# 國 立 交 通 大 學

## 電子工程學系 電子研究所

碩 士 論 文

寬帶通信系統之序列設計

Performance Enhanced Broadband Communication Systems via Sequence Design

研 究 生：江清德

指導教授：馮智豪

中 華 民 國 九 十 九 年 七 月

# 寬帶通信系統之序列設計

# Performance Enhanced Broadband Communication Systems via Sequence Design


研 究 生：江 清 德　　Student: Chin-Te Chiang

指導教授：馮 智 豪 博士　　Advisor: Dr. Carrson C. Fung

**國 立 交 通 大 學**

**電 子 工 程 學 系　電 子 研 究 所**


**碩 士 論 文**

A Thesis

Submitted to Department of Electronics Engineering & Institute of Electronics
College of Electrical and Computer Engineering
National Chiao Tung University
in Partial Fulfillment of the Requirements
for the Degree of Master
in
Electronics Engineering

July 2010

Hsinchu, Taiwan, Republic of China

中華民國九十九年七月

# 寬帶通信系統之序列設計

研究生：江清德　　指導教授：馮智豪 博士

國立交通大學

電子工程學系 電子研究所碩士班

## 摘要

在寬頻通訊系統中，多輸入多輸出技術已經被證明可以提供高速以及增進頻帶效率的重要技術。寬頻通訊系統標準例如 IEEE 802.16e、3GPP LTE 和 LTE Advanced 都採用多天線技術。而這些先進的通訊系統十分依賴適當設計後的訊號，來獲得正確的通道狀態並且執行精確的同步。但因傳送這些非資料訊號，例如像保護位元和指標，使得無線通訊系統會蒙受頻率效率的降低。為了達到多輸入多輸出系統所保證的資料傳輸率增益，準確的空間相關訊息是非常重要的。此外，使用正交分頻多工技術雖然有許多益處像是簡單的等化器設計，但高的峰值因數往往會發生，因而降低了發射機的功率效率。當多輸入多輸出系統和正交分頻多工系統結合時，由於越來越多的傳送端，高峰質因數更容易發生並嚴重地對系統造成影響。

在這篇論文裡頭，利用仿射編碼來設計耐用(robust)疊加訓練序列，即使只有不確定性存在的空間相關訊息，此序列依舊可以幫助預測空間相關的多輸入多輸出通道

狀態。序列是直接加到資料訊號上，不會造成頻率效率的損失。此設計不需要準確的空間相關訊息(矩陣)，在這篇論文中，我們也證明此耐用設計優於之前被提出用來預估空間相關多通道的方法，例如 RMMSE 和 LS-RMMSE。這個耐用設計可以用投影凸集的迭代算法來解決，只要訓練序列的初始化是滿秩的，此迭代保證會收斂。當通道是不空間相關時，我們證明出耐用設計和 RMMSE 是漸近相同的。除此之外，我們也提出一個功率分配方法來達到最佳的資料檢測效能。

由於使用多輸入多輸出正交多頻分工系統，我們提出了疊加序列來執行耐用通道預測並且來降低峰值因數。我們提出了載波仿射編碼來降低高峰質因數，而接收端不需要知道任何的額外訊息來解調資料訊號。相對於之前已知的技術，我們提出的方法降低了很多的額外傳輸開銷，也增進資料檢測效能。即使我們提出的設計也要傳送額外的訊號，但此多餘的訊號可以小到每個載波只要一個符號。此設計可以讓設計者很容易的權衡資料檢測和降低峰值因數的效能。在模擬結果中，我們的方法明顯在降低峰值因數和傳輸效率上都優於 Tone Reservation。

**Performance Enhanced Broadband Communication Systems via**

**Sequence Design**

Student: Chin-Te Chiang Advisor: Dr. Carrson C. Fung

Department of Electronic Engineering
Institute of Electronics
National Chiao Tung University

## Abstract

MIMO technology has proven to be the key enabler of high-speed, bandwidth efficient broadband communication systems such as IEEE 802.16e, 3GPP LTE and LTE-Advanced. Similar to many traditional systems, these advanced communication systems rely heavily on proper signaling in order to obtain correct channel state information and perform precise synchronization. Unfortunately, traditional signaling methods can incur a loss of spectral efficiency due to transmission of overhead data such as preamble, guard bits and pilots. Moreover, accurate spatial correlation information is crucial in achieving the theoretical capacity gain promised by MIMO. Furthermore, with the use of OFDM, high PAPR is often incurred, thus lowering the power efficiency at the transmitter. The problem is worsened when OFDM is combined with MIMO as more RF chains are required for transmission.

In this thesis, a new signaling scheme is proposed for spatially correlated MIMO channels which exploits affine precoding to produce robust superimposed training sequence such that CSI can be accurately obtained even when uncertainty in the spatial correlation matrix exists. The sequence is algebraically added to the data such that there

is no loss of spectral efficiency. The proposed scheme does not require accurate knowledge about the spatial correlation matrix and it is shown to outperform previously proposed robust correlated MIMO channel estimators such as relaxed MMSE (RMMSE) and least-squares-RMMSE (LS-RMMSE). A solution for the sequence can be obtained easily by using a projection on convex sets based iterative algorithm which is guaranteed to converge as long as the training sequence matrix is initialized to have full rank. Furthermore, it is shown that the proposed scheme is asymptotically identical to the RMMSE based schemes when the MIMO channel is spatially uncorrelated. A power allocation scheme is also proposed that can maximize the detection performance.

