]. \tag{48}$$

Comparing with (46), we see immediately that the expected direction of the geodesic is proportional to the gradient. Among all tangential directions, this gradient points in the direction of maximum increase of the cost function defined on the Grassmann manifold. In contrast, the expected direction of the subspace updating in [4] is proportional to the gradient in the Euclidean sense.

To consider the variance of the error in the gradient estimate, we define the error matrix $\boldsymbol{E} = \dot{\boldsymbol{W}}(0) - E\{\dot{\boldsymbol{W}}(0)\}$. Under the same approximation in (41) that is valid when the norm of $\boldsymbol{A}[0]$ is small, it follows straightforwardly that

$$E\{\boldsymbol{E}^H\boldsymbol{E}\} = a^2\Big((N_t - N_s)\boldsymbol{I} - \frac{1}{\pi}\frac{\boldsymbol{W}_0^H\boldsymbol{H}^H\boldsymbol{H}\boldsymbol{Z}_0\boldsymbol{Z}_0^H\boldsymbol{H}^H\boldsymbol{H}\boldsymbol{W}_0}{\|\boldsymbol{Z}_0^H\boldsymbol{H}^H\boldsymbol{H}\boldsymbol{W}_0\|_F^2}\Big). \tag{49}$$

One measurement of the estimation error is related to the second moment of the error matrix through

$$\mathrm{tr}(E\{\boldsymbol{E}^H\boldsymbol{E}\}) = a^2\left((N_t - N_s)N_s - 1/\pi\right). \tag{50}$$

While for the subspace updating of [4], the estimation error embodied by a similarly defined error matrix $\boldsymbol{E}$ [4, Eq. (60)] is given by

$$\mathrm{tr}(E\{\boldsymbol{E}^H\boldsymbol{E}\}) = 2\beta^2\left(N_tN_s - 1/\pi\right), \tag{51}$$

where β is an adaptation parameter controlling the step size. Note that the term $(N_t - N_s)N_s$ in (50) is the number of degrees of freedom of $\boldsymbol{A}[0]$, while the term N_tN_s in (51) is the number of degrees of freedom of the random perturbation in [4]. It can be seen that the degrees of freedom do manifest themselves in the variance of the gradient estimate. Simulation results will show the advantage of our one-bit algorithm over [4].

It should be pointed out that the exact expectation of the direction of the geodesic is usually not proportional to the gradient of the cost function J. This discrepancy is

prominent when entries of $\boldsymbol{A}[0]$ are large enough so that the approximation (41) is no longer valid.

To converge to the principal subspace $\widetilde{\boldsymbol{V}}$ of a static matrix, one might expect a fully-adaptive algorithm to decrease the amplitude of the entries of the velocity matrix as the solution approaches $\widetilde{\boldsymbol{V}}$. We acknowledge that more theoretical analysis is required to fully characterize the convergence behavior of the one-bit algorithm starting from an arbitrary point in the Grassmann manifold. However, given the better performance of the quantization scheme presented in the next section, such effort will be deferred to future research.

3.6 Gaussian VQ

The one-bit quantization scheme updates the transmission weight subspace in a stochastic gradient manner. Our primary purpose in introducing that technique was to show the performance improvement achieved by doing perturbation to the velocity matrices instead of the weight matrices themselves. The performance of the one-bit quantization algorithm is limited by its sequential nature. A more general approach would be to quantize the velocity matrix with more than one bit at a time. In this section, we use such a vector quantization scheme and deterministically quantize the velocity matrix [79]. This VQ scheme will be shown to have much better performance than the one-bit quantization scheme.

Because of the i.i.d. Gaussian assumption on the velocity matrices, the problem has now become one of quantization of i.i.d. Gaussian sources. The variety of VQ techniques is quite large [24, 30]. When designing a quantization scheme for a specific source, considerations include complexity and performance in terms of closeness to the rate/distortion bound [13]. Approaching the rate/distortion bound usually requires quantization on asymptotically long blocks of source vectors. Such an approach is not applicable to situations with a strict delay constraint, which is our case. The receiver cannot buffer several velocity matrices together and perform a long VQ, because the channel state information would become outdated. In practice, numerous structured VQs have been proposed to achieve a good complexity-performance tradeoff. Examples of good VQ schemes for i.i.d. Gaussian sources include lattice quantizers and shape-gain quantizers. However, most of these quantization schemes

only approach their optimality in high resolution/high rate cases. In contrast, this chapter is mainly concerned with low rate quantization with a rate of much less than 1 bit/sample because of the feedback bandwidth constraints. With low-rate VQ and, therefore, a small VQ codebook, the search complexity in the quantization procedure is not much of a concern.

Based on the above considerations, we take a very basic approach to designing the VQ for velocity matrices: the LBG algorithm [40]. The Gaussian VQ codebook was trained with 2×10^5 i.i.d. Gaussian vectors with $\mathcal{N}(0,1)$ entries. The codebook training process involved a distance measure between two vectors (one training vector and one codeword). Choosing which distance measure to use posed an interesting problem. One possibility was to use the distance between the subspaces reached by the geodesics using the two vectors as velocity matrices starting from a fixed subspace. However, such a codebook would probably depend on the average geodesic length and, therefore, the Doppler frequency. To avoid over-complicating the problem, mean-squared error was used in the codebook training process.

An optimal average geodesic length a versus Doppler frequency was utilized to scale the codewords in the Gaussian codebook to minimize the mean tracking error. Blind search was utilized to perform this optimization. Specifically, for each integer-valued Doppler frequency in the range of 1 to 400 Hz, simulations were performed for a taking values from a grid with small step size. Then an eighth-order polynomial was fit to the curve. The resultant values of a^2 versus Doppler frequency, F_D, are shown in Fig. 4. The corresponding polynomial coefficients are listed in Appendix A.3.

Let $\boldsymbol{D}_k$, $k = 0, 1, \ldots, 2^N - 1$ be the codewords in the scaled VQ codebook. At time $n+1$, the codeword chosen by the receiver is defined by

$$\arg\max_k \left\| \boldsymbol{H}[n+1]\boldsymbol{Q}[n] \exp \begin{pmatrix} \boldsymbol{0} & -\boldsymbol{D}_k^H \\ \boldsymbol{D}_k & \boldsymbol{0} \end{pmatrix} \boldsymbol{I}_{N_t,N_s} \right\|_F^2 \tag{52}$$

using the criterion of maximizing the received power. As mentioned in Section 2.3, the matrix exponentials can be efficiently computed using SVDs of the matrices $\{\boldsymbol{D}_k\}$. Computing the SVD of each codeword at run-time can be avoided by storing the SVDs of the codewords at the receiver. The aforementioned scaling only needs to be performed on the singular values (not singular vectors) of the codeword matrices.

3.7 Numerical Results

A Monte Carlo simulation was run to test the performance of the proposed transmit subspace tracking algorithms in the setting of $N_s = N_r = 2$ and $N_t = 8$. The channel estimation error was 20 dB below the channel power in the following numerical examples. The feedback channel was free of error. The channel model was independent Rayleigh flat fading with time correlation generated by Jakes' method. The relationship between the Doppler frequency F_D and the feedback rate F_{FB} was captured by the ratio F_{FB}/F_D.

The numerical results were compared with the feedback schemes in [4, 43]. For the gradient sign algorithm [4], an optimal step size β was utilized, which was determined through numerical search.

In the figure legends, *geodesic* denotes our one-bit quantization algorithm and *N-bit Gaussian VQ* denotes the VQ quantization in Section 3.6 with a 2^N-sized codebook. Regarding the figure legends for the other techniques, *ideal* denotes ideal subspace tracking, *gradient sign* stands for the gradient sign algorithm of [4], *N-bit Grassmann 26 (50)* stands for Grassmannian subspace packing as presented in [43] with linear predictor length 26 (50) and a codebook size of 2^N, and *no CSI* stands for transmission without CSI. The length of the linear prediction filter for Gaussian VQ is always 50.

The average convergence performance for the tracking of constant channels is illustrated in Fig. 5, where the mean values of cost function J from Eq. (15) are shown in dB relative to the cost of ideal subspace tracking. As can be seen, larger values of a and β resulted in faster convergence but larger stable-state error. Our one-bit feedback algorithm showed faster convergence and smaller stable-state error than the gradient sign algorithm.

The ergodic capacities achieved by different algorithms are given in Fig. 6 for $F_{FB}/F_D = 100$ and 1000. The capacity of the subspace tracking algorithms is given by [4]

$$E\left(\log\det\left(\boldsymbol{I} + \frac{E_s}{N_s N_0}\boldsymbol{H}[n]\boldsymbol{W}[n]\boldsymbol{W}[n]^H\boldsymbol{H}[n]^H\right)\right), \tag{53}$$

while the capacity of transmission without CSI is given by [4]

$$E\left(\log\det\left(\boldsymbol{I} + \frac{E_s}{N_t N_0}\boldsymbol{H}[n]\boldsymbol{H}[n]^H\right)\right), \tag{54}$$

where E_s is the total transmission energy per symbol from all transmit antennas. Our one-bit quantization algorithm shows a 2-3 dB gain over the gradient sign algorithm across the medium-to-high SNR range. For the case $F_{FB}/F_D = 1000$, representing high rate feedback or low Doppler frequency, both our one-bit feedback algorithm and the gradient sign feedback outperform the Grassmannian subspace packing algorithm. When $F_{FB}/F_D = 100$, the one-bit feedback algorithm is slightly worse than the Grassmannian subspace packing algorithm.

The comparison of the tracking performance in terms of the cost function J is shown in Fig. 7, where the results are normalized to the performance of ideal subspace tracking (J_{opt}). Our one-bit feedback algorithm performs uniformly better than the gradient sign feedback, and is better than the Grassmannian subspace packing algorithm for F_{FB}/F_D larger than 110. For the range of parameters in these simulations, the Gaussian VQ algorithm is always better than the one-bit quantization algorithm. It can be verified that in the current setting, $J_{opt}[n] = \|\boldsymbol{H}[n]\|_F^2$. While for transmission without CSI, $J[n] = \frac{N_s}{N_t}\|\boldsymbol{H}[n]\|_F^2$, and $J[n]/J_{opt}[n] = \frac{N_s}{N_t} = 0.25$ regardless of the Doppler frequency. Therefore, all the feedback techniques in these simulations are better than transmission without CSI.

Finally, we compare the Gaussian VQ algorithm and Grassmannian subspace packing with different codebook sizes in Fig. 8, where the feedback frequency is fixed to be 6000 Hz and the Doppler frequency varies. Fig. 8 (a) shows the average J/J_{opt} for every time instant, as in Fig. 7. The Gaussian VQ with a codebook size of 32 always performs better than the Grassmannian subspace packing with 1024 codewords, and performs better than 32-codeword Grassmannian subspace packing until $F_{FB}/F_D = 20$. Fig. 8 (b) shows the performance of only those time instants that are predicted and quantized, without counting the time instants when the transmit subspace is held constant before a new feedback codeword arrives. It can be seen that the 10-bit Grassmannian subspace packing eventually performs worse than the Gaussian VQ technique in high Doppler frequencies because its feedback interval is too long. Though 5-bit Grassmannian subspace packing outperforms 5-bit Gaussian VQ when $F_D > 230$ Hz, the advantage was almost canceled by the time instants when the transmit subspace is held constant. We also plotted *bit error rate* (BER)

performance of Gaussian VQ, Grassmannian subspace packing, and transmission without CSI in Fig. 9 for every time instant. The data vector $\boldsymbol{d}[n]$ uses 16-QAM constellation for each entry. For a fair comparison, the transmitter without CSI uses arbitrarily chosen two transmit antennas. The simulated system is uncoded. The maximum likelihood detector is employed at the receiver. Transmission without CSI shows a fixed BER regardless of Doppler frequency. The behaviors of Gaussian VQ and Grassmannian subspace packing codebooks are similar to those observed in Fig. 8. Figs. 8 and 9 bring up interesting questions regarding the optimal size of a codebook versus Doppler frequency and what the transmission subspace should be before the new codeword arrives.

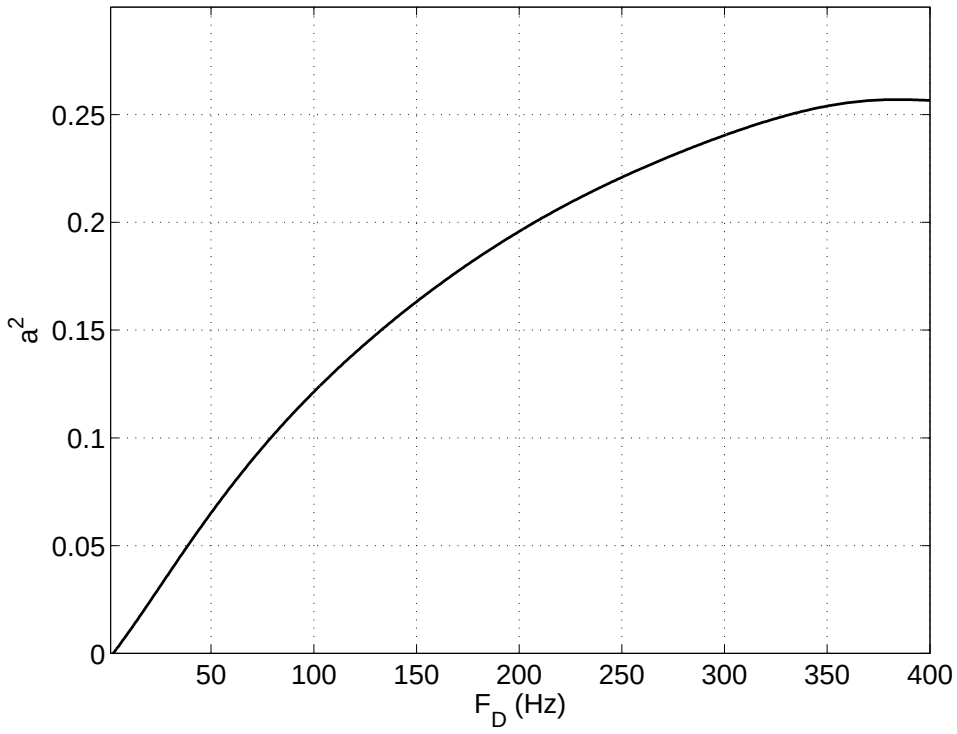


Figure 4: Numerically optimized a^2 versus F_D.

3.8 Conclusion

A low rate feedback algorithm for transmission subspace tracking has been introduced. By treating the variation of the transmit weight matrix as a piece-wise geodesic process in Grassmann manifolds, the tracking algorithms feedback the quantized velocity matrix. The proposed one-bit quantization algorithm tracks fewer parameters than the gradient sign

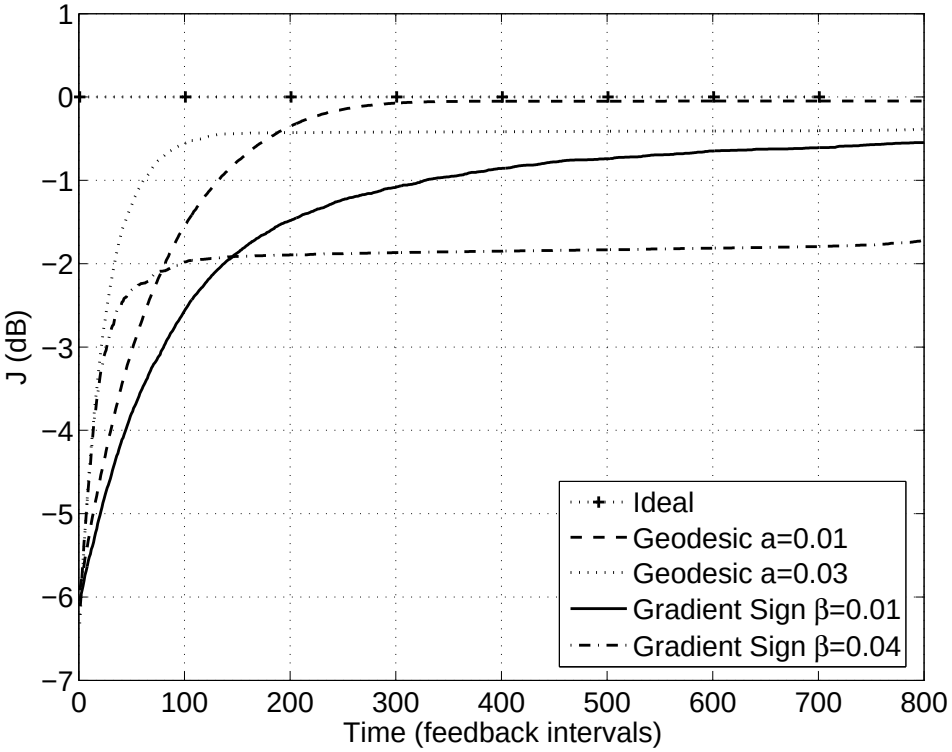
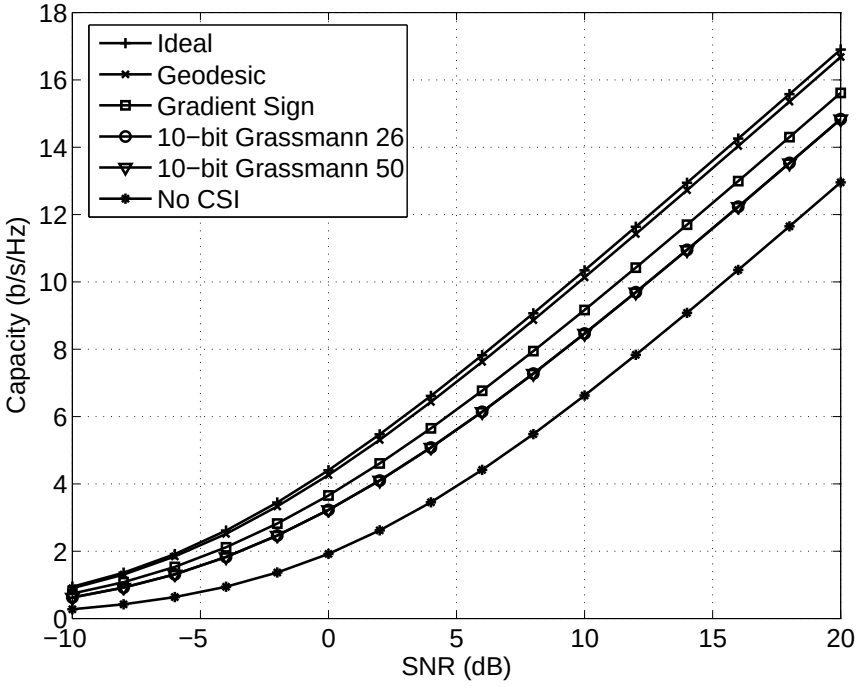
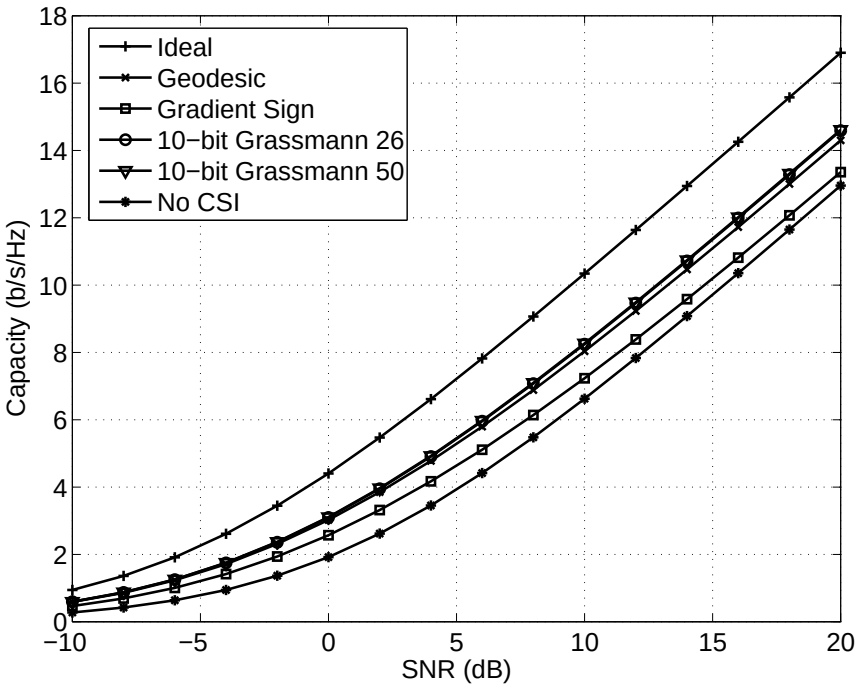


Figure 5: Convergence rate of the received power for the geodesic and gradient sign algorithms as compared to ideal subspace tracking.



(a)



(b)

Figure 6: Ergodic capacity versus SNR. (a) $F_{FB}/F_D = 1000$; (b) $F_{FB}/F_D = 100$.