Next, a joint superimposed sequence design is proposed to jointly perform robust channel estimation and lower the PAPR of MIMO-OFDM systems. A per-tone affine precoding technique is proposed to reduce the PAPR such that no side information is required to be transmitted for the removal of the sequence at the receiver. This is in contrast to previous known techniques which incurs a large amount of transmission overhead, or can dramatically increase the BER. Furthermore, some of these techniques are based on heuristics that cannot optimally lower the PAPR. Even though redundant information has to be sent, this can be as small as 1 symbol/subcarrier. Furthermore, the proposed design allows the designer to easily trade off between BER and PAPR reduction performance. Simulation results have shown that the proposed scheme outperforms the tone reservation scheme not only in PAPR reduction but also in transmit efficiency.

# Acknowledgement

It would be impossible finishing my thesis without support and help from advisor Prof. Carrson C. Fung, lab-mates, friends and my family.

First, I would like to thank Prof. Carrson C. Fung for his patient guidance and constant encouragement in the past three years, which had helped me overcome my hardships and enabled me to finish my present work. I would also like to appreciate my lab-mates Hung-Chih, Yin-Ray, Feng-Chin, Chong-You, Chen-Yen and et al for giving me advice on research and living. Besides, Chi Chun, my lovely girlfriend, always accompanies me and shares my feelings.

Finally, I want to express my deepest gratitude to my beloved parents Shan-Liang Chiang and Ming-Chu Chien. Their endless love and support makes who I am today.

# Contents

# Abbreviations and Notations

| | |
|---|---|
| **ADC** | Analog-to-Digital Converter |
| **AWGN** | Additive White Gaussian Noise |
| **BER** | Bit Error Rate |
| **CP** | Cyclic Prefix |
| **CCDF** | Complementary Cumulative Distribution Function |
| **CSI** | Channel State Information |
| **FFT** | Fast Fourier Transform |
| **IFFT** | Inverse Fast Fourier Transform |
| **LS** | Least-Squares |
| **MIMO** | Multiple-Input and Multiple-Output |
| **MSE** | Mean Square Error |
| **MU** | Multiple Users |
| **OFDM** | Orthogonal Frequency-Division Multiplexing |
| **PAPR** | Peak-to-Average-Power Ratio |
| **PLS** | Partial Transmit Sequence |
| **QPSK** | Quadrature Phase Shift Keying |
| **SISO** | Single-Input and Single-Output |
| **SIT** | Superimposed Training Sequence |
| **SLM** | Selective Mapping |
| **SNR** | Signal-to-Noise Ratio |

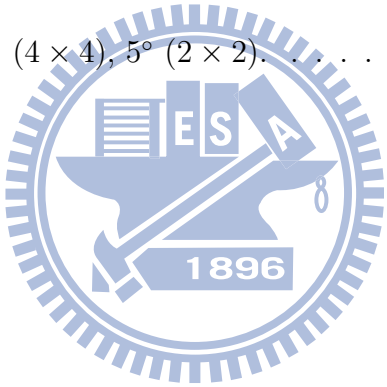
| | |
|---|---|
| **SU** | Single User |
| **STBC** | Space-Time Block Code |
| **STC** | Space-Time Coder |
| **TI** | Tone Injection |
| **TR** | Tone Reservation |

| | |
|---|---|
| $\mathbf{A}$ | upper bold face letter indicate matrix |
| $\mathbf{a}$ | lower bold face letter indicate vector |
| $A^H$ | $\mathbf{A}$'s Hermitian |
| $A^T$ | $\mathbf{A}$'s transposition |
| $A^*$ | $\mathbf{A}$'s conjugation |
| $E[\cdot]$ | expectation |
| $\mathcal{N}(A)$ | the nullspace of $\mathbf{A}$ |
| $vec(\mathbf{A})$ | vectorization of $\mathbf{A}$ |
| $tr(\mathbf{A})$ | the trace of the matrix $\mathbf{A}$ |
| $diag(x)$ | a diagonal matrix with $x$ on its main diagonal |
| $\mathbf{I}_N$ | $N \times N$ identity matrix |
| $\mathbf{0}_{M\times N}$ | $M \times N$ all zero matrix |
| $\bigotimes$ | the Kronecker product |
| $\bigodot$ | the element-wise products of two matrices |
| $\|\mathbf{A}\|_F$ | the Frobenius norm of the matrix $\mathbf{A}$ |
| $\|\mathbf{a}\|_\infty$ | the norm to the infinity of the vector $\mathbf{a}$ |
| $\mathbf{A}(m_i,:)$ | a row vector extracted from the $m_i^{th}$ row of $\mathbf{A}$. |
| $\mathbf{A}(m_i:m_s:m_f,:)$ | a matrix whose elements are taken from the $m_i^{th}$ row to the $m_f^{th}$ row of $\mathbf{A}$ with $m_s$ increment. |

# List of Figures

# List of Tables





# Chapter 1

# Introduction

## 1.1 Research Motivation and Contribution

Rapidly growing demand for wireless services requires wireless communication systems to have faster transmission speed and higher throughput. In the last decade, SU-MIMO has drawn a great deal of attention since it has potential to provide spatial multiplexing gain and achieve higher diversity , which scale linearly with the number of antennas, without sacrificing spectral efficiency [3–5]. To realize such gains, CSI must be obtained accurately. As compared to the SISO channel, the MIMO channel contains more unknown coefficients. Therefore, it is more difficult to estimate the MIMO channel than the SISO one. Although techniques such as differential space-time coding [8, 9] and differential orthogonal space-time block coding [10, 11] have been proposed to blindly demodulate and decode the received signal without CSI, this is done by sacrificing both performance (compared to coherent techniques) and spectrum efficiency. Coherent detection is thus widely used in current systems where CSI is usually obtained using time-multiplexed pilot symbols. This, however, reduces the transmission efficiency, especially in cases where the channel is undergoing fast fading due to the fact that pilot signals have to be placed more frequently [2]. Moreover, the theoretic capacity gain promised by SU-MIMO can only

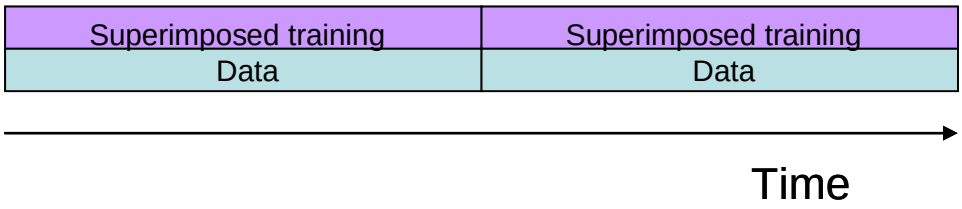



Figure 1.1: structure of superimposed training sequence

be achieved if the the channel is spatially uncorrelated . In scenarios where channels are spatially correlated, this correlation has to be accurately estimated in order to maximize MIMO capacity gain. Increased antenna correlation can be attributed to reduction in antenna spacing or angular spread, which is caused by lack of a rich scattering environment around the transceiver. Hence, some degree of spatial correlation will be experienced at the transmitter and/or receiver.

Many techniques have been proposed to tackle the problem of correlated channel estimation [14,15,23,24,30,55] with some techniques focusing on training sequence design [14, 23, 55] while others dealt with channel estimator design [15, 23, 24, 30]. Despite the tremendous focus that has been placed on this problem, most works have ignored the effect that inaccurate spatial correlation has on the channel estimate. This inaccurately, as shall be shown in the sequel, does have a dramatic impact on the accuracy of the CSI. In this thesis, a robust sequence design is proposed which accounts for the mismatch between the actual and estimated spatial correlation. The sequence is then applied to the MMSE estimator to estimate the MIMO channel. For ease of presentation, such an estimator shall be called RoMMSE estimator. Moreover, the sequence is designed in the context of superimposed training sequence, which was recently proposed by [16] in order to tackle the spectral efficiency problem. Rather than using dedicated timeslots, the SIT sequence is algebraically added to the information bearing signal as illustrated in Figure 1.1. Since the training sequence is overlaid onto the information-bearing signal, it allows for greater spectral efficiency as compared to conventional systems that use time-multiplexed pilot

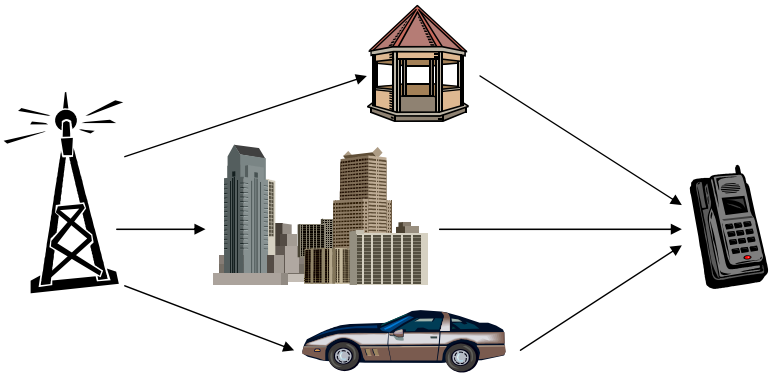


Figure 1.2: multi-path fading in wireless channel

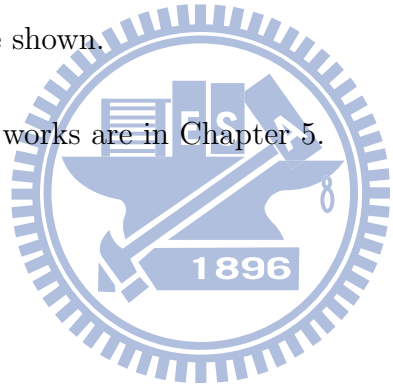
symbols.