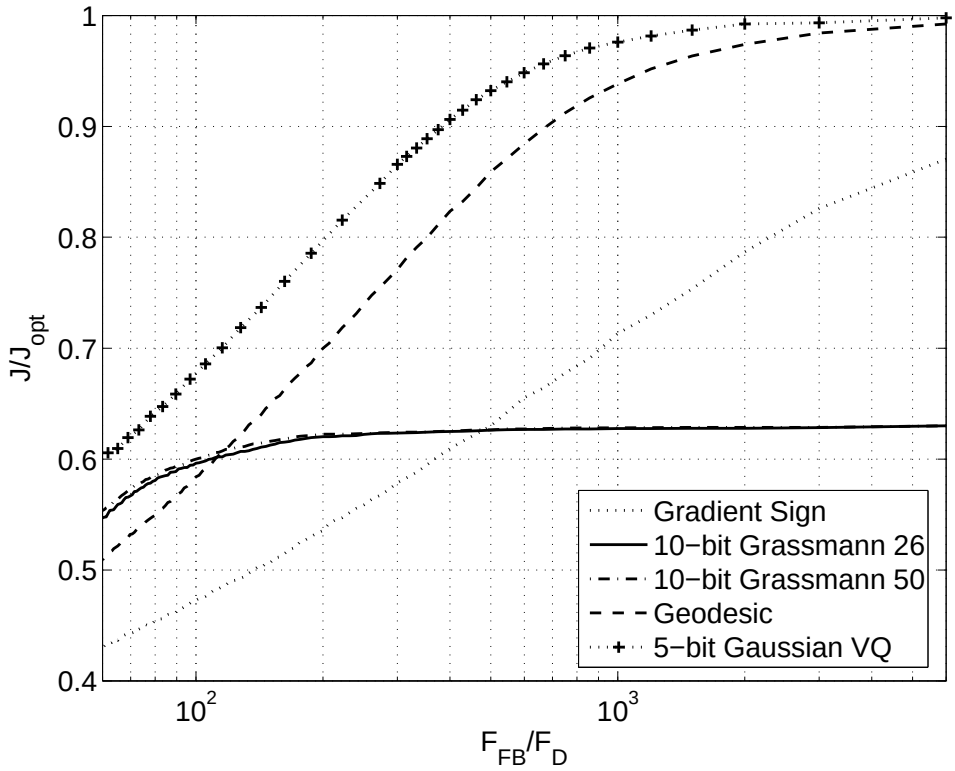
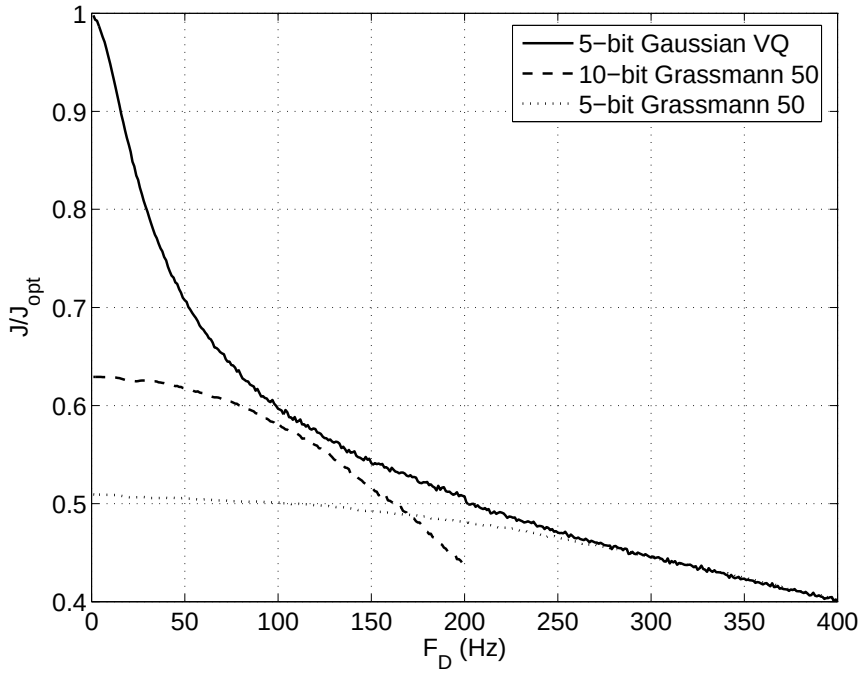
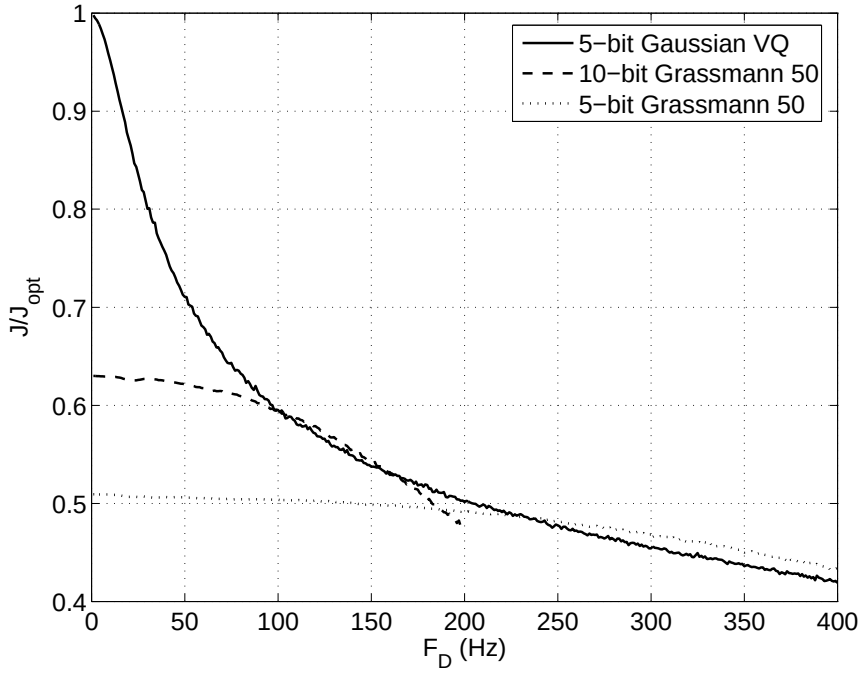


Figure 7: Normalized mean cost function versus F_{FB}/F_D.

algorithm and, consequently, achieves better performance. Numerical results show that the performance of this adaptive algorithm approaches that of ideal subspace tracking for feedback rates on the order of 1000 times the channel Doppler frequency. Compared with a Grassmannian subspace packing quantization algorithm, our one-bit algorithm has better performance at low-to-medium Doppler frequencies and does not incur the complexity of quantization and long-range channel prediction, while the Gaussian VQ always performs better than the one-bit algorithm. The subspace tracking approach was shown to provide significant capacity gains over transmission without CSI.



(a)



(b)

Figure 8: Normalized mean cost function versus F_D with F_{FB}=6000 Hz. (a) Average performance of all time instants; (b) Average performance of predicted time instants only.

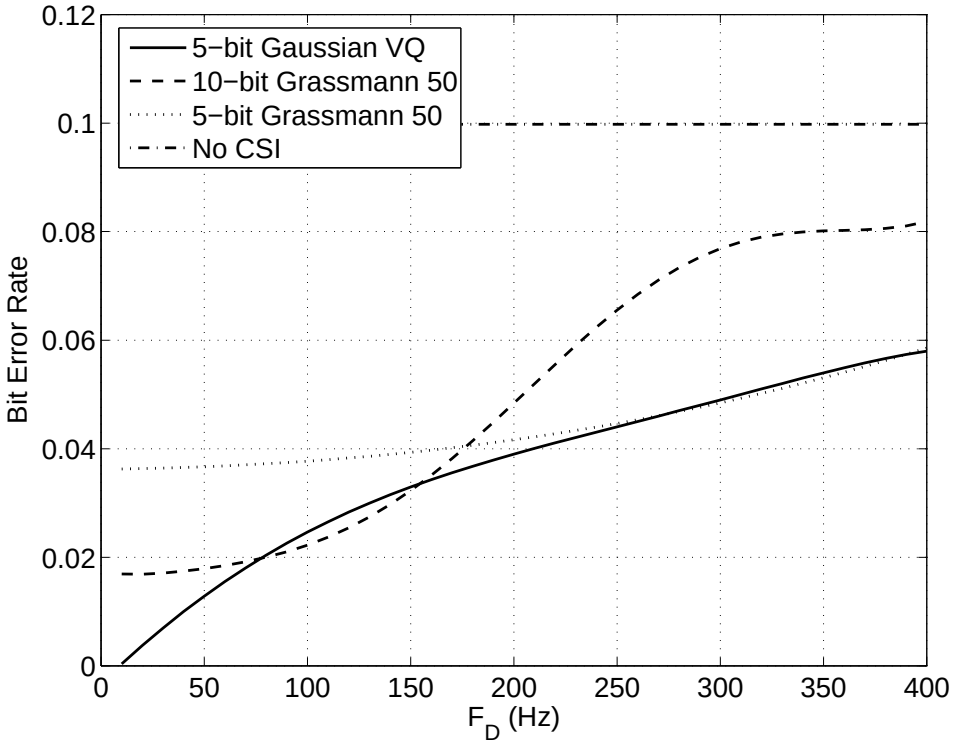


Figure 9: Bit error rate versus F_D with F_{FB}=6000 Hz.

CHAPTER IV

LOW COMPLEXITY OFDM MIMO SYSTEM BASED ON CHANNEL CORRELATIONS

In this chapter, we investigate transmitter design given partial CSI feedback in the context of MIMO OFDM. Traditionally, for a transmitter with perfect CSI, frequency-domain beamforming and combining are used [60]. Since MIMO channels on different subcarriers are different, beamforming and combining vectors are different on each subcarrier, implying that each antenna requires an (I)DFT operation.

Some techniques to reduce the number of DFT blocks were recently proposed [33, 34, 56, 67]. [56, 67] deal with the *Single Input Multiple Output* (SIMO) case, where the receiver uses time domain combining followed by a single DFT. These works are extended in [33] to the MIMO case, where the transmitter also uses a single IDFT followed by time domain beamforming. It was shown that considerable diversity gain can still be achieved with the time domain beamforming and/or combining. [34] trades off 3 dB power gain with the reduced number of DFT blocks.

Most previous works on closed-loop MIMO OFDM systems assume that the transmitter has full knowledge of the underlying channel. When the channel varies rapidly, it is costly to keep the channel knowledge at the transmitter up-to-date. However, the statistics of the channel is time-invariant as long as the channel stays stationary.

In this chapter, we investigate a complexity-constrained OFDM MIMO system, where the transmitter has channel covariance information [78]. The transmitter and the receiver are constrained to use a single (I)DFT per OFDM block (on the average), and the beamforming or combining is performed in the time domain. We show that in the MISO case, under the given channel model, time domain *two-dimensional* (2-D) beamforming does not incur any performance loss. For the MIMO case, we derive design criteria for the transmitter beamforming and receiver combining weighting vectors. Due to analytic complexity,

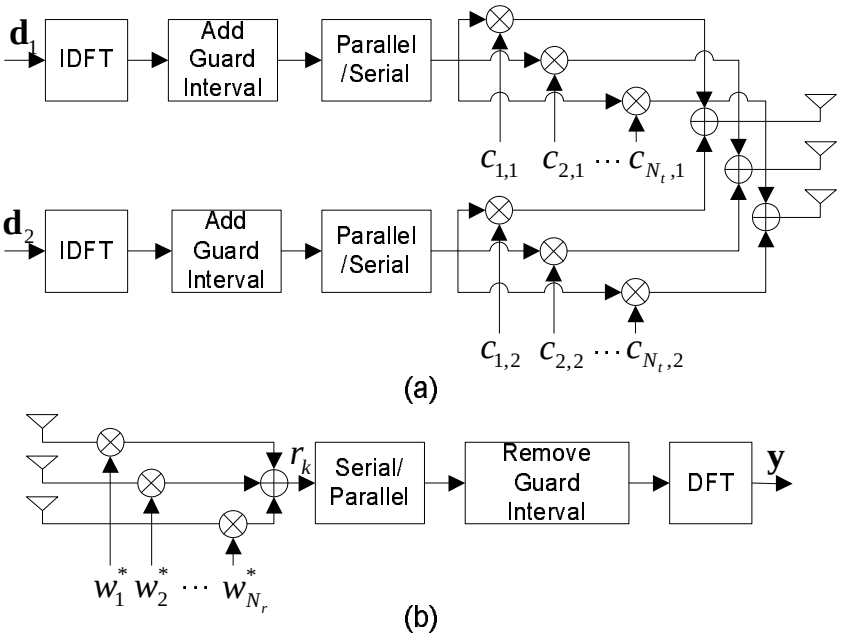


Figure 10: Block diagram of the MIMO OFDM system. (a) Transmitter; (b) Receiver.

only suboptimal solutions are given. The performance of the low-complexity OFDM system is asymptotically analyzed.

4.1 System Model

We consider a MIMO OFDM system as shown in Fig. 10. The numbers of transmit and receive antennas are N_t and N_r, respectively. Denote the time-domain multipath MIMO channel matrix as $\boldsymbol{G}_l$, $l = 0, 1, \ldots, L-1$, where the $N_r \times N_t$ random matrix $\boldsymbol{G}_l$ represents the lth tap of the discrete-time MIMO fading channel impulse response, L is the total number of resolvable multipaths. We focus on the downlink case, i.e., the base station and the *subscriber unit* (SU) assume the roles of transmitter and receiver, respectively. Assuming that the channel decays to zero within the length of cyclic prefix, the MIMO frequency response is given by the DFT of the $\boldsymbol{G}_l$'s:

$$\boldsymbol{H}_k = \sum_{l=0}^{L-1} \boldsymbol{G}_l e^{-j\frac{2\pi}{N}lk},\ k = 0, 1, \ldots, N-1, \tag{55}$$

where N is the total number of subcarriers and $\boldsymbol{H}_k$ is the flat-fading MIMO channel on the kth subcarrier. Organizing the transmitted data symbols into vectors

$$\boldsymbol{d}_k = \left[d_k^{(0)} \; d_k^{(1)} \; \ldots \; d_k^{(N_t-1)} \right]^T$$

with $d_k^{(i)}$ denoting the data symbol transmitted from the ith antenna on the kth subcarrier and $()^T$ standing for transpose, it can be shown that

$$\boldsymbol{r}_k = \boldsymbol{H}_k \boldsymbol{d}_k + \boldsymbol{n}_k, \tag{56}$$

where $\boldsymbol{r}_k$ denotes the $N_r \times 1$ received data vector for the kth subcarrier, and $\boldsymbol{n}_k$ is additive noise. Denote the kth row of $\boldsymbol{G}_l$ as $\boldsymbol{g}_{l,k}$, and the mth entry of it as $g_{l,k,m}$. We adopt the following assumptions regarding the channel:

A-I. The elements of the $\boldsymbol{G}_l$'s are zero-mean circularly symmetric complex Gaussian random variables, implying $E\{g_{l,k,m}^2\} = 0$;

A-II. Different multipath taps are assumed to be uncorrelated, that is,

$$E\{g_{l_1,k_1,m_1} g_{l_2,k_2,m_2}^*\} = 0$$

for $l_1 \neq l_2$;

A-III. Different rows of each $\boldsymbol{G}_l$ are uncorrelated, that is, $E\{g_{l,k_1,m_1} g_{l,k_2,m_2}^*\} = 0$ for $k_1 \neq k_2$;

A-IV. The row $\boldsymbol{g}_{l,k} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{R}_l)$, where $\boldsymbol{R}_l = E\{\boldsymbol{g}_{l,k}^H \boldsymbol{g}_{l,k}\}$, which is independent of k, i.e., the fading statistics are the same for all the receive antennas; Besides, the $\boldsymbol{R}_l$'s are known to the transmitter;

A-V. $\boldsymbol{n}_k$ is zero-mean additive white Gaussian noise that satisfies $E\{\boldsymbol{n}_k \boldsymbol{n}_l^H\} = \sigma_n^2 \boldsymbol{I}_{N_r} \delta[k-l]$, and is independent of both the channel and the data symbols.

In the above, $\boldsymbol{I}_{N_r}$ is an $N_r \times N_r$ identity matrix and δ is Kronecker's delta function. **A-II** follows from the assumption that each multipath tap (or each $\boldsymbol{G}_l$) corresponds to multipaths emanated from a distinct scatterer cluster, and the fading of different scatterer clusters is

independent [8]. **A-III** is reasonable because we assume the SU is surrounded by local scatterers so that fading at the SU antennas is spatially uncorrelated.

Relative to the transmitter, each scatterer cluster has a mean angle of departure $\bar{\theta}_l$ and an angle spread δ_l. The distribution of the actual angle of departure θ_l of the lth path cluster is modeled as Gaussian: $\mathcal{N}(\bar{\theta}_l, \sigma^2_{\theta_l})$, where the variance $\sigma^2_{\theta_l}$ is proportional to the angular spread δ_l. It was shown in [8] that with a uniform linear array at both the transmitter and the receiver, for small angular spread, the covariance matrix can be approximated as

$$[\boldsymbol{R}_l]_{m,n} \approx e^{j2\pi(n-m)d\cos(\bar{\theta}_l)/\lambda} e^{-0.5(2\pi(n-m)d\sin(\bar{\theta}_l)\sigma_{\theta_l}/\lambda)^2}, \tag{57}$$

where $[\boldsymbol{R}_l]_{m,n}$ denotes the entry of $\boldsymbol{R}_l$ on the mth row and nth column, d is the spacing between adjacent transmit antennas, and λ is the wavelength of the radio-frequency signal. Factoring the $N_t \times N_t$ covariance matrix $\boldsymbol{R}_l$ as $\boldsymbol{R}_l = \boldsymbol{R}_l^{\frac{1}{2}} \boldsymbol{R}_l^{\frac{1}{2}}$, it can be shown that

$$\boldsymbol{G}_l \sim \boldsymbol{G}_{w,l} \boldsymbol{R}_l^{\frac{1}{2}}, \tag{58}$$

where $\boldsymbol{G}_{w,l}$ is an $N_r \times N_t$ matrix with *independent and identically distributed* (i.i.d.) circularly symmetric $\mathcal{CN}(0,1)$ entries.

Similar to the approach in [8], it can be shown that all the subcarrier MIMO channels have the same distribution. Specifically, with the definition $\boldsymbol{R} = \sum_{l=0}^{L-1} \boldsymbol{R}_l$, we have

$$\boldsymbol{H}_k \sim \boldsymbol{H}_{w,k} \boldsymbol{R}^{\frac{1}{2}}, k = 0, 1, \ldots, N-1, \tag{59}$$

where $\boldsymbol{H}_{w,k}$ is an $N_r \times N_t$ matrix with i.i.d. circularly symmetric $\mathcal{CN}(0,1)$ entries.

4.2 Time-Domain MISO Eigen-Beamforming

In this section, we show that an optimal 2-D eigen-beamformer combined with Alamouti *space-time block coding* (STBC) can be implemented using one IDFT per block (on the average) at the transmitter for the MISO channels. The system model (56) can be rewritten

as $\boldsymbol{r} = \boldsymbol{H}\boldsymbol{d} + \boldsymbol{n}$, where

$$\begin{aligned}
\boldsymbol{r} &= \left[\boldsymbol{r}_0^T \ \boldsymbol{r}_1^T \ \ldots \ \boldsymbol{r}_{N-1}^T\right]^T, \\
\boldsymbol{d} &= \left[\boldsymbol{d}_0^T \ \boldsymbol{d}_1^T \ \ldots \ \boldsymbol{d}_{N-1}^T\right]^T, \\
\boldsymbol{n} &= \left[\boldsymbol{n}_0^T \ \boldsymbol{n}_1^T \ \ldots \ \boldsymbol{n}_{N-1}^T\right]^T, \\
\boldsymbol{H} &= \mathrm{diag}\left(\boldsymbol{H}_0 \ \boldsymbol{H}_1 \ \ldots \ \boldsymbol{H}_{N-1}\right),
\end{aligned}$$

and diag() is a (block) diagonal matrix with the given entries on the diagonal. Note that in the MISO case, all the $\boldsymbol{G}_l$'s and $\boldsymbol{H}_k$'s degenerate to row vectors. From [74], to maximize the mutual information between $\boldsymbol{d}$ and $\boldsymbol{r}$, $\boldsymbol{d}$ should be circularly symmetric complex Gaussian. With $\boldsymbol{Q} = E\{\boldsymbol{d}\boldsymbol{d}^H\}$, the corresponding mutual information is

$$\mathcal{I}\left(\boldsymbol{d}; (\boldsymbol{r}, \boldsymbol{H})\right) = E\left\{\log\left|\boldsymbol{I} + \frac{1}{\sigma_n^2}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\right|\right\}, \tag{60}$$

where $|.|$ denotes matrix determinant. Note that $\boldsymbol{Q}$ has a block structure with $\boldsymbol{Q}_{i,j} = E\{\boldsymbol{d}_i\boldsymbol{d}_j^H\}$ as composing components.

Lemma 1. *The $\boldsymbol{Q}$ that maximizes the mutual information is block diagonal, i.e., $\boldsymbol{Q}_{i,j} = \boldsymbol{0}$, $i \neq j$.*

Proof. Note that $(\boldsymbol{I} + \frac{1}{\sigma_n^2}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H)$ is positive definite. With any given $\boldsymbol{H}$ and $\boldsymbol{Q}$, from Hadamard's inequality [32],

$$\begin{aligned}
&\left|\boldsymbol{I} + \frac{1}{\sigma_n^2}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\right| \\
&\leq \prod_k \left(1 + \frac{1}{\sigma_n^2}\boldsymbol{H}_k\boldsymbol{Q}_{k,k}\boldsymbol{H}_k^H\right) \\
&= \left|\boldsymbol{I} + \frac{1}{\sigma_n^2}\boldsymbol{H}\,\mathrm{diag}\left(\boldsymbol{Q}_{0,0} \ \boldsymbol{Q}_{1,1} \ \ldots \ \boldsymbol{Q}_{N-1,N-1}\right)\boldsymbol{H}^H\right|.
\end{aligned} \tag{61}$$

□

Following (60) and (61),

$$\mathcal{I}\left(\boldsymbol{d}; (\boldsymbol{r}, \boldsymbol{H})\right) = \sum_{k=0}^{N-1} E\{\log(1 + \frac{1}{\sigma_n^2}\boldsymbol{H}_k\boldsymbol{Q}_{k,k}\boldsymbol{H}_k^H)\}. \tag{62}$$

Now we find the optimal $\boldsymbol{Q}_{k,k}$'s that maximize (62). Since all $\boldsymbol{H}_k$'s have the same distribution (59), the optimal $\boldsymbol{Q}_{k,k}$'s should all be equal, and

$$\mathcal{I}\left(\boldsymbol{d};(\boldsymbol{r},\boldsymbol{H})\right) = NE\{\log(1+\frac{1}{\sigma_n^2}\boldsymbol{H}_0\boldsymbol{Q}_{0,0}\boldsymbol{H}_0^H)\}. \tag{63}$$

Therefore, this model is as if we are dealing with a single flat fading channel, and all the analysis and designs in [35,36,54,64,76,87] are applicable here. Specifically, [35,64,76] imply the following result on $\boldsymbol{Q}_{k,k}$, which we state without proof. Let the eigen-decomposition of $\boldsymbol{Q}_{k,k}$ be $\boldsymbol{Q}_{k,k} = \boldsymbol{U}_d\boldsymbol{\Lambda}_d\boldsymbol{U}_d^H$, where $\boldsymbol{U}_d$ contains the eigenvectors, and the diagonal matrix $\boldsymbol{\Lambda}_d$ has the eigenvalues on its diagonal. Similarly, let the eigen structure of $\boldsymbol{R}$ be $\boldsymbol{R} = \boldsymbol{U}_R\boldsymbol{\Lambda}_R\boldsymbol{U}_R^H$. Then we have

Theorem 1. *With assumptions* **A-I** *through* **A-V**, *the channel capacity-achieving distribution of $\boldsymbol{d}_k$ is circularly symmetric complex Gaussian with zero mean, and the ordered eigenvectors of its covariance matrix are the same as the ordered eigenvectors of $\boldsymbol{R}$, i.e., $\boldsymbol{U}_d = \boldsymbol{U}_R$. The eigenvalues of $\boldsymbol{Q}_{k,k}$ are determined by a water-filling process.*

Theorem 1 implies that if the transmitter is constrained to use *one-dimensional* (1-D) beamforming, $\boldsymbol{d}_k$ is proportional to the dominant eigenvector of $\boldsymbol{R}$. Since the optimal 1-D beamforming vector is common to all subcarriers, optimal 1-D beamforming can be implemented using one IDFT per OFDM block, and the beamforming can be done in the time domain.