Despite having multiple antennas, the fading channel can severely hamper the performance of a wireless system by scattering the transmitted signal, thereby causing the receiver to receive multiple copies of the same signal as illustrated in Figure 1.2. OFDM, which is one of the most popular multi-carrier modulation techniques due to its low computation complexity, offers immunity to the multi-path fading channels by allowing the signal to transmit through multiple flat-fading channels. This is accomplished by using FFT which explain its computation efficiency. Thus, combining MIMO and OFDM offers an attractive solution toward achieving low-complexity high-throughput communication systems . Unfortunately, due to the use of the IFFT at the transmitter, the transmitted signal is no longer constrained, thus causing the signal to have a high peak-to-average power ratio. This demands the use of highly linear and inefficient power amplifiers. This problem is exasperated in MIMO-OFDM systems as multiple RF chains are deployed. Hence, a superimposed sequence is proposed to be added to the data sequence in order to mitigate the PAPR. Unlike previous proposed techniques such as selective mapping and partial transmit sequence [51], the proposed design can easily trade off between bit error rate and throughput in order to increase the effectiveness of the PAPR reduction. This is particularly important as different applications have different throughput, error rate and delay sensitivity requirements.

## 1.2 Thesis Organization

The thesis mainly focuses on the sequence design to enhance system performance. It is organized as follows :

- An introduction to the MIMO and the MIMO-OFDM systems is given in Chapter 2. A review of MIMO channel estimation and PAPR reduction are also included.
- Robust superimposed sequence design for spatially correlated MIMO channels is proposed in Chapter 3. MIMO system with affine precoder is first introduced, followed by problem formulation and simulation results are shown later in this chapter.
- PAPR reduction of MIMO-OFDM using optimal superimposed sequence is proposed in Chapter 4. Detailed system model, problem formulation and simulation results for PAPR reduction are shown.
- Conclusions and future works are in Chapter 5.



## 1.3 Publications

Conference:

Chin-Te Chiang and C.C. Fung, "Robust training sequence design for spatially correlated MIMO channel estimation using affine precoder," *Proc. of the Intl. Conf. on Communications*, May 2010.

Journal:

Chin-Te Chiang and C.C. Fung, "Robust training sequence design for spatially correlated MIMO channel estimation, " *to be submitted to the IEEE Transactions on Vehicular Technology.*

# Chapter 2

# Background

Required background information is overviewed in this chapter. Two wireless communication systems and repective problems are introduced.

- An introduction of MIMO system and MIMO channel estimation is presented first.
- MIMO-OFDM system and PAPR reduction are discussed in Chapter 2.2.

## 2.1 MIMO System

MIMO refers to multiple antennas both at transmitter and receiver. The schematic illustration of a MIMO system with $N_t$ transmit antennas and $N_r$ receive antennas is shown in Figure 2.1 . The spatial dimension is exploited to increase the transmission rate and also offer better reliability. With spatial multiplexing, the transmission rate can be increased by sending multiple data streams with multiple antennas in parallel. On the other hand, data for transmission can also be coded using STC techniques such as space-time trellis coding and space-time block coding. The idea of STC is transmitting multiple and redundant copies of a data stream to achieve transmit diversity. The coding techniques usually applied to MIMO systems with transmit diversity.

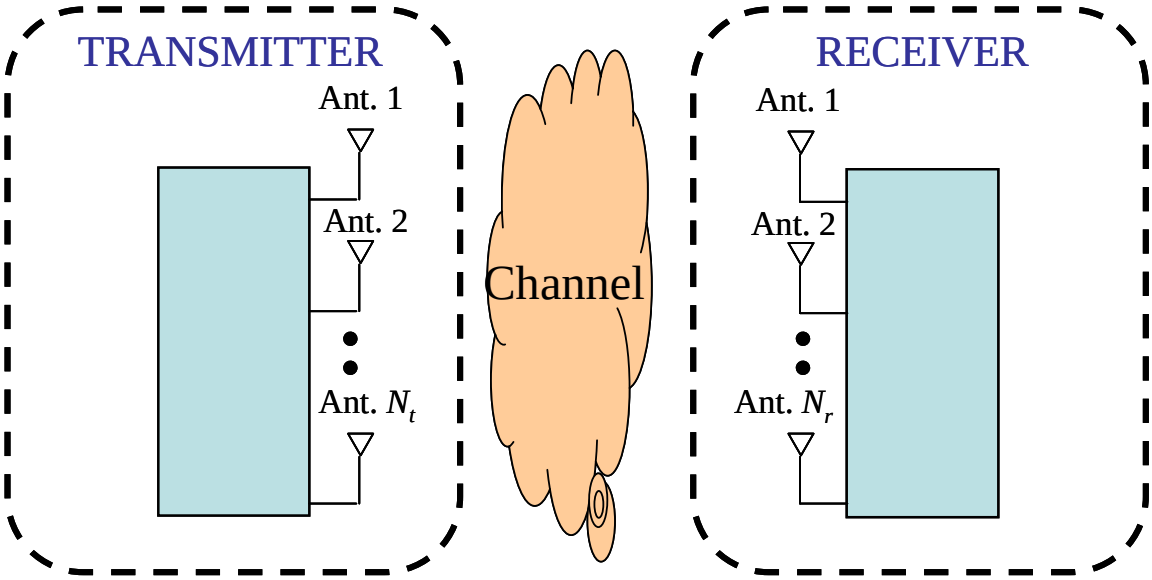



Figure 2.1: Simple schematic illustration of a SU-MIMO system

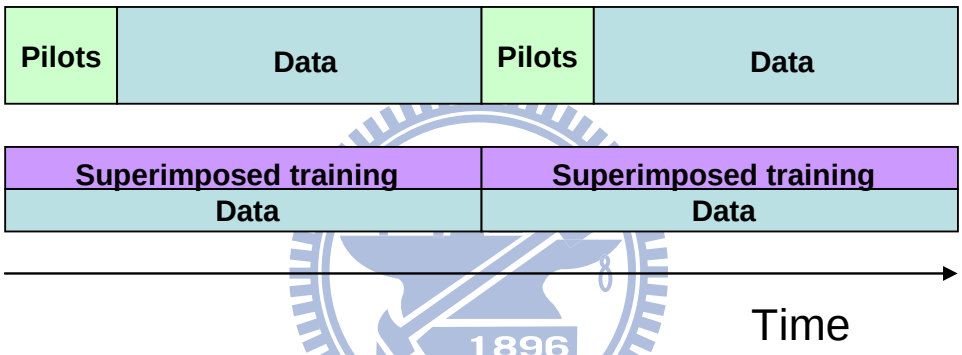



Figure 2.2: comparison of time-multiplexed pilots and superimposed training sequence

### 2.1.1 MIMO Channel Estimation

SU-MIMO [1] has brought tremendous research effort devoted in this area in maximizing diversity and spatial multiplexing gain [6,7] over the last decade. Perfect knowledge about the CSI is usually assumed in the work. But in practice, CSI must be estimated accurately to achieve such gain. Coherent detection is widely used in current MIMO system where CSI is usually estimated using time-multiplexed pilot symbols. Time-multiplexed pilots decrease the transmission rate because it occupies a certain number of times slots. SIT

[1] In this work, only SU-MIMO systems are dealt with as multiuser MIMO systems have different scenarios and performance compared to SU-MIMO systems.

sequence is arithmetically overlaid on the information-bearing data, it is transmitted using all the time slots as the data. The comparison of the two schemes is shown in Figure 2.2. SIT sequence increases the transmission rate as compared to the time-multiplexed pilots. SIT can be extracted at the receiver by using first-order statistics [17,18], but information-bearing data is always viewed as an interference for channel estimation since they share the same time slots. In order to have a better separation of information-bearing data and SIT at the receiver, affine precoder post-multiplies the information bearing data in [54]. It is independent from the channel matrix due to the post multiplication. It also helps the receiver remove the unwanted received signals for channel estimation and so as for data detection. Removal can be done by post-multiplied with well-designed decoupling matrices. A SIT sequence design for spatially correlated MIMO channel using affine precoder is proposed in [55]. Optimal SIT sequence, which is dependent on the spatial correlation and the signal-to-noise ratio, can minimize the mean square error of channel estimation.

### 2.1.2 Spatial Correlation of MIMO channels

Under the ideal assumption of independent and identically distributed(i.i.d.) wireless channel coefficients, the optimal training sequence using MMSE channel estimator has the orthogonal property [29]. However, the assumption does not hold for several conditions such as little spacing between the antennas and small angular spread. Therefore, MIMO channel is spatially correlated. In other words, spatial correlation matrix $\mathbf{R_{hh}}$ of the vectorized channel $\mathbf{h} = vec(\mathbf{H})$ is no more an identity matrix multiplied by the variance of the coefficient.

$$\mathbf{R_{hh}} = E[\mathbf{h}\mathbf{h}^H] \in \mathbb{C}^{N_t N_r \times N_t N_r} \tag{2.1}$$

There are several analytical models to describe the spatial correlated MIMO channel. Kronecker model is the most popular since its simple analytic expression of the correlation

matrix. The model has been widely used for theoretical analysis of MIMO systems [31]. Drawback of the kronecker model is the assumption of physical separation between the transmitter and receiver. Correlation is modeled at the transmitter and receiver sides separately. It neglects the interdependency between the two sides, which is not accurate when the transmitter and receiver are close.

$$\mathbf{R} = \mathbf{\Sigma}_t \otimes \mathbf{\Sigma}_r \tag{2.2}$$

$$\mathbf{H} = \mathbf{\Sigma}_r^{\frac{1}{2}} \mathbf{H}_w \mathbf{\Sigma}_t^{\frac{1}{2}}, \tag{2.3}$$

where $\mathbf{H}_w$ is the random matrix with i.i.d. zero mean complex normal entries with unit variance. $\mathbf{\Sigma}_r^{\frac{1}{2}} \in \mathbb{C}^{N_r \times N_r}$ and $\mathbf{\Sigma}_t^{\frac{1}{2}} \in \mathbb{C}^{N_t \times N_t}$ are the Cholesky factorizations of spatial correlation at the receiver and transmitter, respectively. In order to capture the joint correlation between the transmitter and receiver, Weichselberger model is proposed in [33]. It describes the joint spatial correlation between the two link ends using the coupling matrix $\mathbf{\Omega}$. The spatial correlated MIMO channel can be expressed as,

$$\mathbf{H}_{Weichselberger} = \mathbf{U}_T(\tilde{\mathbf{\Omega}} \bigodot \mathbf{G})\mathbf{U}_R^T. \tag{2.4}$$

$\tilde{\mathbf{\Omega}}$ denotes the element-wise square root of $\mathbf{\Omega}$. $\mathbf{G}$ is an i.i.d random matrix with zero mean and unit variance. $\mathbf{U}_T$ and $\mathbf{U}_R$ denotes the eigenvector matrices of spatial correlation at transmitter and receiver respectively. For correlated MIMO channel estimation, kronecker model is usually considered due to the ease of analysis. It is also used in the thesis.