Combination of 2-D eigen-beamforming and Alamouti STBC [2] can achieve better performance than 1-D eigen-beamforming [87], and can also be implemented using one IDFT per OFDM block on the average. To see this, let $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ be the matrices containing time-domain symbols transmitted in two consecutive OFDM blocks, each row of which contains time-domain symbols transmitted from one antenna, then 2-D eigen-beamforming

with Alamouti STBC can be represented as

$$\boldsymbol{D}_1 = \boldsymbol{U}_2 \begin{bmatrix} f_1 & 0 \\ 0 & f_2 \end{bmatrix} \begin{bmatrix} d_{1,0} & d_{1,1} & \dots & d_{1,N-1} \\ d_{2,0} & d_{2,1} & \dots & d_{2,N-1} \end{bmatrix} \boldsymbol{F}^T, \tag{64}$$

$$\boldsymbol{D}_2 = \boldsymbol{U}_2 \begin{bmatrix} f_1 & 0 \\ 0 & f_2 \end{bmatrix} \begin{bmatrix} d_{2,0}^* & d_{2,1}^* & \dots & d_{2,N-1}^* \\ -d_{1,0}^* & -d_{1,1}^* & \dots & -d_{1,N-1}^* \end{bmatrix} \boldsymbol{F}^T, \tag{65}$$

where the beamforming matrix $\boldsymbol{U}_2$ consists of the two dominant eigenvectors of $\boldsymbol{R}$, f_1 and f_2 are the power loading factors of the two eigen-beams, $d_{m,k}$'s are i.i.d. transmitted symbols on the mth eigen-beam and the kth subcarrier, and $\boldsymbol{F}$ is the IDFT matrix. To transmit $\boldsymbol{D}_1$, two IDFT's are performed. To transmit $\boldsymbol{D}_2$, no more IDFTs are required because the IDFT of $\begin{bmatrix} d_{m,0}^* & d_{m,1}^* & \dots & d_{m,N-1}^* \end{bmatrix}$ can be obtained from the IDFT of $\begin{bmatrix} d_{m,0} & d_{m,1} & \dots & d_{m,N-1} \end{bmatrix}$ by conjugation and shifting [34]. Therefore, for the MISO channel and with the given channel model, combined optimal eigen-beamforming and Alamouti STBC can be implemented with the average complexity of one IDFT per OFDM block.

4.3 Time-Domain Eigen-Beamforming for MIMO Channels

4.3.1 Beamforming and Combining Vectors

In this section, we find transmission strategies for MIMO channels given the constraint that the average complexity is one (I)DFT per OFDM block, at both the transmitter and the receiver. At the transmitter, the best we can do is still 2-D eigen-beamforming, which subsumes 1-D eigen-beamforming as a special case [87]. At the receiver, a time-domain combining is performed before the DFT operation (Fig. 10(b)).

Let the $N_t \times 2$ matrix $\boldsymbol{C} = \begin{bmatrix} \boldsymbol{c}_1 & \boldsymbol{c}_2 \end{bmatrix}$ contain the power-loaded eigen-beamforming vectors. Let $y_{1,k}$ and $y_{2,k}$ be the output of the kth subcarrier at the receiver after the DFT in two consecutive OFDM blocks, then

$$y_{1,k} = \boldsymbol{w}^H \boldsymbol{H}_k \boldsymbol{c}_1 d_{1,k} + \boldsymbol{w}^H \boldsymbol{H}_k \boldsymbol{c}_2 d_{2,k} + \boldsymbol{w}^H \boldsymbol{n}_{1,k}, \tag{66}$$

$$y_{2,k} = \boldsymbol{w}^H \boldsymbol{H}_k \boldsymbol{c}_1 d_{2,k}^* - \boldsymbol{w}^H \boldsymbol{H}_k \boldsymbol{c}_2 d_{1,k}^* + \boldsymbol{w}^H \boldsymbol{n}_{1,k}, \tag{67}$$

where $\boldsymbol{w}$ is the $N_r \times 1$ receiver combining vector and is constrained to have unit norm, and $\boldsymbol{n}_{1,k}$ and $\boldsymbol{n}_{2,k}$ are additive noise on the kth subcarrier. In (66), $d_{m,k}$ denotes the source

symbol transmitted from the mth eigen-beam on the kth subcarrier. A standard Alamouti combining gives the following equivalent channel model:

$$\begin{bmatrix} r_{1,k} \\ r_{2,k} \end{bmatrix} = \sqrt{\boldsymbol{w}^H \boldsymbol{H}_k \boldsymbol{C}\boldsymbol{C}^H \boldsymbol{H}_k^H \boldsymbol{w}} \begin{bmatrix} d_{1,k} \\ d_{2,k} \end{bmatrix} + \begin{bmatrix} n_{1,k} \\ n_{2,k} \end{bmatrix},$$

where the noise components $\begin{bmatrix} n_{1,k} & n_{2,k} \end{bmatrix}^T \sim \mathcal{CN}(\boldsymbol{0}, \sigma_n^2 \boldsymbol{I})$.

Now the task is to design the combining vector $\boldsymbol{w}$ given perfect channel knowledge at the receiver, and the beamforming vectors $\boldsymbol{C}$ given second-order statistics of the channel. The joint design of $\boldsymbol{w}$ and $\boldsymbol{C}$ turns out to be hardly tractable. We take the approach to design them separately. To design $\boldsymbol{C}$, we assume that $\boldsymbol{w}$ is a fixed unit-norm vector. Since $\boldsymbol{H}_k \sim \boldsymbol{H}_{w,k} \boldsymbol{R}^{\frac{1}{2}}$, it can be shown that $\boldsymbol{w}^H \boldsymbol{H}_k \sim \boldsymbol{h}_{w,k} \boldsymbol{R}^{\frac{1}{2}}$, where $\boldsymbol{h}_{w,k}$ is a $1 \times N_t$ row vector with i.i.d. $\mathcal{CN}(0,1)$ entries. Therefore, by treating $\boldsymbol{w}$ as a fixed vector, we obtain an MISO channel model for each subcarrier, whose second-order statistics are known to the transmitter. The conclusion in Theorem 1 can then be applied to design $\boldsymbol{C}$.

The combining vector $\boldsymbol{w}$ can be designed to maximize capacity or minimize bit error rate. However, these are not trivial tasks. Therefore, we resort to maximizing the total SNR of all the subcarriers:

$$\max_{\boldsymbol{w}, \boldsymbol{C}} \sum_{k=0}^{N-1} \mathrm{SNR}_k. \tag{68}$$

One justification of this criterion was given in [33,56], where it was assumed that a repetition code is used across evenly spaced subcarriers, and it was shown that the maximal-ratio combined SNR of each symbol is proportional to $\sum_{k=0}^{N-1} \mathrm{SNR}_k$, if the number of subcarriers used to transmit one symbol is larger than the channel span L. The SNR of the kth subcarrier output is

$$\mathrm{SNR}_k = \frac{\boldsymbol{w}^H \boldsymbol{H}_k \boldsymbol{C}\boldsymbol{C}^H \boldsymbol{H}^H \boldsymbol{w}}{\sigma_n^2}. \tag{69}$$

It is easy to show that

$$\begin{aligned} \sum_{k=0}^{N-1} \mathrm{SNR}_k &= \frac{1}{\sigma_n^2} \boldsymbol{w}^H \sum_{k=0}^{N-1} \left(\boldsymbol{H}_k \boldsymbol{C}\boldsymbol{C}^H \boldsymbol{H}_k^H \right) \boldsymbol{w} \\ &= \frac{N}{\sigma_n^2} \boldsymbol{w}^H \sum_{l=0}^{L-1} \left(\boldsymbol{G}_l \boldsymbol{C}\boldsymbol{C}^H \boldsymbol{G}_l^H \right) \boldsymbol{w}. \end{aligned} \tag{70}$$

Define $\boldsymbol{\mathcal{H}} = \sum_{l=0}^{L-1} \left(\boldsymbol{G}_l \boldsymbol{C}\boldsymbol{C}^H \boldsymbol{G}_l^H \right)$. With the constraint that $\|\boldsymbol{w}\| = 1$, we can immediately get that the optimal $\boldsymbol{w}$ is the dominant eigenvector of the matrix $\boldsymbol{\mathcal{H}}$, and the resultant total SNR is $N\lambda_1(\boldsymbol{\mathcal{H}})/\sigma_n^2$, where $\lambda_1(\boldsymbol{\mathcal{H}})$ is the largest eigenvalue of $\boldsymbol{\mathcal{H}}$.

4.3.2 Performance Analysis

From (58), and with **A-I**, it can be shown that

$$\lambda_1(\boldsymbol{\mathcal{H}}) \sim \lambda_1(\sum_{l=0}^{L-1} \boldsymbol{G}_{w,l} \boldsymbol{R}_l^{\frac{1}{2}} \boldsymbol{C}\boldsymbol{C}^H \boldsymbol{R}_l^{\frac{1}{2}} \boldsymbol{G}_{w,l}^H). \tag{71}$$

To write the above summation in a matrix form, first define

$$\tilde{\boldsymbol{R}}^{\frac{1}{2}} = \text{diag}\left(\boldsymbol{C}^H \boldsymbol{R}_0^{\frac{1}{2}} \ \boldsymbol{C}^H \boldsymbol{R}_1^{\frac{1}{2}} \ \ldots \ \boldsymbol{C}^H \boldsymbol{R}_{L-1}^{\frac{1}{2}} \right), \tag{72}$$

$$\boldsymbol{\mathcal{G}} = \left[\boldsymbol{G}_{w,0} \ \ldots \ \boldsymbol{G}_{w,L-1} \right]^H. \tag{73}$$

Then, we have

$$\sum_{l=0}^{L-1} \boldsymbol{G}_{w,l} \boldsymbol{R}_l^{\frac{1}{2}} \boldsymbol{C}\boldsymbol{C}^H \boldsymbol{R}_l^{\frac{1}{2}} \boldsymbol{G}_{w,l}^H = \boldsymbol{\mathcal{G}}^H \tilde{\boldsymbol{R}}^{\frac{1}{2}H} \tilde{\boldsymbol{R}}^{\frac{1}{2}} \boldsymbol{\mathcal{G}}. \tag{74}$$

Therefore, the following holds true:

$$\lambda_1(\boldsymbol{\mathcal{H}}) \sim \lambda_1(\boldsymbol{\mathcal{G}}^H \tilde{\boldsymbol{R}}^{\frac{1}{2}H} \tilde{\boldsymbol{R}}^{\frac{1}{2}} \boldsymbol{\mathcal{G}}) = \lambda_1(\tilde{\boldsymbol{R}}^{\frac{1}{2}} \boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{G}}^H \tilde{\boldsymbol{R}}^{\frac{1}{2}H}). \tag{75}$$

The $2L \times N_r$ matrix $\tilde{\boldsymbol{R}}^{\frac{1}{2}} \boldsymbol{\mathcal{G}}$ has i.i.d. columns, and each column is distributed as $\mathcal{CN}(\boldsymbol{0}, \tilde{\boldsymbol{R}})$, where

$$\tilde{\boldsymbol{R}} = \tilde{\boldsymbol{R}}^{\frac{1}{2}} \tilde{\boldsymbol{R}}^{\frac{1}{2}H} = \text{diag}\left(\boldsymbol{C}^H \boldsymbol{R}_0 \boldsymbol{C} \ \ldots \ \boldsymbol{C}^H \boldsymbol{R}_{L-1} \boldsymbol{C} \right). \tag{76}$$

Matrix $\tilde{\boldsymbol{R}}^{\frac{1}{2}} \boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{G}}^H \tilde{\boldsymbol{R}}^{\frac{1}{2}H}$ has a (pseudo) Wishart distribution.

Exact comparison of the performance of the low-complexity OFDM system with a usual system which performs N_r DFT's per block is difficult. We hereby give an asymptotic argument. For fixed L and N_t and as N_r gets large, it follows from the law of large numbers that $(1/N_r)\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{G}}^H \to \boldsymbol{I}_{LN_t}$. Hence, in the large N_r limit,

$$\sum_{k=0}^{N-1} \text{SNR}_k = \frac{N}{\sigma_n^2} N_r \lambda_1(\tilde{\boldsymbol{R}}). \tag{77}$$

In contrast, when the combining is performed on each subcarrier separately, then it can be shown that

$$\begin{aligned}\sum_{k=0}^{N-1} \mathrm{SNR}_k &= \frac{1}{\sigma_n^2} \sum_{k=0}^{N-1} \lambda_1(\boldsymbol{H}_k \boldsymbol{C}\boldsymbol{C}^H \boldsymbol{H}_k^H) \\ &\sim \frac{1}{\sigma_n^2} \sum_{k=0}^{N-1} \lambda_1(\boldsymbol{H}_{w,k} \boldsymbol{R}^{\frac{1}{2}} \boldsymbol{C}\boldsymbol{C}^H \boldsymbol{R}^{\frac{1}{2}} \boldsymbol{H}_{w,k}^H) \\ &= \frac{1}{\sigma_n^2} \sum_{k=0}^{N-1} \lambda_1(\boldsymbol{C}^H \boldsymbol{R}^{\frac{1}{2}} \boldsymbol{H}_{w,k}^H \boldsymbol{H}_{w,k} \boldsymbol{R}^{\frac{1}{2}} \boldsymbol{C}) \\ &\to \frac{N}{\sigma_n^2} N_r \lambda_1(\boldsymbol{C}^H \boldsymbol{R} \boldsymbol{C})\end{aligned} \tag{78}$$

as N_r gets large. To gain a rough idea of how these two cases compare with each other, consider when $\boldsymbol{R}_l = \boldsymbol{I}, l = 0, 1, \ldots, L-1$ and $\boldsymbol{C}$ has orthonormal columns. In this case, $\lambda_1(\tilde{\boldsymbol{R}}) = 1$, and $\lambda_1(\boldsymbol{C}^H \boldsymbol{R} \boldsymbol{C}) = L$. Thus, the low-complexity OFDM system has roughly an L-fold power penalty compared with the usual OFDM system, in the asymptote of a large number of receive antennas. However, the number of DFT operations per OFDM block of these two systems is 1 v.s. N_r, a considerable difference when N_r is large.

4.4 Numerical Results

The BER performance of the time domain combining and 2-D eigen-beamforming scheme is evaluated in an OFDM system with 64 subcarriers and a QPSK modulation scheme. The discrete-time channel is modeled as quasi-static fading and has two taps, spaced one OFDM sample apart. To show the effect of partial channel knowledge at the transmitter, we first compare with Alamouti STBC performed on each subcarrier, for the case $N_r = 1$ and $N_t = 2$. The BER performance is shown in Fig. 11. The two channel taps have a power ratio of 16:1, mean angles of departure of 0 and $3\pi/7$, and zero angle spread, giving a highly correlated channel. Here, the SNR is defined as $\mathrm{SNR} = P\mathrm{tr}(\boldsymbol{R})/(N_t\sigma_n^2)$, where P is the total transmitted power and tr() denotes the trace of a matrix. This quantity is equal to the ratio of the total received signal power to the total noise power given that the transmitted signal is i.i.d.. It is shown that the 2-D eigen-beamforming performs uniformly better than the Alamouti code, which does not use any knowledge of the channel.

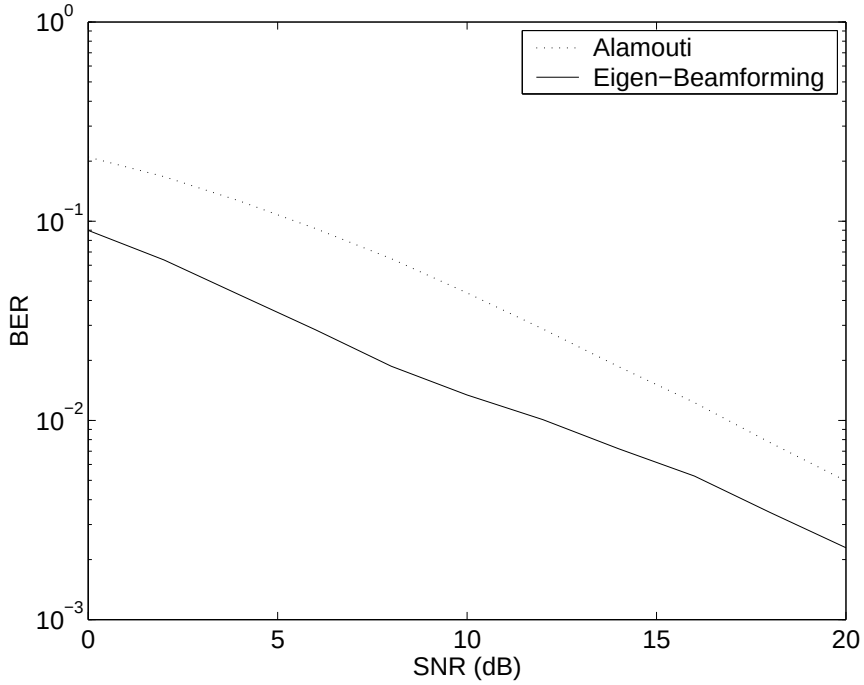


Figure 11: Comparison of 2-D eigen-beamforming with the Alamouti space-time block code.

In Fig. 12, we show the BER performance of time domain combining and 2-D beamforming for MIMO channels, and compare with the receiver-end frequency domain combining. The two channel taps have equal power, mean angles of departure of $-\pi/14$ and $3\pi/7$, and zero angle spread. We use repetition coding to collect multipath diversity, where each symbol is transmitted on two subcarriers, separated by half of the system bandwidth. Similar schemes are used in [33, 56]. As shown in Fig. 12, generally the diversity order increases as the number of transmit or receive antennas increases. Encouragingly, the diversity order achieved by time-domain combining is almost the same as frequency domain combining, although the frequency domain combining requires as many as N_r DFT operations per OFDM block.

4.5 Conclusion

In this chapter, we analyzed a low-complexity MIMO OFDM system where the transmitter has information of the channel correlation. In the low complexity system, the transmitter beamforming and receiver combining are performed in the time domain; thus, only one DFT

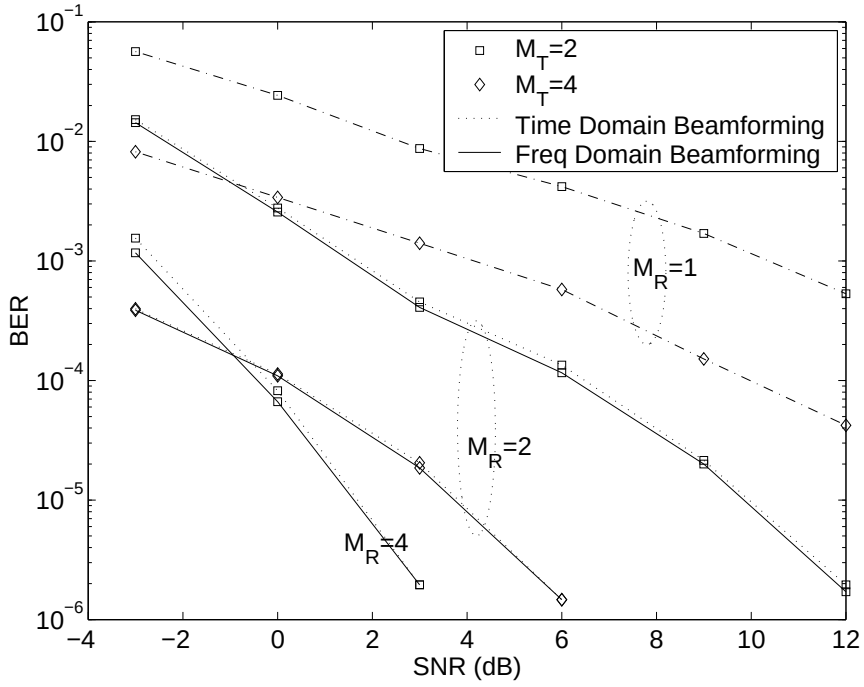


Figure 12: BER performance of time domain and frequency domain combining in MIMO channels

is required. For the MISO case, we show that the optimal 2-D transmit beamforming can be implemented using the low-complexity time domain beamforming. For the MIMO case, we formulate a suboptimal solution of beamforming and combining vectors. Asymptotic analysis shows that the low-complexity combining suffers a power loss. Numerical simulation of a repetition-coded OFDM system demonstrates that the diversity order increases as the number of transmit or receive antennas increases. The time domain combining is able to collect the majority part of the diversity order that the frequency domain combining schemes enjoy, yet with a much smaller complexity.

CHAPTER V

A BROADCAST APPROACH FOR MISO CHANNELS WITH UNCERTAINTY IN THE PARTIAL STATE INFORMATION

5.1 Introduction

Most previous papers on channel mean feedback have not considered uncertainties in the feedback process, such as unexpected delay or error in the feedback channel. Such uncertainties exist in reality and ignoring them results in suboptimal algorithms. In this chapter, we consider mean-feedback with an unknown delay. For channels where the CSI has uncertainty, the transmitter does not know how large an information rate can be reliably transmitted through the underlying channel. We are interested in a scheme that is able to adapt to the quality of the feedback. Ideally, the transmission scheme would work like beamforming, if the CSI is accurate, and would work like transmission without CSI, if the feedback quality is extremely bad. A broadcast approach is one way to deal with such uncertainty.

A broadcast approach for a single user can be developed from a multiuser broadcast channel [13], where a single transmission is directed to a number of receivers, each with possibly different channel conditions, reflected by their SNRs. For a degraded broadcast channel [13], such as a Gaussian broadcast channel with a single antenna at the transmitter, layered coding can be used to achieve the boundary of the capacity region. A receiver can decode all the layers corresponding to receivers having a worse SNR than itself and subtract them before decoding its own code layer [13, Sec. 14.1.3].