### 2.1.3 Previous Work on Correlated MIMO Channel Estimation

Many techniques have been proposed to tackle the problem of correlated MIMO channel estimation. [12] proposed using a state-space approach to estimate and track time-varying correlated MIMO channels, where the channel correlation matrix is estimated from the received data and treated as part of the state variable. In [13], a precoder assisted linear

MMSE estimator was proposed to estimate the channel. In [14], two channel estimators were derived under the MMSE and conditional mutual information criteria by exploiting the virtual channel representation. Unfortunately, there is no closed-form solution, thus the solution has to be computed numerically. [15] derived another MMSE based channel estimator using structured correlation, which allows it to obtain better MSE performance than unstructured based MMSE estimator. One major drawback that is shared among these estimators is that they require exact knowledge about the spatial correlation in order to outperform channel estimators that do not take such correlation into account. Another disadvantage they shared is that they were all derived under the premise that time-multiplex pilot symbols are used, which can drastically reduce transmission efficiency, especially in cases where the channel is undergoing fast fading.

To bypass the second problem, [16] has proposed a channel estimation algorithm which uses a SIT sequence that is arithmetically added into the transmitted signal; this frees up valuable time slots that were previously used by time-multiplex pilot symbols. The training sequence can also be used to deal with problem of synchronization [19]. Improved channel estimation algorithms based on these training sequence have since appeared in literature [17–19]. The sequence itself can be extracted at the receiver by using first-order statistics [17,18] or by using affine precoding [22,54,55]. However, the effectiveness of these algorithms still hinges on acquiring accurate estimates of the spatial correlation, making these methods somewhat infeasible in real situations. To combat against this problem, the relaxed MMSE (RMMSE) and least-squares RMMSE (LS-RMMSE) algorithms have recently been proposed by [23] that can circumvent the dependency on the correlation matrix by using diagonal loading while [24] has proposed a different approach by using basis expansion.

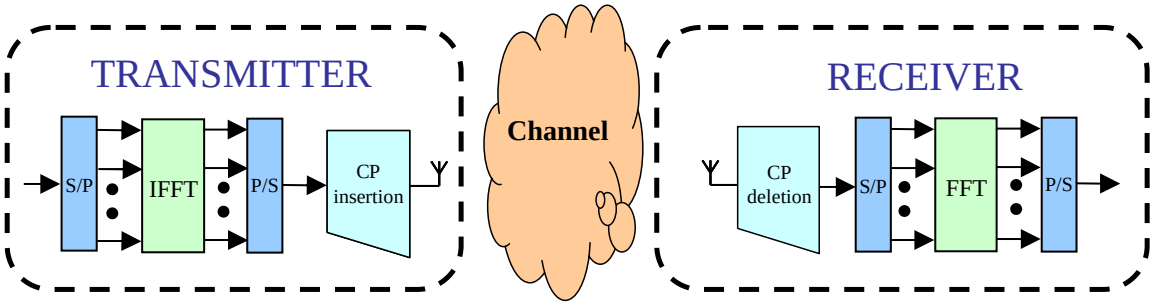



Figure 2.3: Block diagrams of OFDM

## 2.2 MIMO-OFDM System

Mutil-path fading and high transmission rates make the wireless channels to be frequency selective, which is the same case for the MIMO channel. Frequency selective channel brings the unwanted inter-symbol interference and inter-carrier interference. Therefore, OFDM is used to combat the mentioned problems under assumption that the length of cyclic prefix is long enough. Although OFDM is a powerful technique, there are still disadvantages such as sensitivity to synchronization error and high peak-to-average power ratio. The use of multiple antennas is a trend since it provides diversity and spatial multiplexing gain. That is, parallel OFDM transmission, denoted as MIMO-OFDM, is the candidate for the next-generation wireless communication.

### 2.2.1 Concepts of OFDM and PAPR

OFDM, which is one of the multi-carrier modulation, offers immunity to the frequency selective channels and also guarantees for high data transmission efficiency. Hardware implementation can be easily realized using FFT techniques. That is, it has been adopted in many wireless communication standards, such as IEEE 802.11a, IEEE 802.16e, 3GPP LTE and the LTE-Advanced. . OFDM system can turn the frequency selective channels into parallel frequency flat subchannels when the CP is greater than or equal to the order of the channel. Figure 2.4 shows the structure of CP, which prefixes a repetition of the

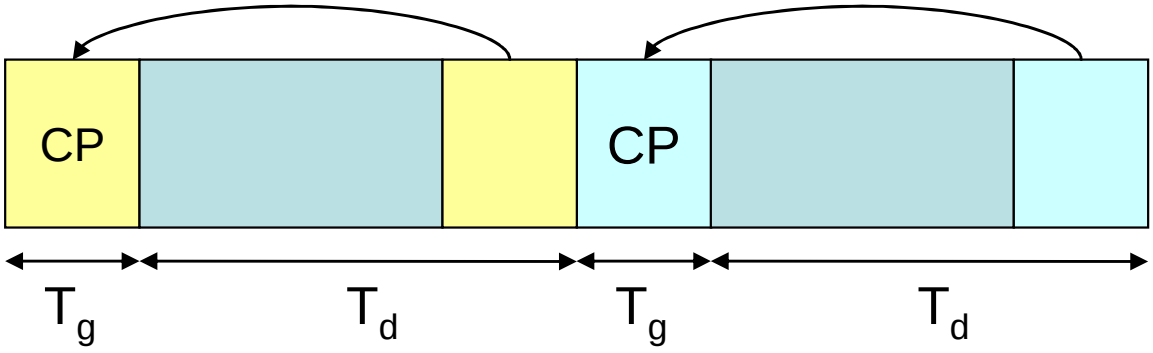



Figure 2.4: cyclic preix

tail part of the OFDM symbol. With the CP, linear convolution of the equivalent channel, which is composed of the CP insertion, removal and the frequency selective channels, has the circulant property. The property is favored since it makes the equivalent channel be diagonalized by the IFFT and FFT blocks, which makes the frequency selective channel to be parallel frequency flat.

Since IFFT block transforms the signals from frequency domain into time domain, sum of the Gaussian-like time-domain waveforms contributes to the high PAPR. High PAPR does not only decrease the efficiency of the power amplifier but also introduce the out-of-band noise. Definition of PAPR for the signal $s(t)$ is defined as,

$$PAPR\,(s(t)) = \frac{\max |s_(t)|^2}{E(|s_(t)|^2)}. \tag{2.5}$$

### 2.2.2 Structure of MIMO-OFDM system

In order to have a higher transmission rate, OFDM can be combined with MIMO. Consider a MIMO-OFDM system with $N_t$ transmit antennas and $N_r$ receive antennas. A simple schematic presentation is shown in Figure 2.5. The input data of the serial to parallel can be the modulation using space time frequency block codes or space frequency block codes techniques. The techniques mainly map information symbols to antennas and tones as a way for using both spatial and frequency diversity. The frequency-selective MIMO channels can be further flatten as in OFDM system. At the receiver, diversity from

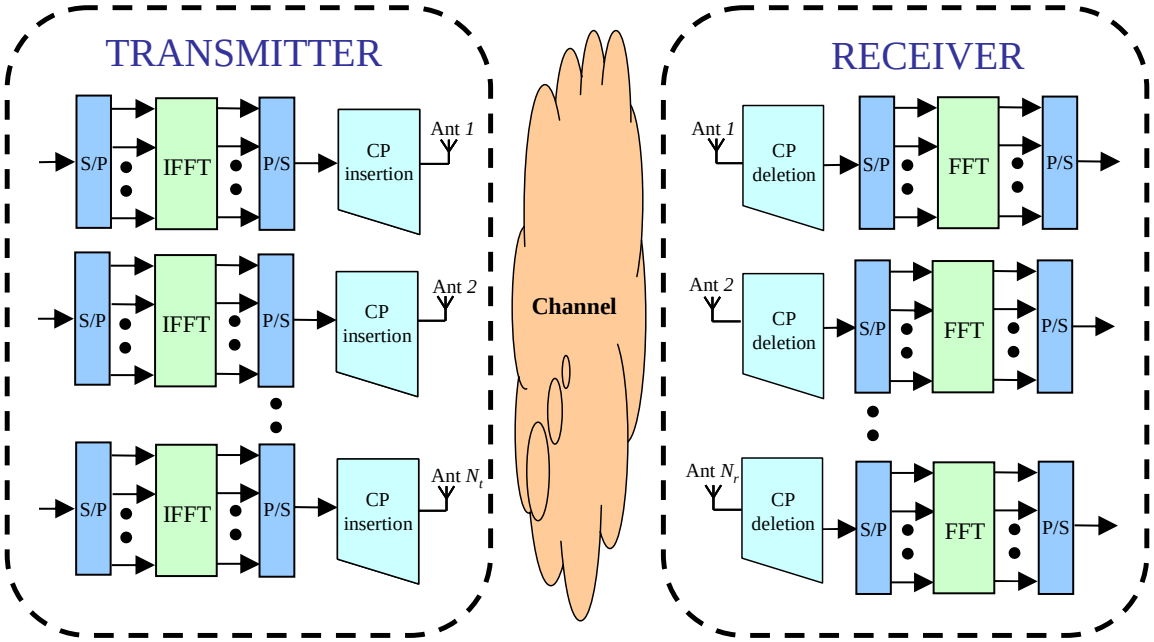



Figure 2.5: MIMO-OFDM system

multiple antennas can boost the data detection performance.