Cover [14] originally discussed broadcasting for the single user compound channel. A compound channel is a possibly infinite collection of channel transfer functions. The realization of a compound channel, or the index in the collection, is determined by nature

and is unknown to the transmitter [6, 14]. When the transmitter has no CSI, the fading channel can be viewed as a compound channel with the channel gain as a parameter of the compound channel. A broadcast approach for a single user where the transmitter has no CSI was first proposed by Shamai and Steiner [63]. In this broadcast approach, the receiver consists of a continuum of ordered users, each corresponding to a specific fading channel condition. The transmitter sends layered coded information. Each of the imaginary users at the receiver decodes a code layer if its channel realization allows. The maximum number of layers successively decoded is dictated by the fading channel realization. This single user broadcast strategy facilitates reliable transmission rates adapted to the actual channel conditions. Liu et al. [41] showed that two-layer superposition coding is adequate to achieve most of the throughput gain achievable by an infinite number of layers. All previous papers on broadcast approaches [41, 63, 70] are for channels where the transmitter has no CSI. We extend the broadcast approach to a mean-feedback MISO channel with uncertainties. Here, the unknown delay is the parameter of the compound channel.

In reality, the delay of the feedback channel is likely within a limited range. Our approach with the assumption that the delay is unknown serves as a lower bound in terms of achievable rate compared to the real life case.

The system model is given in Section 5.2, and the outage approach is shown in Section 5.3. Section 5.4 presents a two-layer coded broadcast approach. Numerical results on achievable rates and conclusions are given in Sections 5.5 and 5.6, respectively.

5.2 The Model

We assume a block fading model in which the fading is fixed during each transmission of a codeword. However, the channel may vary during the CSI feedback process. The complex baseband equivalent model of a flat fading MISO channel with a single receive antenna and N_t transmit antennas can be expressed as

$$y = \boldsymbol{h}^H \boldsymbol{x} + n,$$

where $\boldsymbol{h}$ is an $N_t \times 1$ channel transfer vector, $\boldsymbol{x}$ is an $N_t \times 1$ transmit vector, y is the received signal, and n is an additive noise sample distributed as $\mathcal{CN}(0, 1)$. We assume that

the constraint on the total transmitted power is $N_t P$, where P denotes the average SNR per transmit antenna. As shown by [74], for a given instantiation of $\boldsymbol{h}$, if the receiver knows the channel, among all input distributions with a given covariance matrix $P\boldsymbol{Q}$, where $\mathrm{tr}(\boldsymbol{Q}) = N_t$, the Gaussian vector $\boldsymbol{x} \sim \mathcal{CN}(\boldsymbol{0}, P\boldsymbol{Q})$ maximizes the mutual information between $\boldsymbol{x}$ and y, which is given by

$$I((y, \boldsymbol{h}); \boldsymbol{x}) = \log(1 + \boldsymbol{h}^H P\boldsymbol{Q}\boldsymbol{h}). \tag{79}$$

In the mean-feedback model, the transmitter knows that

$$\boldsymbol{h} \sim \mathcal{CN}(\boldsymbol{g}, \alpha^2 \boldsymbol{I}) \tag{80}$$

given the feedback $\boldsymbol{g}$. With this channel distribution and input covariance matrix $P\boldsymbol{Q}$, Eq. (64) of [52] gives

$$I((y, \boldsymbol{h}); \boldsymbol{x}) = \int_0^\infty \frac{e^{-y}}{y} \left[1 - \frac{\exp\left(-\boldsymbol{g}^H \left(\frac{\boldsymbol{Q}Py}{\boldsymbol{I} + \alpha^2 \boldsymbol{Q}Py} \right) \boldsymbol{g} \right)}{\det(\boldsymbol{I} + \alpha^2 \boldsymbol{Q}Py)} \right] dy. \tag{81}$$

Visotsky and Madhow [76] proved that to maximize the mutual information in (81), the optimal $\boldsymbol{Q}$ satisfies the following lemma.

Lemma 2. *The dominant eigenvector of $\boldsymbol{Q}$ points in the direction of $\boldsymbol{g}$. All non-dominant eigenvalues are equal to each other. Non-dominant eigenvectors are arbitrarily chosen, except for the restriction that all eigenvectors are orthonormal.*

Let the largest eigenvalue be q_0 and all other eigenvalues be equal to q, where $q_0 + (N_t - 1)q = N_t$, then (81) can be rewritten as [52]

$$I((y, \boldsymbol{h}); \boldsymbol{x}) = \int_0^\infty \frac{e^{-\frac{y}{\alpha^2 P}}}{y} \left[1 - \frac{\exp\left(-\frac{\gamma q_0 y}{1 + q_0 y} \right)}{(1 + q_0 y)(1 + qy)^{N_t - 1}} \right] dy, \tag{82}$$

where $\gamma = \frac{\|\boldsymbol{g}\|^2}{\alpha^2}$ is a measure of feedback quality. The optimal values of q_0 and q can be determined through numerical optimization.

Let $\boldsymbol{h}(t)$ be a first-order autoregressive random process with forgetting factor c, $\boldsymbol{h}(t) = c\boldsymbol{h}(t-1) + \boldsymbol{w}(t)$, where $\boldsymbol{w}(t)$ is an $N_t \times 1$ vector of i.i.d. circularly symmetric Gaussians, each of variance σ_w^2. The marginal distribution of $\boldsymbol{h}(t)$ is $\mathcal{CN}(\boldsymbol{0}, \alpha^2 \boldsymbol{I})$ for any t, where $\alpha^2 = \frac{\sigma_w^2}{1 - c^2}$.

At time t, the receiver feeds back $\boldsymbol{g} = \boldsymbol{h}(t)$ to the transmitter. Assuming a delay d in the feedback channel, Visotsky and Madhow [76] showed that conditioned on $\boldsymbol{g}$, the distribution of true channel vector $\boldsymbol{h}$ is given by

$$\boldsymbol{h} \sim \mathcal{CN}(\eta \boldsymbol{g}, (1-\eta^2)\alpha^2 \boldsymbol{I}), \tag{83}$$

where $\eta = c^d$. When the delay d is unknown, the transmitter may assume that η is uniformly distributed in $[0, 1]$ without loss of generality.

The validity of (83) is not limited to first-order autoregressive channels. Zhou et al. [85] showed that when the variation of each channel coefficient follows Clarke's two-dimensional isotropic scattering model [71], (83) also holds true. Although we formulate the uncertainty in the channel mean feedback problem as an unknown delay, the broadcast approach proposed in this chapter can be similarly applied to other types of uncertainties in channel mean feedback. For example, the variance α^2 in (80) can be used to model quantization errors [85]. Uncertainty in the variance of quantization errors arises when the law of channel variation changes while the quantization codebook is fixed at the receiver. In those cases, the description of uncertainty is not necessarily the same as (83).

We observe in (83) that CSI deteriorates from a deterministic $\boldsymbol{g}$ to $\mathcal{CN}(\boldsymbol{0}, \alpha^2 \boldsymbol{I})$ as d varies from 0 to infinity. If η is deemed as the parameter of a compound channel, a broadcast approach can be used to adapt to the particular realization of η without knowing it at the transmitter.

5.3 An Outage Approach

With an unknown delay, the most straightforward approach would be to choose a transmission rate and a corresponding outage probability such that the expected throughput is maximized. For each fixed η, the rate achievable by an optimal input covariance matrix $\boldsymbol{Q}$ satisfying Lemma 2 is given by

$$\begin{aligned} R'_O(\eta) =& E_{\boldsymbol{h} \sim \mathcal{CN}(\eta \boldsymbol{g}, (1-\eta^2)\alpha^2 \boldsymbol{I})} \log(1 + \boldsymbol{h}^H P \boldsymbol{Q} \boldsymbol{h}) \\ =& \int_0^\infty \frac{e^{-\frac{y}{(1-\eta^2)\alpha^2 P}}}{y} \left[1 - \frac{\exp\left(-\frac{\eta^2 \|\boldsymbol{g}\|^2 q_0 y}{\alpha^2 (1-\eta^2)(1+q_0 y)} \right)}{(1+q_0 y)(1+qy)^{N_t - 1}} \right] dy. \end{aligned} \tag{84}$$

The probability that this rate can be supported is $1-\eta$. The outage rate, defined as the expected throughput, is therefore $(1-\eta)R'_O(\eta)$ for a given transmit covariance matrix. To maximize the outage rate, we need to optimize jointly over η and q_0:

$$R_O = \underset{\eta, q_0}{\text{maximize}}\ (1-\eta)R'_O(\eta). \tag{85}$$

5.4 Two-Layer Broadcast Approach

We consider a two-layer coded broadcast approach with two imaginary users at the receiver. Although the broadcast channel in which different users are defined by different values of η in (83) is not necessarily a degraded one, predetermined ordering of the users can still be achieved because a code layer decodable with CSI of a certain quality can also be decoded at better CSI, and layered coding and decoding can still be applied.

Let user 1 be associated with lower-quality CSI than user 2. Each of the two users has to decode at a fractional rate. For some realization of η, only the first user can decode its fractional rate R_1. With CSI that has a better quality (larger η), the second user decodes initially the interference of rate R_1 and then subtracts the codeword for the first user from the received signal and decodes its own at rate R_2. Hence, the total achievable rate for a realization of η is either R_1 or R_1+R_2.

Let $P\boldsymbol{Q}_1$ and $P\boldsymbol{Q}_2$ be the covariance matrices of the two code layers, where the power constraint requires that $\mathrm{tr}(\boldsymbol{Q}_1+\boldsymbol{Q}_2)=N_t$. For the transmission of user 1 always to be able to get through the channel, user 1 has to assume CSI of $\mathcal{CN}(\boldsymbol{0},\alpha^2\boldsymbol{I})$. Since user 2 is always an interference to user 1, the achievable rate of user 1 is equal to

$$R_1 = E_{\boldsymbol{h}\sim\mathcal{CN}(\boldsymbol{0},\alpha^2\boldsymbol{I})}\{\log[1+P\boldsymbol{h}^H(\boldsymbol{Q}_1+\boldsymbol{Q}_2)\boldsymbol{h}] - \log(1+P\boldsymbol{h}^H\boldsymbol{Q}_2\boldsymbol{h})\}. \tag{86}$$

For a better channel condition $\boldsymbol{h}\sim\mathcal{CN}(\eta\boldsymbol{g},(1-\eta^2)\alpha^2\boldsymbol{I})$, the achievable rate of user 2 is equal to

$$R_2 = E_{\boldsymbol{h}\sim\mathcal{CN}(\eta\boldsymbol{g},(1-\eta^2)\alpha^2\boldsymbol{I})}\log(1+P\boldsymbol{h}^H\boldsymbol{Q}_2\boldsymbol{h}). \tag{87}$$

The probability that only R_1 can get through the channel is η. The probability that R_2 can also get through the channel is $1-\eta$. The expected throughput is $\eta R_1+(1-\eta)(R_1+R_2) = R_1+(1-\eta)R_2$.

Table 2: Covariance matrices for two code layers

	$\boldsymbol{Q}_1$	$\boldsymbol{Q}_2$
WW	$\beta\boldsymbol{I}$	$\xi\boldsymbol{I}$
WB	$\beta\boldsymbol{I}$	$N_t\xi\boldsymbol{g}\boldsymbol{g}^H/\|\boldsymbol{g}\|^2$
BB	$N_t\beta\boldsymbol{g}\boldsymbol{g}^H/\|\boldsymbol{g}\|^2$	$N_t\xi\boldsymbol{g}\boldsymbol{g}^H/\|\boldsymbol{g}\|^2$

We have to design the optimal covariance matrices for the two layers to maximize the throughput. Numerical examples in [76] showed that the rate achieved by the optimal covariance matrix in mean-feedback is either very close to that achieved by beamforming or very close to that achieved by a white covariance matrix that is proportional to an identity matrix, depending on the quality of CSI. Based on this observation, we take a suboptimal approach by permitting the covariance matrices of both layers to be either beamforming or white. Thus, there are four combinations in total. We denote the case where both layers use white covariance matrices as WW, the case where user 1 uses a white covariance matrix while user 2 uses beamforming as WB, the case where both users use beamforming as BB. The remaining case where user 1 uses beamforming and user 2 uses a white covariance matrix is not considered because user 2 always enjoys CSI of a better quality than user 1. Confined to the above three cases, we only need to optimize the power allocation between the two code layers. Let $\mathrm{tr}(\boldsymbol{Q}_1) = N_t\beta$, and $\mathrm{tr}(\boldsymbol{Q}_2) = N_t\xi$, then we have $\beta + \xi = 1$. When $\boldsymbol{Q}_1$ is white, $\boldsymbol{Q}_1 = \beta\boldsymbol{I}$; when user 1 uses beamforming, $\boldsymbol{Q}_1 = N_t\beta\boldsymbol{g}\boldsymbol{g}^H/\|\boldsymbol{g}\|^2$. Similarly, $\boldsymbol{Q}_2 = \xi\boldsymbol{I}$ when it is white and $\boldsymbol{Q}_2 = N_t\xi\boldsymbol{g}\boldsymbol{g}^H/\|\boldsymbol{g}\|^2$ when it is beamforming. The covariance matrices for the three cases are summarized in Table 2. For each of WW, WB, and BB, the throughput can be maximized by jointly optimizing η and β:

$$\underset{\beta,\eta}{\text{maximize}}\ R_1 + (1-\eta)R_2. \tag{88}$$

Finally, a maximum can be chosen among the three cases. Denote the achievable rates of each of the three cases as R_{WW}, R_{WB}, and R_{BB}, respectively. The analytic expressions of (88) for each of the three cases are given in Appendix B.

5.5 Numerical Results

From (82), we see that the mutual information is a function of $\gamma = \|\boldsymbol{g}\|^2/\alpha^2$ and $\alpha^2 P$. Without loss of generality, we set $\alpha = 1$. We also define P as SNR. We performed numerical optimizations (85) for the outage approach and (88) for each of the three cases of broadcast approach in a range of γ and SNR. Since these are non-convex problems, standard line search algorithms with multiple random initial points were utilized.

Experiments showed that WW is better than WB when γ is small. BB is better than WB when γ is large and the SNR is small. But the maximum value of $R_{WW} - R_{WB}$ is merely 1.7764×10^{-15} bits/channel use, and the maximum of $R_{BB} - R_{WB}$ is just 1.3015×10^{-5} bits/channel use, which are negligible. Therefore, WB is optimal among the three two-layer coding schemes across almost all of the simulated ranges of γ and SNR.

The achievable rates of outage approach and the two-layer broadcast approach WB are shown in Fig. 13 for the case $\gamma = 21$ dB. An SNR gain of more than 6 dB is evident in the high SNR region. As a holistic picture, Fig. 14 shows the difference $R_{WB} - R_O$ for a range of γ and SNR. The optimized values of R_{WB}, β, and η for WB are shown in Fig. 15. The optimized values of R_O, q_0, and η for the outage approach are shown in Fig. 16. From Fig. 16 (b), we see that the optimal covariance matrix for the outage approach is either beamforming ($q_0 = 4$) or white ($q_0 = q = 1$), which further validates our choice of covariance matrices for the broadcast approach in Section 5.4.

5.6 Conclusion

We studied increasing the throughput of the mean-feedback MISO channel with an unknown delay in the feedback channel. We used a two-layer broadcast approach where the first layer has a white covariance matrix and the second layer performs beamforming. Significant gain was observed relative to the outage approach in some regions of feedback quality and SNR.

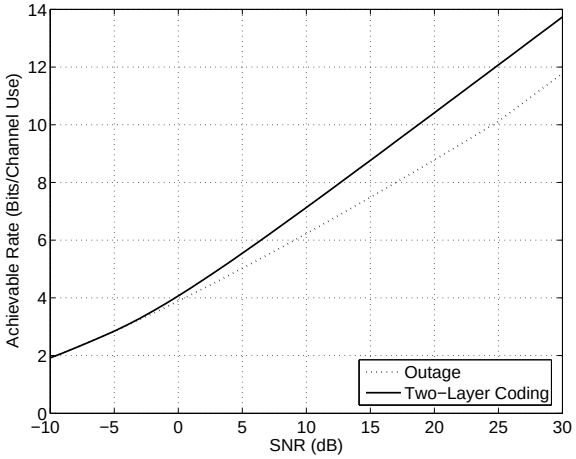


Figure 13: Achievable rates for the outage approach and two-layer coding when $\gamma = 21$ dB.

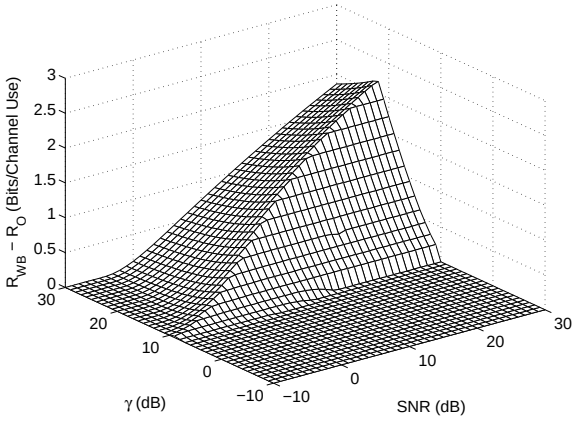
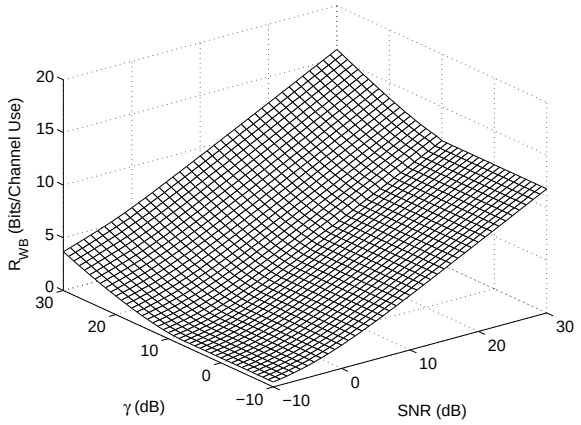
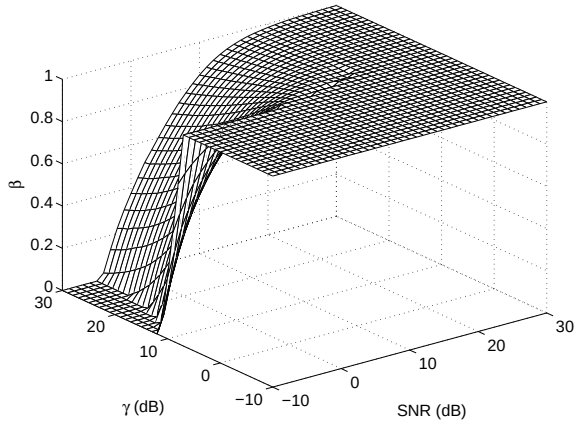


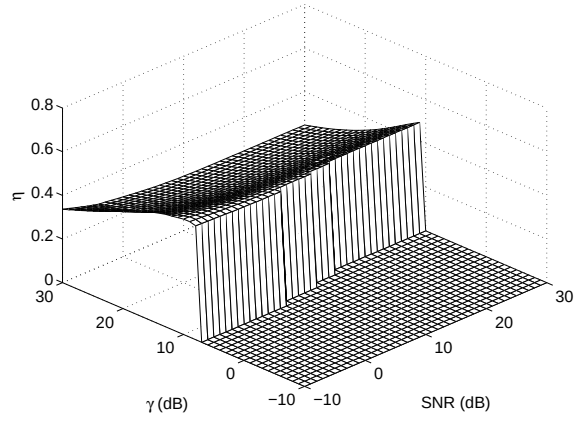
Figure 14: The difference between the achievable rates of two-layer coding and the outage approach.



(a)

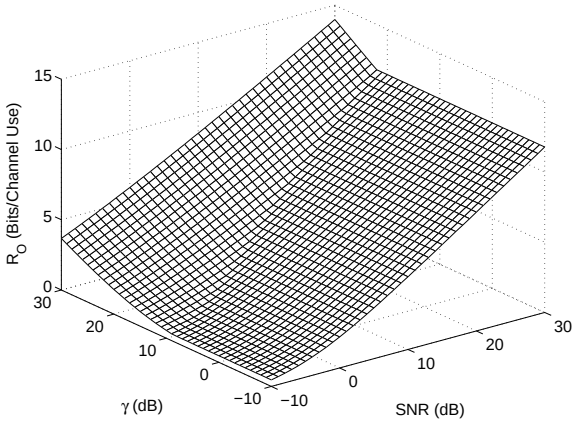


(b)

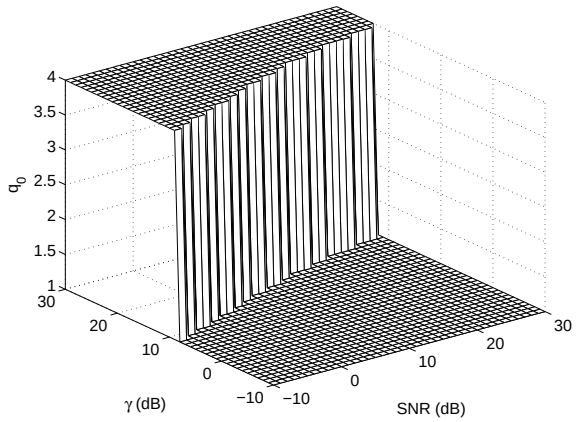


(c)

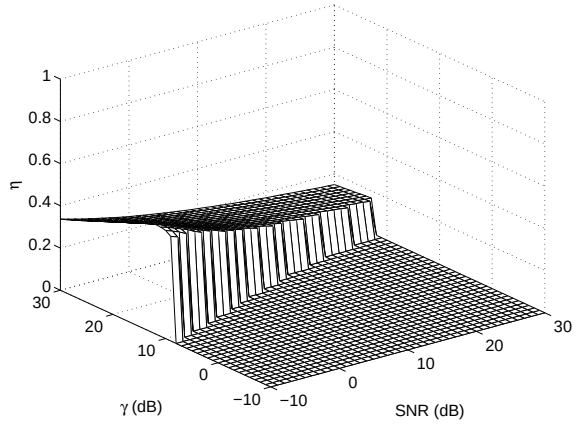
Figure 15: Optimization results for the two-layer broadcast approach where the first layer uses a white covariance matrix and the second layer uses beamforming. (a) R_{WB}; (b) Optimal β versus γ and SNR; (c) Optimal η versus γ and SNR.



(a)



(b)



(c)

Figure 16: Optimization results for the outage approach. (a) R_O; (b) Optimal q_0 versus γ and SNR; (c) Optimal η versus γ and SNR.

CHAPTER VI

DATA-DEPENDENT SUPERIMPOSED TRAINING FOR NONCOHERENT CHANNELS

6.1 Introduction

In systems considered in previous chapters, the receivers are assumed to have perfect CSI or accurate estimates of CSI. This situation is only part of the panorama of high speed mobile communications, because such an assumption is only reasonable when the channel fading is slow and the coherence time is long relative to the symbol duration. On the other hand, reliable digital communication over fast time-varying wireless channels is becoming more and more important in view of the rapid growth of cellular and personal communications systems. Such systems are called noncoherent systems because the transmitter and the receiver do not have channel state information.

Information theoretically optimal noncoherent coding for block fading channels was considered in [31, 82]. At high SNR, the optimal codewords were shown to be so-called *unitary space-time codes* and, therefore, are unstructured. At the receiver, decoding can only be done through an exhaustive search through the codebook. Considering decoding complexity, training based schemes are attractive. Besides, a training based scheme has been shown to be superior to unitary space-time codes in certain situations [16].

If the coherence time in a block-fading channel is very long, TDM training together with Gaussian codes are asymptotically (in coherence time) optimal. Therefore, superimposed training is more of interest in fast fading channels. For a continuously time-varying fast fading channel, Dong and Tong [18] showed that, if *linear minimum mean-square error* (LMMSE) channel estimators are used at the receiver, superimposed training results in better channel estimates than TDM training, although the power allocation between training symbols and data symbols was not optimized there. In fast fading channels, lots of training

symbols are required to get an accurate estimate of the channel. A TDM training system has to insert frequent training symbols and leaves few time slots for data transmission. On the other hand, superimposed training may be suitable for fast fading channels, because every symbol contains a training component. For superimposed training systems, despite the universal presence of training symbols, it is imaginable that channel estimation accuracy is harmed by the data symbols because the data symbols act as high-variance noise for the channel estimator.

In this chapter, we improve the performance of superimposed training systems for fast fading channels by proposing a data-dependent[1] superimposed training scheme. The transmitter is equipped with multiple candidate training sequences. Given a block of data, the training sequence that minimizes the channel estimation error is selected and added together with the data sequence to form the transmitted codeword. The channel estimation error is expected to be reduced compared to traditional superimposed training because of this dynamic selection of training sequences, assuming the receiver knows which training sequence is used at the transmitter. To reduce the probability that the receiver makes a wrong decision regarding which training sequence is actually used by the transmitter, the set of training sequences are designed to minimize the PEP between the transmitted codewords consisting of combined data and training. We show through simulations that, if the receiver employs a close-to-optimal detector, the data-dependent superimposed training scheme has better performance than TDM training and traditional superimposed training.

The remainder of this chapter is organized as follows. The channel model and traditional training schemes are outlined in Section 6.2. We discuss the design of multiple training sequences in Section 6.3. Design examples and performance comparison between our proposed scheme and traditional training schemes are given in Section 6.4. Finally, Section 6.5 draws conclusions and provides some discussion.

[1]The notion of data-dependent training was first used by the authors of [25, 26]. Despite the notion, the nature of that scheme is fixed FDM training for frequency-selective channels.

6.2 Channel Models and Training Schemes

6.2.1 Channel Models

Without loss of generality, we consider a system with a single receive antenna. The scalar signal received by the receiver at time t is modeled as

$$y(t) = \boldsymbol{x}(t)\boldsymbol{h} + n(t),\ t = 0, 1, \ldots, N-1, \tag{89}$$

where N is the number of time samples of a space-time codeword, $\boldsymbol{h} = [\, h_0\ h_1\ \ldots\ h_{N_t-1}\,]^T$ is a column vector of channel coefficients, and $\boldsymbol{x}(t)$ is a row vector of transmitted signal values. The noise samples $n(t)$ are assumed to be i.i.d. *additive white Gaussian noise* (AWGN) distributed as $\mathcal{CN}(0, \sigma_n^2)$. The channel coefficients are assumed constant over the transmission of one codeword, but are allowed to vary independently between blocks. We assume that the channel is i.i.d. Rayleigh fading, and each coefficient h_k, $k = 0,\ 1,\ \ldots,\ N_t - 1$ is distributed as $\mathcal{CN}(0, 1)$. Moreover, we assume that each signal entry in $\boldsymbol{x}(t)$ has average power of P and define the channel SNR as

$$\text{SNR} = \frac{P}{\sigma_n^2}. \tag{90}$$

Stacking all N samples into a single column vector, the above set of equations can be written as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{h} + \boldsymbol{n}, \tag{91}$$

where

$$\boldsymbol{y} = [\, y(0)\ y(1)\ \ldots\ y(N-1)\,]^T \tag{92}$$

$$\boldsymbol{X} = [\, \boldsymbol{x}^T(0)\ \boldsymbol{x}^T(1)\ \ldots\ \boldsymbol{x}^T(N-1)\,]^T \tag{93}$$

$$\boldsymbol{n} = [\, n(0)\ n(1)\ \ldots\ n(N-1)\,]. \tag{94}$$

Matrix $\boldsymbol{X}$ is called a codeword transmitted during a channel coherence interval.

Both flat fading and frequency-selective fading channels with a single transmit antenna can be described by (89). For a flat fading channel with a single transmit antenna, the channel vector $\boldsymbol{h}$ degrades to a scalar h, and signal vector $\boldsymbol{x}(t)$ degrades to a scalar signal

$x(t)$. Therefore, the channel model (91) becomes

$$\boldsymbol{y} = \boldsymbol{x}h + \boldsymbol{n}, \tag{95}$$

where $\boldsymbol{x} = [\, x(0)\ x(1)\ \ldots\ x(N-1)\,]^T$.

In the case of a frequency-selective fading channel with a single transmit antenna, $\{h_0, h_1, \ldots, h_{N_t-1}\}$ are the channel coefficients of a tapped-delay-line model. Assuming that there is a protection gap or trailing zeros between the transmission of codewords, inter-block interference can be ignored. For a frequency-selective channel with a single transmit antenna and trailing zeros, the codeword matrix $\boldsymbol{X}$ has the following form:

$$\boldsymbol{X} = \begin{pmatrix} x(0) & 0 & 0 & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ x(N-N_t) & \ldots & x(0) & 0 & \ldots & 0 \\ 0 & x(N-N_t) & \ldots & x(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & x(N-N_t) & \ldots & x(0) \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & x(N-N_t) \end{pmatrix}. \tag{96}$$

6.2.2 Training Schemes

When the receiver does not have knowledge of the channel, training based schemes can be used in which pilot symbols known by the receiver are transmitted to help with channel estimation. In TDM training schemes, training symbols are periodically inserted between data symbols [18]. Without loss of generality, we consider fast fading where only one training symbol can be put into each codeword. Then, in the flat fading case, a codeword has the form

$$\boldsymbol{x} = [\, x_\tau\ x_d(0)\ x_d(1)\ \ldots\ x_d(N-2)\,]^T, \tag{97}$$

where x_τ is the training symbol, and $\{x_d(k), k = 0, 1, \ldots, N-2\}$ are data symbols. Since we assume a pseudo-static block fading channel, the position of the training symbol inside the codeword does not matter. For frequency-selective channels, to avoid *intersymbol*

interference (ISI) between training symbols and data symbols, a strict-sense TDM training scheme has to put trailing zeros after the training symbol [1], resulting in the codeword

$$\boldsymbol{X} = \begin{pmatrix} x_\tau & 0 & 0 & 0 & \dots & 0 \\ 0 & x_\tau & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 & x_\tau \\ x_d(0) & 0 & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ x_d(N-2N_t) & \dots & x_d(0) & 0 & \dots & 0 \\ 0 & x_d(N-2N_t) & \dots & x_d(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & x_d(N-2N_t) & \dots & x_d(0) \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 & x_d(N-2N_t) \end{pmatrix}. \tag{98}$$

Clearly, only $N-2N_t+1$ data symbols can be transmitted in TDM training because of the trailing zeros after the training symbol.

For superimposed training, each symbol in the codeword is the summation of a training component and a signal component:

$$x(t) = x_\tau(t) + x_d(t). \tag{99}$$

The vector channel model (91) can be written as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{h} + \boldsymbol{n} = (\boldsymbol{X}_\tau + \boldsymbol{X}_d)\boldsymbol{h} + \boldsymbol{n}, \tag{100}$$

where $\boldsymbol{X}_\tau$ and $\boldsymbol{X}_d$ are comprised of $x_\tau(t)$ and $x_d(t)$, respectively, formed in the same way as $\boldsymbol{X}$ is formed.

For superimposed training, the average power per training symbol is defined as

$$P_\tau = E|x_\tau(t)|^2. \tag{101}$$

Similarly, the average power per data symbol is

$$P_d = E|x_d(t)|^2. \tag{102}$$

Naturally, the total transmit power per symbol is $P = P_\tau + P_d$.

For our data-dependent superimposed training, the selection of training sequences is dependent upon data sequences. Therefore, strictly speaking, the training sequence and the data sequence are not statistically independent. However, as the number of training sequences is small compared to the data sequences, and each training sequence is still associated with a large number of data sequences, statistical independency holds approximately, and so does the following relationship:

$$E|x(t)|^2 = P_\tau + P_d. \tag{103}$$

The parameter P_τ/P_d and the set of training sequences are to be designed for the data-dependent superimposed training scheme. For TDM training, the power of the training symbol is defined as

$$P_\tau = |x_\tau|^2, \tag{104}$$

and is the only parameter that needs to be optimized. Note that the transmit power of TDM training in an ISI channel should be properly scaled to account for the effect of trailing zeros.

6.2.3 Detectors

We assume that the channel coefficients are unknown both at the transmitter and at the receiver. However, when dealing with ISI channels, we assume that the channel span N_t is known or correctly upper-bounded. In this chapter, we consider an uncoded system and assume that the data symbols $\{x_d(t)\}$ are chosen i.i.d. from a *pulse amplitude modulation* (PAM) or QAM constellation. Equivalently, we can think of the data-dependent superimposed training system as a rate-1 coding without adding any redundancy.

Without any channel state information at the receiver, conditioned on the transmitted signal matrix $\boldsymbol{X}$, the received signal $\boldsymbol{y}$ is distributed as complex Gaussian with zero-mean and covariance matrix $\boldsymbol{\Lambda} = \boldsymbol{X}\boldsymbol{X}^H + \sigma_n^2\boldsymbol{I}$. Therefore, the *probability density function* (pdf) of received signal $\boldsymbol{y}$ conditioned on $\boldsymbol{X}$ is given by [46]

$$p(\boldsymbol{y}|\boldsymbol{X}) = \frac{\exp(-\boldsymbol{y}^H\boldsymbol{\Lambda}^{-1}\boldsymbol{y})}{\pi^N \det(\boldsymbol{\Lambda})}. \tag{105}$$

Let the size of the data symbol constellation be Γ, and the number of codewords in the codebook be K. In the example of the flat fading channel (95), $K = \Gamma^N$. In the remainder of this section we use the general channel model (91). Denote the codewords as $\{\boldsymbol{X}_j\}$, $j \in \mathcal{J}$, where $\mathcal{J} = \{0, 1, \ldots, K-1\}$ is the index set of the codewords. Denote the decision on the codeword at the receiver as $\hat{\boldsymbol{X}}$.

For the channel in (91), the maximum likelihood (ML) decoder at the receiver makes a decision on the index of the transmitted codeword by

$$\hat{j} = \arg\max_{j \in \mathcal{J}} p(\boldsymbol{y}|\boldsymbol{X}_j), \tag{106}$$

which requires knowledge of statistics of the channel coefficients and also the noise variance σ_n^2.

Another well-known receiver that does not require knowledge of channel statistics is the *generalized likelihood ratio test* (GLRT) receiver [10, 66]. This receiver maximizes the likelihood of the received signal conditioned on the fading coefficients $\boldsymbol{h}$ and transmitted signal $\boldsymbol{X}_j$, first over the fading coefficients and subsequently over the transmitted signal, yielding

$$\hat{j} = \arg\max_{j \in \mathcal{J}} \boldsymbol{y}^H \boldsymbol{X}_j (\boldsymbol{X}_j^H \boldsymbol{X}_j)^{-1} \boldsymbol{X}_j^H \boldsymbol{y}. \tag{107}$$

Since $\boldsymbol{X}_j(\boldsymbol{X}_j^H \boldsymbol{X}_j)^{-1}\boldsymbol{X}_j^H$ is the projection matrix onto the jth signal's subspace, the detector compares the energy of the received signal in the signals' subspaces. The GLRT decoder is known to be simpler than the ML decoder and suffers a slight performance degradation compared to the ML decoder.

A simpler yet suboptimal decoder for the training based schemes is to first obtain an estimate of the channel $\hat{\boldsymbol{h}}$ based on the received signal and pilot symbols, and then perform coherent detection assuming that $\hat{\boldsymbol{h}}$ is the true channel. If the noise is AWGN, this detector is equivalent to the following minimum distance detector

$$\hat{j} = \arg\min_{j \in \mathcal{J}} \|\boldsymbol{y} - \boldsymbol{X}_j \hat{\boldsymbol{h}}\|_F. \tag{108}$$

The purpose of training symbols is to help the receiver obtain channel state information. Therefore, for training based schemes, the coherent detector (108) following channel

estimation seems to be the only reasonable thing to do. However, we can also think of the training scheme as an inner code, and usually there is another outer code in a communication system. The code might be iteratively decoded starting from the decisions given by (108). During the decoding process, the channel can also be iteratively estimated using hard decisions [15, 39, 55] or soft decisions [57, 81] on the data symbols. If the outer code is properly designed, the iterative decoding can approach the performance of an ML decoder. Based on these considerations, we consider the performance of ML decoders when designing the inner training codes.

6.3 Design of Data-Dependent Training Sequences

6.3.1 Channel Estimation and Training Sequence Selection

6.3.1.1 Flat Fading

We first consider the flat fading channel (95). We next find an LMMSE estimate of channel h based on $\boldsymbol{y}$ and $\boldsymbol{x}_\tau$. Denoting the channel estimation filter (vector) as $\boldsymbol{c}$, the linear channel estimate is given by $\hat{h} = \boldsymbol{cy}$. The LMMSE criterion minimizes

$$E|h - \hat{h}|^2 = E|h - \boldsymbol{cy}|^2. \tag{109}$$

It follows straightforwardly that $\boldsymbol{c} = E\{h\boldsymbol{y}^H\}E\{\boldsymbol{y}\boldsymbol{y}^H\}^{-1}$. Assume that the data components $\{x_d(t)\}$ are i.i.d. with zero mean. From (100), we have $E\{h\boldsymbol{y}^H\} = E\{hh^*\}\boldsymbol{x}_\tau^H = \boldsymbol{x}_\tau^H$, and

$$\begin{aligned} E\{\boldsymbol{y}\boldsymbol{y}^H\} &= E\{(\boldsymbol{x}_\tau + \boldsymbol{x}_d)hh^*(\boldsymbol{x}_\tau + \boldsymbol{x}_d)^H\} + \sigma_n^2\boldsymbol{I} \\ &= E_{\boldsymbol{x}_d}E_{h|\boldsymbol{x}_d}\{(\boldsymbol{x}_\tau + \boldsymbol{x}_d)hh^*(\boldsymbol{x}_\tau + \boldsymbol{x}_d)^H\} + \sigma_n^2\boldsymbol{I} \\ &= E_{\boldsymbol{x}_d}\{(\boldsymbol{x}_\tau + \boldsymbol{x}_d)(\boldsymbol{x}_\tau + \boldsymbol{x}_d)^H\} + \sigma_n^2\boldsymbol{I} \\ &= \boldsymbol{x}_\tau\boldsymbol{x}_\tau^H + (P_d + \sigma_n^2)\boldsymbol{I}. \end{aligned} \tag{110}$$

It follows that

$$\boldsymbol{c} = \boldsymbol{x}_\tau^H \left[\boldsymbol{x}_\tau \boldsymbol{x}_\tau^H + (P_d + \sigma_n^2)\boldsymbol{I}\right]^{-1} \tag{111}$$

$$= \frac{1}{P_d + \sigma_n^2} \boldsymbol{x}_\tau^H [\boldsymbol{I} - \boldsymbol{x}_\tau (P_d + \sigma_n^2 + \boldsymbol{x}_\tau^H \boldsymbol{x}_\tau)^{-1} \boldsymbol{x}_\tau^H] \tag{112}$$

$$= \frac{1}{P_d + \sigma_n^2} \boldsymbol{x}_\tau^H (\boldsymbol{I} - \frac{1}{P_d + \sigma_n^2 + P_\tau N} \boldsymbol{x}_\tau \boldsymbol{x}_\tau^H) \tag{113}$$

$$= \frac{1}{P_d + \sigma_n^2} (\boldsymbol{x}_\tau^H - \frac{\boldsymbol{x}_\tau^H \boldsymbol{x}_\tau}{P_d + \sigma_n^2 + P_\tau N} \boldsymbol{x}_\tau^H) \tag{114}$$

$$= \frac{1}{P_d + \sigma_n^2 + P_\tau N} \boldsymbol{x}_\tau^H, \tag{115}$$

where (112) follows from the *Sherman-Morrison-Woodbury formula* [28, Eq. (2.1.4)], and (113) and (115) follow from the fact that $\boldsymbol{x}_\tau^H \boldsymbol{x}_\tau = P_\tau N$.

To clearly see the effect of data sequence $\boldsymbol{x}_d$ on the channel estimation error, we need to evaluate the channel estimation error deeming the data sequence as deterministic. The expectation of the squared channel estimation error with respect to channel h and noise $\boldsymbol{n}$ is given by

$$\begin{aligned}
&E_{h,\boldsymbol{n}}\{(\boldsymbol{c}\boldsymbol{y} - h)(\boldsymbol{c}\boldsymbol{y} - h)^H\} \\
=&E_{h,\boldsymbol{n}}\{[\boldsymbol{c}(\boldsymbol{x}_\tau + \boldsymbol{x}_d) - 1]h + \boldsymbol{c}\boldsymbol{n}\}\{[\boldsymbol{c}(\boldsymbol{x}_\tau + \boldsymbol{x}_d) - 1]h + \boldsymbol{c}\boldsymbol{n}\}^H \\
=&[\boldsymbol{c}(\boldsymbol{x}_\tau + \boldsymbol{x}_d) - 1][\boldsymbol{c}(\boldsymbol{x}_\tau + \boldsymbol{x}_d) - 1]^H + \sigma_n^2 \boldsymbol{c}\boldsymbol{c}^H \\
=&\left[\frac{1}{P_d + \sigma_n^2 + P_\tau N} \boldsymbol{x}_\tau^H (\boldsymbol{x}_\tau + \boldsymbol{x}_d) - 1\right] \left[\frac{1}{P_d + \sigma_n^2 + P_\tau N} \boldsymbol{x}_\tau^H (\boldsymbol{x}_\tau + \boldsymbol{x}_d) - 1\right]^H \\
&+ \sigma_n^2 \frac{1}{(P_d + \sigma_n^2 + P_\tau N)^2} \boldsymbol{x}_\tau^H \boldsymbol{x}_\tau \\
=&\left[\frac{\boldsymbol{x}_\tau^H \boldsymbol{x}_d}{P_d + \sigma_n^2 + P_\tau N} - \frac{P_d + \sigma_n^2}{P_d + \sigma_n^2 + P_\tau N}\right] \left[\frac{\boldsymbol{x}_\tau^H \boldsymbol{x}_d}{P_d + \sigma_n^2 + P_\tau N} - \frac{P_d + \sigma_n^2}{P_d + \sigma_n^2 + P_\tau N}\right]^H \\
&+ \frac{\sigma_n^2 P_\tau N}{(P_d + \sigma_n^2 + P_\tau N)^2}.
\end{aligned} \tag{116}$$

To gain more insight into the relationship between data sequences and channel estimation MMSE, we consider an asymptotically long codeword. From (116), we can show that

$$\lim_{N \to \infty} E_{h,\boldsymbol{n}}\{(\boldsymbol{c}\boldsymbol{y} - h)(\boldsymbol{c}\boldsymbol{y} - h)^H\} = \frac{\|\boldsymbol{x}_\tau^H \boldsymbol{x}_d\|_F^2}{(P_\tau N)^2}. \tag{117}$$

This result shows that as the codeword length increases, the squared inner product between the training sequence and the data sequence is the dominant term in the channel estimation

error. Therefore, for the purpose of helping the receiver obtain a better channel estimate, a transmitter can be supplied with multiple candidate training sequences. Given a sequence of data symbols, the transmitter should select a training sequence that is most orthogonal to the data sequence. Although this orthogonality criterion is only asymptotically accurate at long codeword lengths, we adopt this criterion even if the codeword is short for simplicity reasons.

To describe the training sequence selection procedure, let M be the number of training sequences available at the transmitter. Denote the set of training sequences as

$$\{\boldsymbol{x}_{\tau 1}, \boldsymbol{x}_{\tau 2}, \ldots, \boldsymbol{x}_{\tau M}\}.$$

Given a data sequence $\boldsymbol{x}_d$, the transmitter chooses the mth training sequence if

$$m = \arg\min_{l} \|\boldsymbol{x}_{\tau l}^H \boldsymbol{x}_d\|_F^2. \tag{118}$$

The mapping from a data sequence to the final transmitted codeword is defined as

$$\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{x}_d) = \boldsymbol{x}_{\tau m} + \boldsymbol{x}_d, \text{ where } m \text{ is given by (118).} \tag{119}$$

6.3.1.2 Frequency-Selective Fading

For a general superimposed training scheme, the training part of the codeword $\boldsymbol{X}_\tau$ has exactly the same form as $\boldsymbol{X}$ in (96). However, it is hard to obtain any insight and a simple matching rule between the data and training sequences. Moreover, as will be shown in Section 6.4.3, our effort at optimizing the single training sequence to minimize the PEP resulted in a training sequence that is nonzero only in the first entry. Motivated by these factors, we apply the constraint that the training symbols be free from ISI caused by other training symbols, i.e., $N_t - 1$ trailing zeros be inserted between two training symbols. Under this constraint, $\boldsymbol{X}_\tau$ will be in the form of stacked identity matrices, each of which is scaled

by a training symbol. For example,

$$\boldsymbol{X}_\tau = \begin{pmatrix} x_\tau(1) & 0 & 0 \\ 0 & x_\tau(1) & 0 \\ 0 & 0 & x_\tau(1) \\ x_\tau(2) & 0 & 0 \\ 0 & x_\tau(2) & 0 \\ 0 & 0 & x_\tau(2) \\ \vdots & \vdots & \vdots \end{pmatrix}, \tag{120}$$

if $N_t = 3$.

Similar to the flat fading case, we design an LMMSE channel estimator $\boldsymbol{C}$ to minimize the channel estimation error

$$E\|\boldsymbol{h} - \boldsymbol{C}\boldsymbol{y}\|_F^2. \tag{121}$$

The optimal $\boldsymbol{C}$ can be shown to be

$$\boldsymbol{C} = \boldsymbol{X}_\tau^H \left(\boldsymbol{X}_\tau \boldsymbol{X}_\tau^H + E\{\boldsymbol{X}_d \boldsymbol{X}_d^H\} + \sigma_n^2 \boldsymbol{I}\right)^{-1}. \tag{122}$$

The data covariance matrix $E\{\boldsymbol{X}_d \boldsymbol{X}_d^H\}$ is a diagonal matrix but is not proportional to an identity matrix because of the Toeplitz structure of $\boldsymbol{X}_d$. However, if the codeword is long enough, the effect of boundary components on the diagonal can be ignored, and $E\{\boldsymbol{X}_d \boldsymbol{X}_d^H\} \cong N_t P_d \boldsymbol{I}$. Under this simplification,

$$\begin{aligned} \boldsymbol{C} &= \boldsymbol{X}_\tau^H \left(\boldsymbol{X}_\tau \boldsymbol{X}_\tau^H + (N_t P_d + \sigma_n^2)\boldsymbol{I}\right)^{-1} \\ &= \frac{1}{N_t P_d + \sigma_n^2} \boldsymbol{X}_\tau^H \left(\boldsymbol{I} - \frac{1}{N_t P_d + \sigma_n^2 + P_\tau N/N_t} \boldsymbol{X}_\tau \boldsymbol{X}_\tau^H\right) \\ &= \frac{1}{N_t P_d + \sigma_n^2 + P_\tau N/N_t} \boldsymbol{X}_\tau^H, \end{aligned} \tag{123}$$

where we used [28, Eq. (2.1.4)], and repeatedly used the fact that $\boldsymbol{X}_\tau^H \boldsymbol{X}_\tau = P_\tau N/N_t \boldsymbol{I}$, which holds if N is an integer multiple of N_t.

Now we can examine the effect of the data sequence on the channel estimation error

deeming the data sequence as deterministic. The covariance matrix of the channel estimation error vector is given by

$$\begin{aligned}
&E(\boldsymbol{C}\boldsymbol{y}-\boldsymbol{h})(\boldsymbol{C}\boldsymbol{y}-\boldsymbol{h})^H \\
=&E\{[\boldsymbol{C}(\boldsymbol{X}_\tau+\boldsymbol{X}_d)-\boldsymbol{I}]\boldsymbol{h}+\boldsymbol{C}\boldsymbol{n}\}\{[\boldsymbol{C}(\boldsymbol{X}_\tau+\boldsymbol{X}_d)-\boldsymbol{I}]\boldsymbol{h}+\boldsymbol{C}\boldsymbol{n}\}^H \\
=&[\boldsymbol{C}(\boldsymbol{X}_\tau+\boldsymbol{X}_d)-\boldsymbol{I}][\boldsymbol{C}(\boldsymbol{X}_\tau+\boldsymbol{X}_d)-\boldsymbol{I}]^H+\sigma_n^2\boldsymbol{C}\boldsymbol{C}^H \\
=&\left(\frac{1}{N_tP_d+\sigma_n^2+P_\tau N/N_t}\boldsymbol{X}_\tau^H\boldsymbol{X}_d-\frac{N_tP_d+\sigma_n^2}{N_tP_d+\sigma_n^2+P_\tau N/N_t}\boldsymbol{I}\right) \\
&\left(\frac{1}{N_tP_d+\sigma_n^2+P_\tau N/N_t}\boldsymbol{X}_\tau^H\boldsymbol{X}_d-\frac{N_tP_d+\sigma_n^2}{N_tP_d+\sigma_n^2+P_\tau N/N_t}\boldsymbol{I}\right)^H+\sigma_n^2\boldsymbol{C}\boldsymbol{C}^H.
\end{aligned} \tag{124}$$

The total variance of the channel estimation error is just the trace of the above covariance matrix:

$$\begin{aligned}
&E\|\boldsymbol{h}-\boldsymbol{C}\boldsymbol{y}\|_F^2 \\
=&\,\mathrm{tr}\;E[(\boldsymbol{C}\boldsymbol{y}-\boldsymbol{h})(\boldsymbol{C}\boldsymbol{y}-\boldsymbol{h})^H] \\
=&\,\mathrm{tr}\left[\left(\frac{1}{N_tP_d+\sigma_n^2+P_\tau N/N_t}\boldsymbol{X}_\tau^H\boldsymbol{X}_d-\frac{N_tP_d+\sigma_n^2}{N_tP_d+\sigma_n^2+P_\tau N/N_t}\boldsymbol{I}\right)\right. \\
&\left.\left(\frac{1}{N_tP_d+\sigma_n^2+P_\tau N/N_t}\boldsymbol{X}_\tau^H\boldsymbol{X}_d-\frac{N_tP_d+\sigma_n^2}{N_tP_d+\sigma_n^2+P_\tau N/N_t}\boldsymbol{I}\right)^H\right]+\sigma_n^2\,\mathrm{tr}(\boldsymbol{C}\boldsymbol{C}^H).
\end{aligned} \tag{125}$$

Similarly to the flat fading case, we consider asymptotically long codewords, and we have

$$\lim_{N\to\infty}\mathrm{tr}\;E[(\boldsymbol{C}\boldsymbol{y}-\boldsymbol{h})(\boldsymbol{C}\boldsymbol{y}-\boldsymbol{h})^H]=\frac{\|\boldsymbol{X}_\tau^H\boldsymbol{X}_d\|_F^2}{(P_\tau N/N_t)^2}. \tag{126}$$

So, under the constraint that each training symbol is followed by N_t-1 trailing zeros, we get a similar asymptotic result in the frequency-selective fading case as in the flat fading case. Therefore, the mapping between training sequences and data sequences can be defined similarly to (118) and (119), i.e., for each Toeplitz matrix formed by a data sequence, choose the training sequence whose Toeplitz construction (120) is most orthogonal to the data matrix.

6.3.2 Design of the Set of Training Sequences

Although the transmitter can dynamically select a training sequence to minimize the MMSE of the channel estimate, the receiver has to figure out which training sequence is used

through a decision process in order to achieve this gain of better channel estimation. Obviously, errors can happen in this decision process, and a different set of training sequences probably leads to a different error probability. Therefore, we need to design the set of training sequences to minimize the probability that the receiver erroneously makes a decision about which training sequence is used by the transmitter. This probability is the error probability between one set of codewords and another set of codewords. Another interesting quantity is the intra-set error probability, which should also be minimized. Therefore, we take the approach of minimizing the overall error probability under the mapping defined in (119).

Analytical assessment of the overall error probability seems intractable. Therefore, we resort to using a bound on the error probability as the design criterion. As a common practice, we employ the union bound for this purpose, which is known to be tight at high channel SNRs [58].

The codeword error probability can be upper-bounded using the union bound according to

$$\begin{aligned} P_e &= \frac{1}{K}\sum_{i\in\mathcal{J}} \Pr(\hat{\boldsymbol{X}} \neq \boldsymbol{X}_i | \boldsymbol{X}_i \text{ transmitted}) \\ &\leq \frac{1}{K}\sum_{i\in\mathcal{J}}\sum_{j\in\mathcal{J}, j\neq i} p_{j|i}, \end{aligned} \tag{127}$$

where $p_{j|i}$ is the PEP between codewords $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ conditioned on $\boldsymbol{X}_i$ being transmitted.

The form of $p_{j|i}$ depends on the detection algorithms utilized by the receivers. As mentioned before, we consider ML decoder performance in this chapter. The expression of PEP for ML decoders in a noncoherent channel was derived in [10]. A Chernoff bound on this PEP was given in [17]. Since the Chernoff bound of PEP has a simpler form and is easier to evaluate than the PEP itself, we chose to use the Chernoff bound on PEP as our performance metric when designing the training codes. We cite the Chernoff bound result in [17, Eq. (3.12)] in the following.

Lemma 3. *At asymptotically high SNR, the Chernoff upper bound of PEP (127) for the*

channel model (91) is given by

$$P_{CB}^{i,j}(\lambda) = \frac{1}{2} \frac{\det(\boldsymbol{X}_i^H \boldsymbol{X}_i)^{\lambda} \det(\boldsymbol{X}_j^H \boldsymbol{X}_j)^{1-\lambda}}{\det\left[\begin{pmatrix} \lambda \boldsymbol{I} & \\ & (1-\lambda)\boldsymbol{I} \end{pmatrix}^{\frac{1}{2}} \begin{pmatrix} \boldsymbol{X}_i^H \\ \boldsymbol{X}_j^H \end{pmatrix} \begin{pmatrix} \boldsymbol{X}_i & \boldsymbol{X}_j \end{pmatrix} \begin{pmatrix} \lambda \boldsymbol{I} & \\ & (1-\lambda)\boldsymbol{I} \end{pmatrix}^{\frac{1}{2}} \right]}. \tag{128}$$

In (128), λ is a parameter that can be optimized to minimize the Chernoff bound and make it tight. The value of $\lambda = 1/2$ is optimal for various cases [17], for example, the so-called equal energy case, where $\boldsymbol{X}_i^H \boldsymbol{X}_i = \boldsymbol{X}_j^H \boldsymbol{X}_j$. For more general cases, $\lambda = 1/2$ is not optimal and the Chernoff bound $P_{CB}^{i,j}(1/2)$ is less tight. However, $P_{CB}^{i,j}(1/2)$ still serves as an upper bound and is used in this chapter as a performance metric. $P_{CB}^{i,j}(1/2)$ has the following expression

$$P_{CB}^{i,j}(1/2) = \frac{1}{2} \frac{\det(\boldsymbol{X}_i^H \boldsymbol{X}_i)^{\frac{1}{2}} \det(\boldsymbol{X}_j^H \boldsymbol{X}_j)^{\frac{1}{2}}}{\det\left[\frac{1}{2} \begin{pmatrix} \boldsymbol{X}_i^H \\ \boldsymbol{X}_j^H \end{pmatrix} \begin{pmatrix} \boldsymbol{X}_i & \boldsymbol{X}_j \end{pmatrix} \right]} \tag{129}$$

and will be denoted as $P_{CB}^{i,j}$ from now on. By observation, the roles played by $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ in $P_{CB}^{i,j}$ are symmetric. Therefore, the union bound in (127) has become

$$P_e \leq \frac{1}{K(K-1)/2} \sum_{i=0}^{K-2} \sum_{j=i+1}^{K-1} P_{CB}^{i,j}. \tag{130}$$

At this point, the design problem can be stated. Given a signal constellation and a channel model with a specific coherence interval, we want to design a set of training sequences so that under the mapping of (119) at the transmitter, the Chernoff upper bound on the union bound is minimized.

To design the multiple training sequences, we first consider properties these training sequences should have, because a structured set can make the numerical optimization efforts easier. For the flat fading channel, we will empirically reason that from the channel estimation point of view, if $M \leq N$, the training sequences should be chosen to be orthogonal. Recall that we assumed that the data symbols are picked i.i.d. from a QAM or PAM constellation, it follows that the data sequences are evenly spaced in an N-d Euclidean space. Since the criterion of matching data sequences with training sequences is orthogonality, the

training sequences should also be placed evenly in the space. When the number of training sequences is less than the dimension of the space, the most even way to place them in the space is to make them orthogonal. For frequency-selective fading, with a similar argument, we can show that when the training sequences have the form stated in (120), and under the condition that $M \leq N/N_t$, the non-zero-padded training sequences should be orthogonal to each other.

Now that we have shown the training sequences should be orthogonal to each other, we also constrain that all training sequences have equal power. If we define $\boldsymbol{\mathcal{T}}$ to be the matrix whose column vectors are the training sequences, then $\boldsymbol{\mathcal{T}}$ is a scaled orthonormal matrix because $\boldsymbol{\mathcal{T}}^H\boldsymbol{\mathcal{T}} = NP_\tau \boldsymbol{I}_M$. The optimization problem has become one with an orthonormality constraint. Various gradient descent types of numerical optimization algorithms exist for problems with an orthogonality constraint [19, 45]. Since we do not have analytical expressions for the gradient of the sum PEP with respect to training sequences and the training power allocation, we chose to parameterize the orthonormal matrices, thereby change the problem into one without orthonormal constraints, and then optimize these parameters utilizing a line search algorithm implemented using the MATLAB function *fmincon*.

A natural modeling of orthonormal matrices is the Stiefel manifold [9], in which each point is an orthonormal matrix. The number of degrees of freedom of an $N \times M$ orthonormal complex matrix, where $N \geq M$, is less then NM. An $N \times M$ orthonormal matrix has the following parametrization

$$\exp \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B}^H \\ \boldsymbol{B} & \boldsymbol{0} \end{pmatrix} \boldsymbol{I}_{N,M}, \tag{131}$$

where $\boldsymbol{B}$ is an arbitrary $(N-M) \times M$ complex matrix and $\boldsymbol{A}$ is an $M \times M$ skew Hermitian matrix, which means that $\boldsymbol{A}^H = -\boldsymbol{A}$. The matrix $\boldsymbol{A}$ can be further parameterized by its upper triangular or lower triangular components including the main diagonal.

The optimization of the set of training sequences for the flat fading channel are summarized as following, and optimization for frequency-selective channels can be similarly inferred:

- Training sequences have equal power;

- If $M \leq N$, training sequences should be orthogonal to each other;
- The matrix having normalized training sequences as columns can be parameterized as in (131);
- The parameters from (131) and the power allocation between data and training sequences are optimized to minimize the Chernoff upper bound on the union bound (130), through a line search algorithm.

6.4 Numerical Examples

In this section, we compare our proposed data-dependent superimposed training with TDM training and traditional superimposed training with a single training sequence. For TDM training, we optimized the power allocation between the training symbol and data symbols. For superimposed training schemes, we optimize both the power allocation embodied by P_τ/P_d and the training sequences. The receiver uses a GLRT receiver because it performs similarly to ML receivers but is slightly simpler.

We first point out that if the training sequence is not properly designed for the traditional superimposed training with a single training sequence, the noncoherent system may not work at all. For example, consider BPSK modulated data symbols. Since we consider very short coherence intervals and codeword length, the data sequences $\{-1, -1, \ldots, -1\}$ and $\{1, 1, \ldots, 1\}$ can happen. In this case, if the training sequence is a repetition of some symbol, say x_τ, then the resultant codewords would be $\{-1 + x_\tau, -1 + x_\tau, \ldots, -1 + x_\tau\}$ and $\{1 + x_\tau, 1 + x_\tau, \ldots, 1 + x_\tau\}$. These two sequences cannot be discerned at the receiver by the GLRT receiver (107) because the projection matrices onto the signal subspace are exactly the same. Therefore, for fast fading, the single superimposed training sequence should be optimized, which has seldom been done before.

In the figures' legends, *TDM* denotes TDM training. An integer in the legend indicates the number of orthogonal training sequences used when constructing the codebook. In all of the following figures, the curves inside the same figure correspond to the same information rate for fairness.

6.4.1 Flat Fading, $N = 2$, QPSK vs 16-QAM

In the first example, we consider a flat-fading channel with coherence interval $N = 2$. For TDM training, the first symbol is a training symbol and the second symbol is picked from a 16-QAM constellation. For superimposed training, data symbols are QPSK modulated. For a single training sequence, the optimized $P_\tau = 0.3178P$, and the normalized training sequence is

$$\boldsymbol{x}_\tau = \begin{pmatrix} 0 \\ 0.2064 + 0.9785i \end{pmatrix}.$$

Interestingly, the superimposed training sequence is only nonzero in one of the two entries. For a system with two training sequences, $P_\tau = 0.1134P$, and the normalized optimized training sequences are written as columns of the following matrix

$$\begin{pmatrix} -0.3054 - 0.6376i & -0.6376 + 0.3059i \\ -0.3059 + 0.6376i & -0.6376 - 0.3054i \end{pmatrix}.$$

For TDM training, the optimized $P_\tau = 0.4718P$. The performance comparison in terms of codeword error probability is shown in Fig. 17. Note that, although we plot the performance across a range of SNRs, the optimization was only performed for asymptotically high SNRs. We can see that two training sequences offer about 1 dB gain compared to a single superimposed training sequence and also offer some gain relative to TDM training.

6.4.2 Flat Fading, $N = 4$, 4-PAM vs 6-PAM

In this example, we consider a real channel with coherence interval $N = 4$. The superimposed training systems utilize a 4-PAM data constellation, while the TDM training system uses 6-PAM. For a single superimposed training sequence, $P_\tau = 0.0311P$, and the normalized training sequence is

$$(\, -0.6015 \;\; -0.3694 \;\; 0.7083 \;\; -0.0006 \,)^T.$$

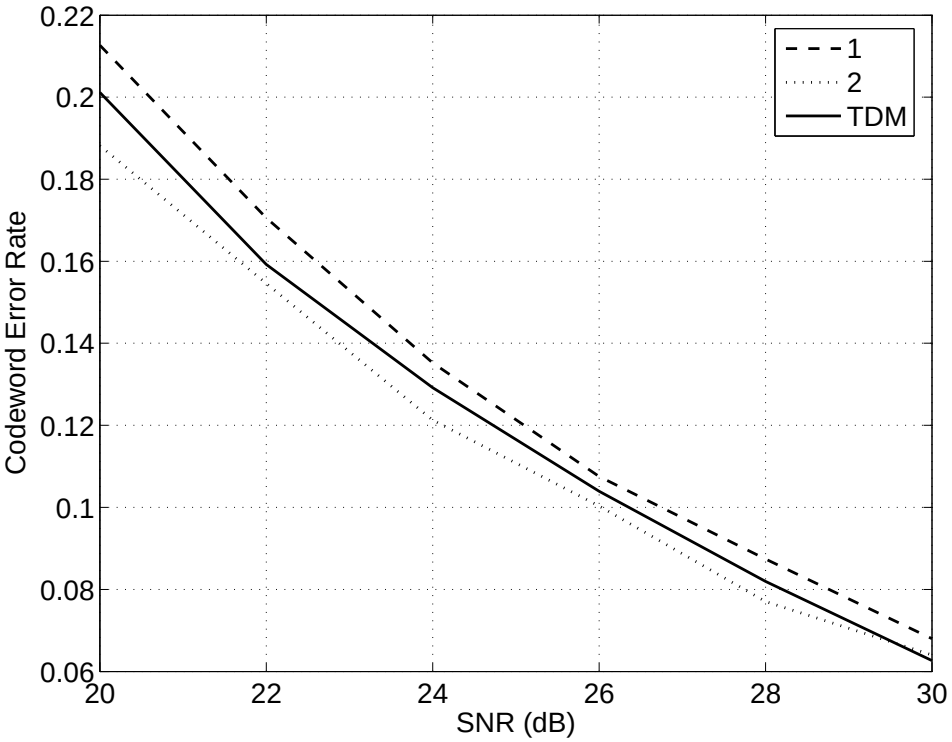


Figure 17: Codeword error rate example with $N = 2$. The TDM training scheme uses 16-QAM data symbols, and the superimposed training schemes use QPSK data symbols.

For the case of four superimposed training sequences, $P_\tau = 0.6087P$, and the normalized training sequences are columns of

$$\begin{pmatrix} -0.5823 & -0.7477 & 0.1199 & 0.2958 \\ 0.3660 & -0.1701 & -0.7092 & 0.5780 \\ -0.6222 & 0.6386 & -0.1181 & 0.4372 \\ -0.3739 & -0.0649 & -0.6846 & -0.6224 \end{pmatrix}.$$

For TDM training with a single training symbol placed at the beginning of the codeword, $P_\tau = 0.2081P$.

We first compare the codeword error probability between TDM training and four-sequence superimposed training in Fig. 18. Note that in the setting of this example, the rate of superimposed training is more than that of TDM training. Therefore, for Fig. 18, we sequentially removed those codewords that have largest PEP with other codewords, until the number of codewords is equal to that of TDM training. This process is not optimal in any sense and is disadvantageous for four-sequence superimposed training. Even so, we

can see from Fig. 18 that four-sequence superimposed training has more than 1 dB gain over TDM training.

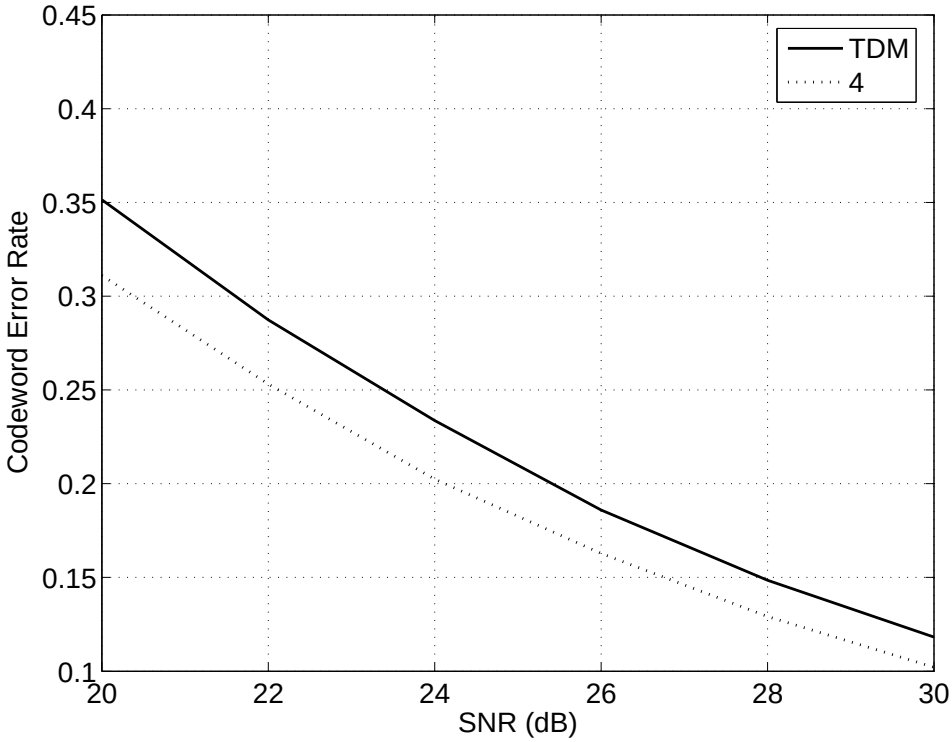
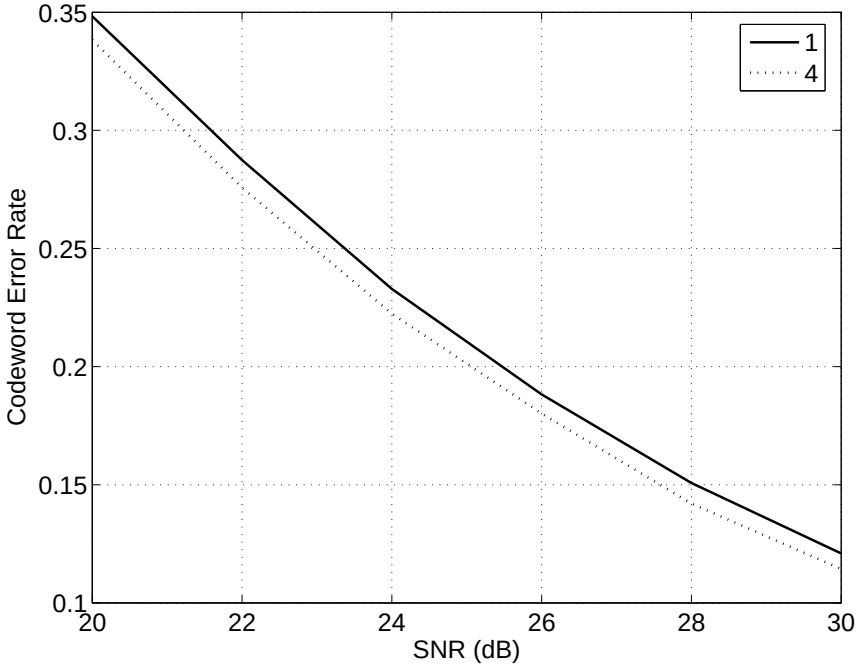


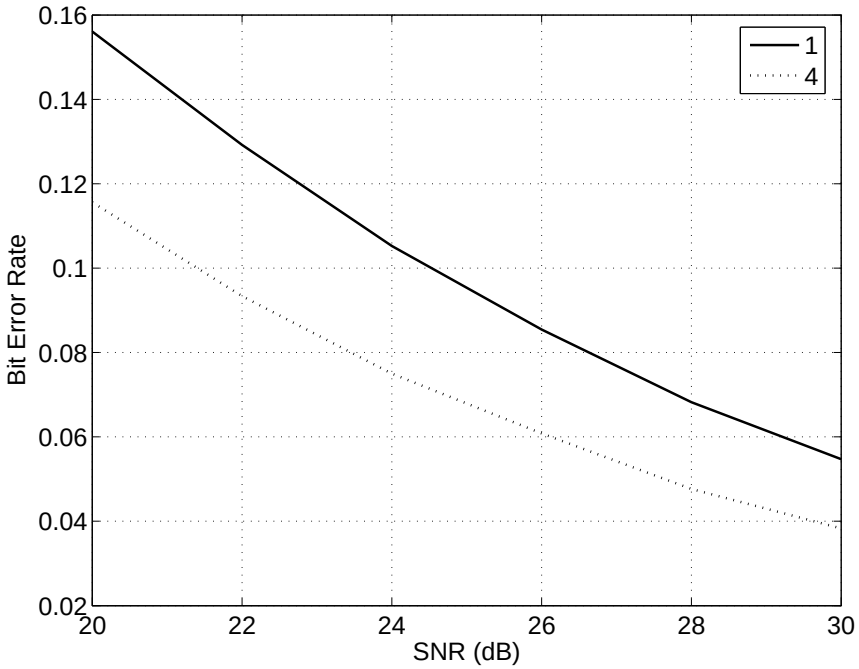
Figure 18: Codeword error rate example with $N = 4$. The TDM training scheme uses 6-PAM data symbols, and the superimposed training scheme uses 4-PAM data symbols.

The performance between four-sequence and a single sequence superimposed training in terms of codeword error rate and bit error rate is compared in Fig. 19. Here the mapping between bits and 4-PAM symbols is the Gray code [58]. The four-sequence superimposed training has a large gain over the single training sequence case in terms of bit error rate.

The LMMSE channel estimation error of various schemes is plotted in Fig. 20. Note that the LMMSE channel estimate only relies on the training sequence or symbol and the variance of data symbols and noise samples. For the multiple training sequences case, the channel estimation is performed using the correct training sequence associated with each codeword. We observed that the four-sequence system has a much better channel estimation performance than the single superimposed training sequence system. This gain comes not only from the orthogonality criterion in the mapping (119), but also from the fact that a larger percentage of the total power is allocated to the training part in this example. We



(a)



(b)

Figure 19: Error rates for superimposed training with $N = 4$ and 4-PAM data symbols. (a) Codeword error rate; (b) Bit error rate.

also observe that TDM training benefits from high SNR, while the superimposed training has an error floor at high SNR because the data symbols appear as high-variance noise to the channel estimator.

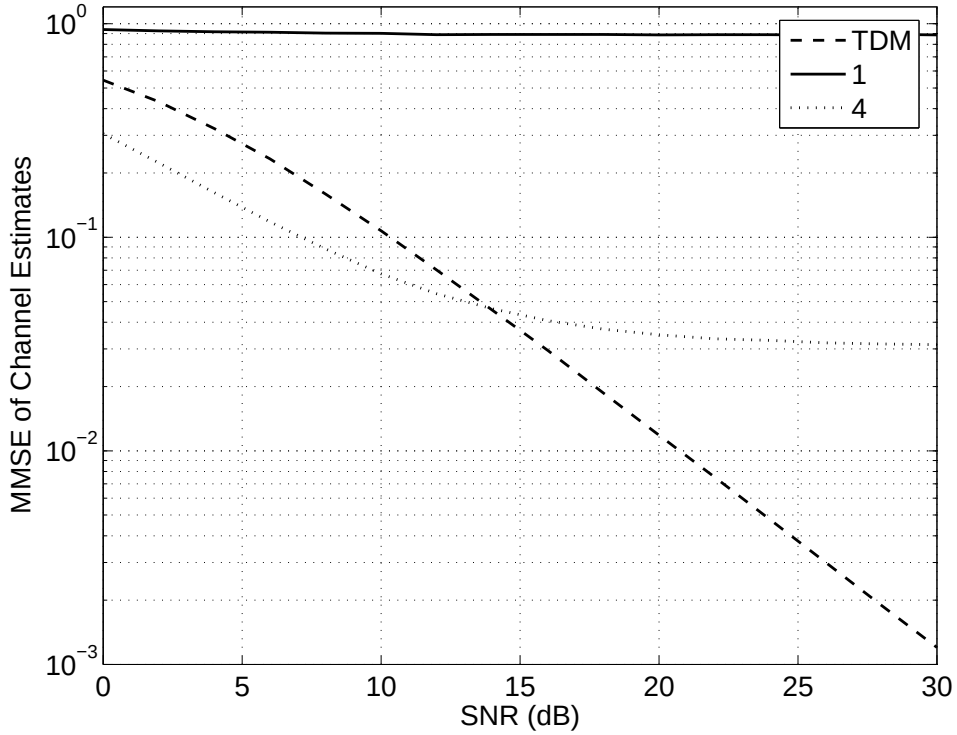


Figure 20: MMSE of channel estimates with $N = 4$. TDM training scheme uses 6-PAM data symbols, and superimposed training schemes use 4-PAM data symbols.

6.4.3 Frequency-Selective Fading, $N = 6$, $N_t = 3$, QPSK vs 256-QAM

As a last example, we consider a complex frequency-selective fading channel with $N_t = 3$ symbol-spaced taps and a coherence interval of $N = 6$. While the superimposed training system can transmit four symbols followed by two trailing zeros, the TDM training system can only transmit one training symbol followed by two trailing zeros, and then one data symbol followed by two trailing zeros. If the superimposed training system uses QPSK modulation for the data, the TDM system has to use 256-QAM to get the same information rate. For single sequence superimposed training, $P_\tau = 0.0644$, and the normalized training sequence without accounting for the last two trailing zeros is

$$(\, 0.6676 - 0.7445i \;\; 0 \;\; 0 \;\; 0 \,)^T.$$

This is an interesting result because the optimization process was initiated from a general length-4 sequence but ended with only one nonzero entry. This partly motivated our constraint (120) on the structure of other multiple training sequences. For two superimposed training sequences, $P_\tau = 0.0873P$, and the training sequences are the columns of

$$\begin{pmatrix} 0.3982 + 0.0013i & 0.6456 - 0.6517i \\ 0 & 0 \\ 0 & 0 \\ -0.6536 - 0.6436i & 0.3981 - 0.0062i \end{pmatrix},$$

without accounting for the last two trailing zeros. The optimized TDM training has $P_\tau = 0.5287P$. The codeword error performance is shown in Fig. 21. The superimposed training systems have a large gain compared to TDM training, most likely because of the high modulation order of data symbols in the TDM training setting. The two-sequence superimposed training has a modest gain relative to the single sequence case.

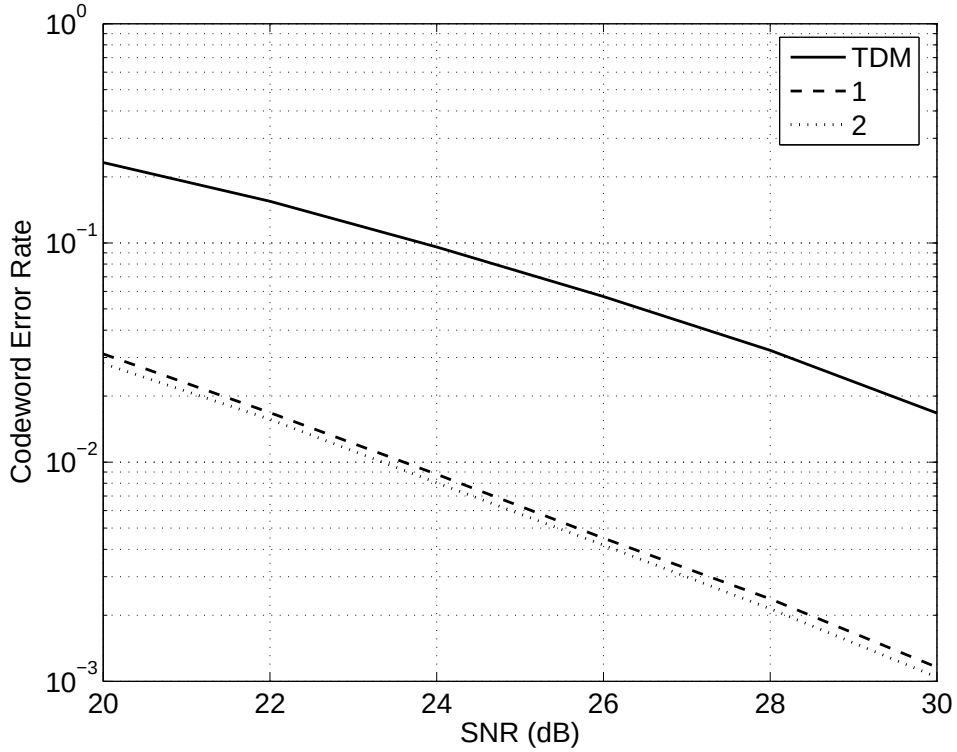


Figure 21: Codeword error rate example with $N = 6$. The TDM training scheme uses 256-QAM data symbols, and the superimposed training schemes use QPSK data symbols.

6.5 Conclusion and Discussion

We considered the design of a data-dependent superimposed training system for a fast fading noncoherent channel. Given a data codeword at the transmitter, a training codeword that is most orthogonal to the data codeword is selected and added to the data to form the final codeword. The training sequences are designed to minimize the Chernoff upper bound on the union bound of codeword error rate. For GLRT receivers in a frequency-selective channel, the superimposed training systems show great performance gain over the TDM training system. For flat fading channels, data-dependent superimposed training system shows about 1 dB gain over TDM training. The performance improvement for multiple training sequences over the single training sequence is more prominent in flat fading channels rather than frequency-selective channels.

We only considered the analytic Chernoff bound on pairwise error probability for a maximum likelihood receiver in this chapter. Other criteria can be used, for example, packet error rate or bit error rate. We can also optimize the set of training sequences for other detection schemes, for example, a suboptimal and simpler one that first estimates the channel and then performs coherent detection taking the channel estimate as the true channel. We envision a structure with multiple parallel channel estimation/data detection branches, each for a training sequence, followed by a final decision device. Since analytical expressions of these error probabilities are unlikely to exist, simulation based optimization methods may be necessary [68]. Extending the current work to multiple transmit antennas would also be interesting. Besides channel estimation, the superimposed training scheme may also be helpful for frequency and timing synchronization purposes because of the universal presence of the training sequence.

CHAPTER VII

CONCLUSIONS

Wireless communication systems equipped with multiple transmit and receive antennas are able to achieve higher throughput and/or reliability than single antenna systems without increasing bandwidth. The spectral efficiency of a MIMO system can be further improved if the transmitter has access to channel state information. Perfect CSI at the transmitter is hard to acquire in FDD systems. For the transmitter to achieve better performance than transmission without CSI in such cases, the receiver has to feedback partial channel state information to the transmitter.

Grassmannian subspace packing based vector quantization is a strong candidate for CSI feedback codebook design in pseudo-static block fading channels. For continuous fading, which is more realistic, Grassmannian packing based VQ scheme does not exploit the time-domain correlation of channel states. In Chapter 3, we proposed new codebook design and CSI feedback schemes that utilize both time-domain channel correlation and intrinsic properties of the variation of subspaces. Instead of quantizing the subspaces themselves, we proposed to quantize the geodesic trajectory connecting two subspaces. More specifically, we quantize a key entity that characterizes a geodesic arc: the velocity matrix. Our schemes achieved significant performance improvement compared to the gradient sign algorithm, because we compactly represent the source using piecewise geodesic modeling and quantize a smaller number of parameters. For slow fading and strong time-domain correlation, our scheme is also better than Grassmannian subspace packings. Our perspective of CSI feedback codebook design based on channel Doppler frequency is also novel.

We designed a MIMO OFDM transmitter with partial CSI at the transmitter under special complexity constraints in Chapter 4. To limit the number of (I)DFTs per OFDM symbol to be one, time-domain beamforming and combining are performed at the transmitter and the receiver, respectively. The performance penalty is analyzed relative to a

full-complexity system that requires as many (I)DFTs as the number of antennas.

Imperfections in the CSI feedback channel have rarely been treated in the literature. We considered an arbitrary delay in the feedback channel in Chapter 5 and proposed a broadcast approach to adapt the reliably decodable information rate to the quality of CSI. This is achieved by transmitting a first layer that has a white covariance matrix and a second layer along the direction as stated by the CSI. Power allocation between the two layers as well as information rates were optimized to maximize the expected throughput. Significant throughput gain was observed in some regions of CSI quality and SNRs.

Finally, we relaxed the assumption that the receiver always has perfect CSI and considered noncoherent training codes for fast fading channels in Chapter 6. Training based schemes are more attractive than unstructured unitary space-time codes for complexity considerations because training symbols help the receiver obtain an estimate of the channel and initiate further decoding processes. We generalized the idea of superimposed training to multiple training sequences. For each data sequence, the transmitter dynamically matches the data with a training sequence that is most orthogonal to it. This mapping reduces channel estimation error if the receiver knows which training sequence is used by the transmitter. To reduce the probability that the receiver assumes a wrong training sequence, the set of training sequences was optimized to minimize the Chernoff bound on sum PEPs. The superimposed training schemes showed significant gain over TDM training schemes for fast fading frequency-selective channels. Multiple training sequence codes showed better performance than single training sequence codes in all cases, but the gain varies depending on the setting.

In future research, effective CSI feedback schemes for MIMO OFDM systems can be considered. This problem is more challenging than flat fading channels because each subcarrier has a different MIMO channel matrix. However, in a MIMO OFDM system, the power-delay profile varies more slowly than time-domain taps. The correlation of channel matrices on different subcarriers, as determined by the power-delay profile, also varies slowly. Proper utilization of this prior information constitutes interesting research topics. Another interesting area is optimal transmitter design given partial CSI in the setting of

multiuser downlink channels, where the transmitter tries to transmit independent information to multiple receivers simultaneously.

In terms of training codes, future research can probe into training sequence design for suboptimal but simpler receivers that perform channel estimation followed by coherent detection. With the existence of an outer code, the receiver structure and the convergence behavior of an iterative detector/channel estimator can be investigated.

APPENDIX A

SUPPLEMENTARY FOR CHAPTER III

A.1

From (26),

$$\begin{aligned}
&\boldsymbol{Q}[0]\exp(\boldsymbol{B}[0])\boldsymbol{I}_{N_t,N_s}\\
=&(\ \boldsymbol{W}[0]\ \ \boldsymbol{Z}[0]\ )\begin{pmatrix}\boldsymbol{U}_1\boldsymbol{C}\boldsymbol{U}_1^H\\ \boldsymbol{U}_2\boldsymbol{S}\boldsymbol{U}_1^H\end{pmatrix}\\
=&\boldsymbol{W}[0]\boldsymbol{U}_1\boldsymbol{C}\boldsymbol{U}_1^H+\boldsymbol{Z}[0]\boldsymbol{U}_2\boldsymbol{S}\boldsymbol{U}_1^H.
\end{aligned}\tag{132}$$

Similarly,

$$\boldsymbol{Q}[0]\exp(-\boldsymbol{B}[0])\boldsymbol{I}_{N_t,N_s}=\boldsymbol{W}[0]\boldsymbol{U}_1\boldsymbol{C}\boldsymbol{U}_1^H-\boldsymbol{Z}[0]\boldsymbol{U}_2\boldsymbol{S}\boldsymbol{U}_1^H.\tag{133}$$

Using (132) and (133), the terms inside the signum function in (37) can be expanded as

$$\begin{aligned}
&\|\boldsymbol{H}\boldsymbol{Q}[0]\exp(\boldsymbol{B}[0])\boldsymbol{I}_{N_t,N_s}\|_F^2-\|\boldsymbol{H}\boldsymbol{Q}[0]\exp(-\boldsymbol{B}[0])\boldsymbol{I}_{N_t,N_s}\|_F^2\\
=&\operatorname{tr}(\boldsymbol{I}_{N_t,N_s}^H\exp(\boldsymbol{B}[0])^H\boldsymbol{Q}^H[0]\boldsymbol{H}^H\boldsymbol{H}\boldsymbol{Q}[0]\exp(\boldsymbol{B}[0])\boldsymbol{I}_{N_t,N_s})\\
&-\operatorname{tr}(\boldsymbol{I}_{N_t,N_s}^H\exp(-\boldsymbol{B}[0])^H\boldsymbol{Q}^H[0]\boldsymbol{H}^H\boldsymbol{H}\boldsymbol{Q}[0]\exp(-\boldsymbol{B}[0])\boldsymbol{I}_{N_t,N_s})\\
=&4\operatorname{tr}\Re(\boldsymbol{U}_1\boldsymbol{C}\boldsymbol{U}_1^H\boldsymbol{W}^H[0]\boldsymbol{H}^H\boldsymbol{H}\boldsymbol{Z}[0]\boldsymbol{U}_2\boldsymbol{S}\boldsymbol{U}_1^H).
\end{aligned}\tag{134}$$

The binary decision $s[1]$ is simply the sign of the above expression

$$s[1]=\operatorname{sign}\operatorname{tr}\Re(\boldsymbol{U}_1\boldsymbol{C}\boldsymbol{U}_1^H\boldsymbol{W}^H[0]\boldsymbol{H}^H\boldsymbol{H}\boldsymbol{Z}[0]\boldsymbol{U}_2\boldsymbol{S}\boldsymbol{U}_1^H).\tag{135}$$

Substituting $\boldsymbol{C}\approx\boldsymbol{I}$ and $\boldsymbol{S}\approx\boldsymbol{\Theta}$ into (135), we have

$$\begin{aligned}
s[1]\approx&\operatorname{sign}\operatorname{tr}\Re(\boldsymbol{W}^H[0]\boldsymbol{H}^H\boldsymbol{H}\boldsymbol{Z}[0]\boldsymbol{U}_2\boldsymbol{\Theta}\boldsymbol{U}_1^H)\\
=&\operatorname{sign}\operatorname{tr}\Re(\boldsymbol{W}^H[0]\boldsymbol{H}^H\boldsymbol{H}\boldsymbol{Z}[0]\boldsymbol{A}[0]).
\end{aligned}\tag{136}$$

A.2

Here we cite one result of Appendix A of [4].

Lemma 4. *Let* $\boldsymbol{G}$ *be a non-random complex matrix, and* $\boldsymbol{P}$ *be a random matrix, having the same dimensions as* $\boldsymbol{G}$*, with* $\mathcal{CN}(0,1)$ *entries. If*

$$\boldsymbol{Z} \equiv [\operatorname{sign} \operatorname{tr} \Re(\boldsymbol{G}^H \boldsymbol{P})]\boldsymbol{P}, \tag{137}$$

then

$$E\{\boldsymbol{Z}\} = \frac{1}{\sqrt{\pi}} \frac{\boldsymbol{G}}{\|\boldsymbol{G}\|_F}. \tag{138}$$

Now substitute $\boldsymbol{P}$ in Lemma 4 with $\boldsymbol{A}[0] = a\boldsymbol{A}_w[0]$ and $\boldsymbol{G}$ with $\boldsymbol{Z}^H[0]\boldsymbol{H}^H\boldsymbol{H}\boldsymbol{W}[0]$, and then (46) follows straightforwardly.

A.3

$$a^2 = p_1 {F_D}^8 + p_2 {F_D}^7 + p_3 {F_D}^6 + p_4 {F_D}^5 + p_5 {F_D}^4 + p_6 {F_D}^3 + p_7 {F_D}^2 + p_8 {F_D}^1 + p_9. \tag{139}$$

p_1	1.7258e-20
p_2	-2.9536e-17
p_3	2.0865e-14
p_4	-7.8741e-12
p_5	1.7062e-9
p_6	-2.0741e-7
p_7	1.0147e-5
p_8	0.0012022
p_9	-0.0029007

APPENDIX B

SUPPLEMENTARY FOR CHAPTER V

B.1

$$
\begin{aligned}
R_{WW} =& E_{\boldsymbol{h}\sim\mathcal{CN}(\boldsymbol{0},\alpha^2\boldsymbol{I})}\{\log[1+P\boldsymbol{h}^H(\beta\boldsymbol{I}+\xi\boldsymbol{I})\boldsymbol{h}]-\log(1+P\boldsymbol{h}^H\xi\boldsymbol{I}\boldsymbol{h})\} \\
&+(1-\eta)E_{\boldsymbol{h}\sim\mathcal{CN}(\eta\boldsymbol{g},(1-\eta^2)\alpha^2\boldsymbol{I})}\log[1+P\boldsymbol{h}^H\xi\boldsymbol{I}\boldsymbol{h}] \\
=& \int_0^\infty \frac{e^{-\frac{y}{\alpha^2 P}}}{y}\left[1-\frac{1}{(1+y)^{N_t}}\right]dy-\int_0^\infty \frac{e^{-\frac{y}{\alpha^2\xi P}}}{y}\left[1-\frac{1}{(1+y)^{N_t}}\right]dy \\
&+(1-\eta)\int_0^\infty \frac{e^{-\frac{y}{(1-\eta^2)\alpha^2\xi P}}}{y}\left[1-\frac{\exp\left(-\frac{\eta^2\|\boldsymbol{g}\|^2 y}{\alpha^2(1-\eta^2)(1+y)}\right)}{(1+y)^{N_t}}\right]dy.
\end{aligned}
\tag{140}
$$

$$
\begin{aligned}
R_{WB} =& E_{\boldsymbol{h}\sim\mathcal{CN}(\boldsymbol{0},\alpha^2\boldsymbol{I})}\{\log[1+P\boldsymbol{h}^H(\beta\boldsymbol{I}+N_t\xi\frac{\boldsymbol{g}\boldsymbol{g}^H}{\boldsymbol{g}^H\boldsymbol{g}})\boldsymbol{h}]-\log(1+P\boldsymbol{h}^H N_t\xi\frac{\boldsymbol{g}\boldsymbol{g}^H}{\boldsymbol{g}^H\boldsymbol{g}}\boldsymbol{h})\} \\
&+(1-\eta)E_{\boldsymbol{h}\sim\mathcal{CN}(\eta\boldsymbol{g},(1-\eta^2)\alpha^2\boldsymbol{I})}\log[1+P\boldsymbol{h}^H N_t\xi\frac{\boldsymbol{g}\boldsymbol{g}^H}{\boldsymbol{g}^H\boldsymbol{g}}\boldsymbol{h}] \\
=& \int_0^\infty \frac{e^{-\frac{y}{\alpha^2 P}}}{y}\left[1-\frac{1}{(1+(\beta+N_t\xi)y)(1+\beta y)^{N_t-1}}\right]dy-\int_0^\infty \frac{e^{-\frac{y}{\alpha^2 N_t\xi P}}}{1+y}dy \\
&+(1-\eta)\int_0^\infty \frac{e^{-\frac{y}{(1-\eta^2)\alpha^2 P}}}{y}\left[1-\frac{\exp\left(-\frac{\eta^2\|\boldsymbol{g}\|^2 N_t\xi y}{\alpha^2(1-\eta^2)(1+N_t\xi y)}\right)}{1+N_t\xi y}\right]dy.
\end{aligned}
\tag{141}
$$

$$
\begin{aligned}
R_{BB} =& E_{\boldsymbol{h}\sim\mathcal{CN}(\boldsymbol{0},\alpha^2\boldsymbol{I})}\{\log[1+P\boldsymbol{h}^H(N_t\beta\frac{\boldsymbol{g}\boldsymbol{g}^H}{\boldsymbol{g}^H\boldsymbol{g}}+N_t\xi\frac{\boldsymbol{g}\boldsymbol{g}^H}{\boldsymbol{g}^H\boldsymbol{g}})\boldsymbol{h}]-\log(1+P\boldsymbol{h}^H N_t\xi\frac{\boldsymbol{g}\boldsymbol{g}^H}{\boldsymbol{g}^H\boldsymbol{g}}\boldsymbol{h})\} \\
&+(1-\eta)E_{\boldsymbol{h}\sim\mathcal{CN}(\eta\boldsymbol{g},(1-\eta^2)\alpha^2\boldsymbol{I})}\log[1+P\boldsymbol{h}^H N_t\xi\frac{\boldsymbol{g}\boldsymbol{g}^H}{\boldsymbol{g}^H\boldsymbol{g}}\boldsymbol{h}] \\
=& \int_0^\infty \frac{e^{-\frac{y}{\alpha^2 P}}}{y}\left[1-\frac{1}{1+N_t y}\right]dy-\int_0^\infty \frac{e^{-\frac{y}{\alpha^2\xi P}}}{y}\left[1-\frac{1}{1+N_t y}\right]dy \\
&+(1-\eta)\int_0^\infty \frac{e^{-\frac{y}{(1-\eta^2)\alpha^2\xi P}}}{y}\left[1-\frac{\exp\left(-\frac{\eta^2\|\boldsymbol{g}\|^2 N_t y}{\alpha^2(1-\eta^2)(1+N_t y)}\right)}{1+N_t y}\right]dy.
\end{aligned}
\tag{142}
$$

APPENDIX C

ACRONYMS

Acronyms	Description
$A \sim B$	A is distributed as B
AWGN	Additive white Gaussian noise
BER	Bit-error rate
BPSK	Binary phase-shift keying
CDM	Code-division multiplexed
CDMA	Code-division multiple access
CSI	Channel state information
DFT	Discrete Fourier Transform
FDD	Frequency-division duplex
FDM	Frequency-division multiplexed
GLRT	Generalized likelihood ratio test
IDFT	Inverse Discrete Fourier Transform
i.i.d.	Independent and identically distributed
ISI	Intersymbol interference
MIMO	Multiple-input multiple-output
MISO	Multiple-input single-output
ML	Maximum likelihood
n-D	n-dimensional
OFDM	Orthogonal frequency division multiplexing
PAM	Pulse amplitude modulation
PEP	Pairwise error probability
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature phase-shift keying

Acronyms	**Description**
SIMO	Single input multiple output
SNR	Signal-to-noise ratio
STBC	Space-time block coding
SU	Subscriber unit
SVD	Singular value decomposition
TDM	Time-division multiplexed
VQ	Vector quantization
$\mathbb{C}^n$	n-dimensional complex space
$\mathcal{CN}(\mu, \sigma^2)$	The complex Gaussian distribution with mean μ and variance σ^2
$\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{Q})$	The complex Gaussian vector distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{Q}$
$\text{diag}(\boldsymbol{x})$	A (block) diagonal matrix with $\boldsymbol{x}$ on its diagonal
$E\{\cdot\}$	Expectation
$E_{A\sim B}\{\cdot\}$	Expectation with regard to A that is distributed as B
$\boldsymbol{I}$	Identity matrix
$\boldsymbol{I}_n$	$n \times n$ identity matrix
$\boldsymbol{I}_{N_t,N_s}$	Matrix Containing the first N_s columns of $\boldsymbol{I}_{N_t}$
$\Re\{\cdot\}$	Real part of complex entries
$\text{sign}(\cdot)$	Signum function
$\text{tr}\{\cdot\}$	Trace of a matrix
$(\cdot)^*$	Conjugation
$(\cdot)^H$	Hermitian transpose
$\|\cdot\|_F$	Frobenius norm of a matrix
$[\,\cdot\,]_p$	The pth entry of a vector
$[\,\cdot\,]_{p,q}$	The (p,q)th entry of a matrix
$\mathbf{0}$	All-zero vector or matrix
$\mathbf{0}_{m\times n}$	$m \times n$ all-zero matrix

REFERENCES

[1] Adireddy, S., Tong, L., and Viswanathan, H., "Optimal placement of training for frequency-selective block-fading channels," *IEEE Trans. Inform. Theory*, vol. 48, pp. 2338–2353, Aug. 2002.

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

[3] Bahai, A. R. S., Saltzberg, B. R., and Ergen, M., *Multi-Carrier Digital Communications Theory and Applications of OFDM.* New York, NY: Springer, 2nd ed., 2004.

[4] Banister, B. C. and Zeidler, J. R., "Feedback assisted transmission subspace tracking for MIMO systems," *IEEE J. Selected Areas Commun.*, vol. 21, pp. 452–463, Apr. 2003.

[5] Banister, B. C. and Zeidler, J. R., "A simple gradient sign algorithm for transmit antenna weight adaptation with feedback," *IEEE Trans. Signal Processing*, vol. 51, pp. 1156–1171, May 2003.

[6] Blackwell, D., Breiman, L., and Thomasian, A. J., "The capacity of a class of channels," *The Annals of Mathematical Statistics*, vol. 30, pp. 1229–1241, Dec. 1959.

[7] Blum, R. S. and Winters, J. H., "On optimum MIMO with antenna selection," in *Proc. IEEE Int. Conf. Commun.*, vol. 1, (New York, NY), pp. 386–390, Apr. 2002.

[8] Bölcskei, H., Gesbert, D., and Paulraj, A. J., "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Trans. Commun.*, vol. 50, pp. 225–234, Feb. 2002.

[9] Boothby, W. M., *An Introduction to Differentiable Manifolds and Riemannian Geometry.* New York: Academic Press, 2nd ed., 1986.

[10] Brehler, M. and Varanasi, M. K., "Asymptotic error probability analysis of quadratic receivers in rayleigh-fading channels with applications to a unified analysis of coherent and noncoherent space-time receivers," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2383–2399, Sept. 2001.

[11] Conway, J. H., Hardin, R. H., and Sloane, N. J. A., "Packing lines, planes, etc.: Packings in grassmannian spaces," *Experimental Mathematics*, vol. 5, no. 2, pp. 139–159, 1996.

[12] Coskun, O. and Chugg, K. M., "Combined coding and training for unknown ISI channels," *IEEE Trans. Commun.*, vol. 53, pp. 1310–1322, Aug. 2005.

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

[14] Cover, T. M., "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. 18, pp. 2–14, Jan. 1972.

[15] Cozzo, C. and Hughes, B. L., "Joint channel estimation and data detection in space-time communications," *IEEE Trans. Commun.*, vol. 51, pp. 1266–1270, Aug. 2003.

[16] Dayal, P., Brehler, M., and Varanasi, M. K., "Leveraging coherent space-time codes for noncoherent communication via training," *IEEE Trans. Inform. Theory*, vol. 50, pp. 2058–2080, Sept. 2004.

[17] Dogandzić, A., "Chernoff bounds on pairwise error probabilities of space-time codes," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1327–1336, May 2003.

[18] Dong, M., Tong, L., and Sadler, B. M., "Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels," *IEEE Trans. Signal Processing*, vol. 52, pp. 1403–1418, May 2004.

[19] Edelman, A., Arias, T. A., and Smith, S. T., "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix. Anal. Appl.*, vol. 20, pp. 303–353, Oct. 1998.

[20] Farhang-Boroujeny, B., "Pilot-based channel identification: Proposal for semi-blind identification of communication channels," *Electronics Letters*, vol. 31, pp. 1044–1046, June 1995.

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

[22] Gallivan, K. A., Srivastava, A., Liu, X., and Dooren, P. V., "Efficient algorithms for inferences on Grassmann manifolds," in *Proc. IEEE Workshop on Statistical Signal Processing*, pp. 315–318, Sept. 2003.

[23] Gamal, H. E., Caire, G., and Damen, M. O., "Lattice coding and decoding achieve the optimal diversitymultiplexing tradeoff of MIMO channels," *IEEE Trans. Inform. Theory*, vol. 50, pp. 968–985, June 2004.

[24] Gersho, A. and Gray, R. M., *Vector Quantization and Signal Compression*. Boston, MA: Kluwer Academic Publishers, 1992.

[25] Ghogho, M., McLernon, D., Alameda-Hernandez, E., and Swami, A., "Channel estimation and symbol detection for block transmission using data-dependent superimposed training," *IEEE Signal Processing Letters*, vol. 12, pp. 226–229, Mar. 2005.

[26] Ghogho, M., McLernon, D., Alameda-Hernandez, E., and Swami, A., "SISO and MIMO channel estimation and symbol detection using data-dependent superimposed training," in *Proc. IEEE Int. Conf. Acoust. Speech. Signal Processing*, (Philadelphia, PA), Mar. 2005.

[27] Goldsmith, A., Jafar, S. A., Jindal, N., and Vishwanath, S., "Capacity limits of MIMO channels," *IEEE J. Selected Areas Commun.*, vol. 21, pp. 684–702, June 2003.

[28] Golub, G. H. and Loan, C. F. V., *Matrix Computations*. Baltimore: Johns Hopkins University Press, 3rd ed., 1996.

[29] Gore, D. A. and Paulraj, A. J., "Mimo antenna subset selection with space-time coding," *IEEE Trans. Signal Processing*, vol. 50, pp. 2580–2588, Oct. 2002.

[30] Gray, R. M. and Neuhoff, D. L., "Quantization," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2325–2383, Oct. 1998.

[31] Hochwald, B. M. and Marzetta, T. L., "Unitary space-time modulation for multiple-antenna communications in rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 46, pp. 543–564, Mar. 2000.

[32] Horn, R. A. and Johnson, C. R., *Matrix Analysis*. Cambridge University Press, 1985.

[33] Huang, D. and Letaief, K. B., "A reduced complexity coded OFDM system with MIMO antennas for broadband wireless communications," in *Proc. IEEE Global Telecommun. Conf.*, Nov. 2002.

[34] Huang, D., Letaief, K. B., and Lu, J., "A low complexity receive space diversity architecture for broadband OFDM systems," in *Proc. IEEE Vehicular Tech. Conf.*, pp. 1225–1229, Sept. 2002.

[35] Jafar, S. A. and Goldsmith, A., "On optimality of beamforming for multiple antenna systems with imperfect feedback," in *Proc. IEEE ISIT.*, pp. 321–321, June 2001.

[36] Jöngren, G., Skoglund, M., and Ottersten, B., "Combining transmit beamforming and orthogonal space-time block codes by utilizing side information," in *Proc. of 1st IEEE Sensor Array and Multichannel Signal Proc. Workshop*, (Boston, MA), pp. 153–157, Mar. 2000.

[37] Kay, S. M., *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice-Hall, 1993.

[38] Li, Y. G., Chuang, J. C., and Sollenberger, N. R., "Transmitter diversity for OFDM systems and its impact on high-rate data wireless networks," *IEEE J. Selected Areas Commun.*, vol. 17, pp. 1233–1243, July 1999.

[39] Li, Y., Georghiades, C. N., and Huang, G., "Iterative maximum-likelihood sequence estimation for space-time coded systems," *IEEE Trans. Commun.*, vol. 49, pp. 948–951, June 2001.

[40] Linde, Y., Buzo, A., and Gray, R. M., "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. COM-28, pp. 84–95, Jan. 1980.

[41] Liu, Y., Lau, K. N., Takeshita, O. Y., and Fitz, M. P., "Optimal rate allocation for superposition coding in quati-static fading channels," in *Proc. IEEE ISIT.*, (Lausanne, Switzerland), p. 111, June 2002.

[42] Love, D. J. and Heath, Jr., R. W., "Equal gain transmission in multiple-input multiple-output wireless systems," *IEEE Trans. Commun.*, vol. 51, pp. 1102–1110, July 2003.

[43] Love, D. J. and Heath, Jr., R. W., "Limited feedback precoding for spatial multiplexing systems using linear receivers," in *Proc. IEEE MILCOM*, (Boston, MA), Oct. 2003.

[44] Love, D. J., Heath, Jr., R. W., and Strohmer, T., "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2735–2747, Oct. 2003.

[45] Manton, J. H., "Optimization algorithms exploiting unitary constraints," *IEEE Trans. Signal Processing*, vol. 50, pp. 635–650, Mar. 2002.

[46] Marzetta, T. L. and Hochwald, B. M., "Capacity of a mobile multiple-antenna communication link in rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 543–564, Jan. 1999.

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

[48] Molisch, A. F., Win, M. Z., and Winters, J. H., "Reduced-complexity transmit/receive-diversity systems," *IEEE Trans. Signal Processing*, vol. 51, pp. 2729–2738, Nov. 2003.

[49] Mondal, B. and Heath, Jr., R. W., "Adaptive feedback for MIMO beamforming systems," in *Proc. IEEE 5th Workshop on Signal Processing Advances in Wireless Communications*, pp. 213–217, July 2004.

[50] Mondal, B. and Heath, Jr., R. W., "A lower bound on outage probability of limited feedback mimo beamforming systems," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, (Pacific Grove, CA), Nov. 2004.

[51] Mondal, B. and Heath, Jr., R. W., "Performance bounds for limited feedback mimo beamforming systems," in *Proc. of Allerton Conf. on Comm. Control and Comp.*, (Monticello, IL), Sept. 2004.

[52] Moustakas, A. L. and Simon, S. H., "Optimizing multiple-input single-output (MISO) communication systems with general gaussian channels: Nontrivial covariance and nonzero mean," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2770–2780, Oct. 2003.

[53] Mukkavilli, K. K., Sabharwal, A., Erkip, E., and Aazhang, B., "On beamforming with finite rate feedback in multiple-antenna systems," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2562–2579, Oct. 2003.

[54] Narula, A., Lopez, M. J., Trott, M. D., and Wornell, G. W., "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE J. Selected Areas Commun.*, vol. 16, pp. 1423–1436, Oct. 1998.

[55] Nefedov, N., Pukkila, M., Visoz, R., and Berthet, A. O., "Iterative data detection and channel estimation for advanced TDMA systems," *IEEE Trans. Commun.*, vol. 51, pp. 141–144, Feb. 2003.

[56] Okada, M. and Komaki, S., "Pre-DFT combining space diversity assisted COFDM," *IEEE Trans. Vehicular Technology*, vol. 50, pp. 487–496, Mar. 2001.

[57] Otnes, R. and Tüchler, M., "Iterative channel estimation for turbo equalization of time-varying frequency-selective channels," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1918–1923, Nov. 2004.

[58] Proakis, J. G., *Digital Communications*. New York, NY: McGraw-Hill, 3rd ed., 1995.

[59] Raghothaman, B., "Deterministic perturbation gradient approximation for transmission subspace tracking in FDD-CDMA," in *Proc. IEEE Int. Conf. Commun.*, vol. 4, pp. 2450–2454, May 2003.

[60] Raleigh, G. G. and Cioffi, J. M., "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, pp. 357–366, Mar. 1998.

[61] Roh, J. C. and Rao, B. D., "An efficient feedback method for MIMO systems with slowly time-varying channels," in *Proc. IEEE WCNC*, vol. 2, (Atlanta, GA), pp. 760–764, Mar. 2004.

[62] Scaglione, A., Stoica, P., Barbarossa, S., Giannakis, G. B., and Sampath, H., "Optimal designs for space-time linear precoders and decoders," *IEEE Trans. Signal Processing*, vol. 50, pp. 1051–1064, May 2002.

[63] Shamai, S. and Steiner, A., "A broadcast approach for a single-user slowly fading MIMO channel," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2617–2635, Oct. 2003.

[64] Simon, S. H. and Moustakas, A., "Optimizing MIMO antenna systems with channel covariance feedback," *Bell Laboratories Technical Memorandum*, 2002.

[65] Skoglund, M., Giese, J., and Parkvall, S., "Code design for combined channel estimation and error protection," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1162–1171, May 2002.

[66] Skoglund, M., Giese, J., and Parkvall, S., "Code design for combined channel estimation and error protection," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1162–1171, May 2002.

[67] Slimane, S. B., "A low complexity antenna diversity receiver for OFDM based systems," in *Proc. IEEE Int. Conf. Commun.*, pp. 1147–1151, 2001.

[68] Spall, J. C., *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. Hoboken, NJ: Wiley, 2003.

[69] Srivastava, A. and Klassen, E., "Bayesian and geometric subspace tracking," *Adv. Appl. Prob.*, vol. 36, no. 1, pp. 43–56, 2004.

[70] Steiner, A. and Shamai, S., "Hierarchical coding for a MIMO channel," in *IEEE Convention of Electrical and Electronics Engineers in Israel*, pp. 72–75, Sept. 2004.

[71] Stüber, G. L., *Principles of Mobile Communication*. Boston, MA: Kluwer Academic Publishers, 2nd ed., 2001.

[72] Tarokh, V., Jafarkhani, H., and Calderbank, A. R., "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1456–1467, July 1999.

[73] Tarokh, V., Seshadri, N., and Calderbank, A. R., "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, Mar. 1998.

[74] Telatar, E., "Capacity of multi-antenna gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[75] Vaswani, N., Roy-Chowdhury, A. K., and Chellappa, R., "'Shape activity': A continuous-state HMM for moving/deforming shapes with application to abnormal activity detection," *IEEE Trans. Image Processing*, vol. 14, pp. 1603–1616, Oct. 2005.

[76] Visotsky, E. and Madhow, U., "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2632–2639, Sept. 2001.

[77] WiMAX Forum, "Mobile WiMAX Part I: A technical overview and performance evaluation," tech. rep., WiMAX Forum, Feb. 2006.

[78] Yang, J. and Li, Y. G., "Low complexity OFDM MIMO system based on channel correlations," in *Proc. IEEE Global Telecommun. Conf.*, vol. 2, (San Francisco, CA), pp. 591–595, Dec. 2003.

[79] Yang, J. and Williams, D. B., "Transmission subspace tracking for mimo systems with low rate feedback," *submitted to IEEE Trans. Commun., revised*.

[80] Yang, J. and Williams, D. B., "MIMO transmission subspace tracking with low rate feedback," in *Proc. IEEE Int. Conf. Acoust. Speech. Signal Processing*, (Philadelphia, PA), Mar. 2005.

[81] Yeap, B. L., Wong, C. H., and Hanzo, L., "Reduced complexity in-phase/quadrature-phase M-QAM turbo equalization using iterative channel estimation," *IEEE Trans. Wireless Commun.*, vol. 2, pp. 2–10, Jan. 2003.

[82] Zheng, L. and Tse, D. N. C., "Communication on the grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inform. Theory*, vol. 48, pp. 359–383, Feb. 2002.

[83] Zheng, L. and Tse, D. N. C., "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1073–1096, May 2003.

[84] Zhou, G. T., Viberg, M., and McKelvey, T., "A first-order statistical method for channel estimation," *IEEE Signal Processing Letters*, vol. 10, pp. 57–60, Mar. 2003.

[85] Zhou, S. and Giannakis, G. B., "Optimal transmitter eigen-beamforming and space-time block coding based on channel mean feedback," *IEEE Trans. Signal Processing*, vol. 50, pp. 2599–2613, Oct. 2002.

[86] Zhou, S. and Giannakis, G. B., "Optimal transmitter eigen-beamforming and space-time block coding based on channel correlations," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1673–1690, July 2003.

[87] Zhou, S. and Giannakis, G. B., "Optimal transmitter eigen-beamforming and space-time block coding based on channel correlations," in *Proc. IEEE Int. Conf. Commun.*, vol. 1, (New York, NY), pp. 553–557, Apr. 2002.

VITA

Jingnong Yang was born in Taiyuan, Shanxi Province, P. R. China. He received his B.S. Degree in Electronic Engineering from Tsinghua University, Beijing, China in 1996 and his M.S. Degree in Electrical and Computer Engineering from Georgia Institute of Technology in 2004. He plans to receive his Ph.D. Degree in Electrical and Computer Engineering from Georgia Institute of Technology in 2006. His main theoretical interests lie in the realm of signal processing for wireless communications, with an emphasis on MIMO signal processing.