Although MIMO-OFDM have the advantages from both MIMO and OFDM, it still suffers from the high PAPR. Definition of the PAPR in MIMO-OFDM system is more general as compared to the one introduced in equation . It is defined as the maximum PAPR among all the transmit antennas [53].

$$PAPR\,(s_l(t)) = \frac{\max|s_l(t)|^2}{E(|s_l(t)|^2)},$$

$$PAPR_{MIMO} = \max_{l=1,2\cdots N_t} PAPR(s_l(t)) \tag{2.6}$$

where $l$ is the index for the transmitter. $s_l(t)$ denotes the signal in the time domain at antenna $l$. It can also describe the PAPR of the SISO-OFDM system.

### 2.2.3 Previous Works on PAPR reduction

It is well known that nonlinearity in high power amplifier causes distortion in the transmit signal. Such distortion can lead to undesirable spectral regrowth, thus interfering with signals in the neighboring subcarriers. Large amount of distortion caused by high PAPR

can cause in-band self-interference, which increases bit error rate. It is customary for power amplifiers to operate with a certain power backoff, which is defined as the ratio of maximum saturation output power to lower average output power [34]. However, such backoff schemes lowers the efficiency of the power amplifier and increases overall power consumption. Such problem can be avoided by employing clipping [35], it can be implemented by setting a saturation level. Once the signal exceeds the threshold, the amplitude is set as the level without changing the phase.

$$B(s(t)) = \begin{cases} s(t), & |s(t)| \leq A, \\ Ae^{j\theta(s(t))} & |s(t)| > A \end{cases} \tag{2.7}$$

$B$ is the clipping operation of the signal $s(t)$. A is the amplitude threshold. Clipping does not decrease the transmission rates, but its distortion contributes to the bit-error rate degradation and also brings out-of-band noise. A blocking coding scheme was proposed in [37] in which codewords that cause high PAPR are avoided and are instead coded with a different set of (longer) codewords. Since the signal is not distorted, such technique does not increase the BER. Unfortunately, it decreases the spectral efficiency and requires changes in the transmit frame structure to allow error-free decoding at the receiver. Distortionless techniques are prefered, but there is still drawbacks, which leads to the low transmission efficiency and high computation complexity. Tone reservation and tone injection [47] techniques are data-dependent methods which adds signal to the information-bearing signal in order to lower the PAPR. Tone reservation in Figure 2.6 reserves certain subcarriers such that they can be set to optimal values to minimize the PAPR without affecting the information-bearing subcarriers. Unlike the TI method, the TR method does not distort the original signal as the added signal is injected into set of subcarriers that have been reserved for PAPR reduction. This, of course, lowers the spectral efficiency of the system. TI avoids this problem by increasing the constellation size and injecting extra data into the new constellation points which tend to decrease the


Figure 2.6: tone reservation, the black blocks are the reserved frequency subcarriers

PAPR. Unfortunately, the injected data occupy the same subcarrier as the information-bearing signal which may adversely affect BER performance. The added signal may also increase transmit power. The TR method was extended by [48] such that the power injected into the reserved subcarriers for PAPR reduction is formulated as a power allocation problem which can be efficiently solved using linear programming. The active constellation extension(ACE) method [49] exploits the constant modulus structure by dynamically extending the some of the other signal constellation point such that the PAPR of the information-bearing signal can be reduced. This technique does not increase the BER and no side information is required to be sent. However, the degree of PAPR reduction is inversely proportional to the constellation size of the modulation, thus limiting its ability to lower PAPR in systems employing high order modulation. Furthermore, similar to the TI method, it may also increase the transmit signal power. But side information is still needed for data decoding. A more extensive treatment of previous proposed PAPR reduction techniques are given in [51, 52]. Advantages and disadvantages of the mentioned techniques are discussed in several criterion such as distortion, power increase, data rate loss and implementation complexity. The table 2.2.3 shows the comparison in different criterion. The techniques mentioned can be easily extended from SISO-OFDM to MIMO-OFDM, but it has not taken use of the spatial domain. A PAPR reduction scheme specifically designed for MIMO-OFDM systems called cross-antenna rotation and inversion, or CARI, was proposed in [50]. The technique perform subblockwise rotation and inversion across all antennas for subcarriers to achieve the PAPR reduction. The tech-

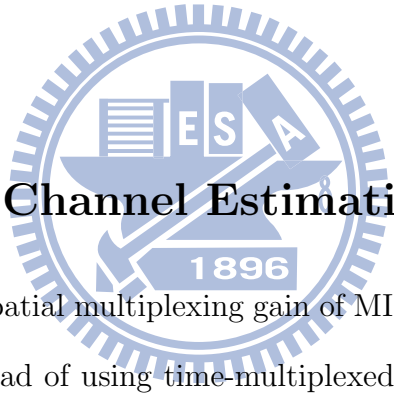
Table 2.1: comparison of different PAPR reduction techniques

| | distortion | power increase | data rate loss |
|---|---|---|---|
| clipping | Yes | No | No |
| TR/TI | No | Yes | Yes |
| PTS/SLM | No | No | Yes |
| ACE | No | Yes | No |

nique utilizes additional degrees of freedom in the spatial domain to decrease the PAPR. But side information is needed for data decoding. Nevertheless, every PAPR technique has its own advantages and disadvantages, it has no optimal solution for all multicarrier transmission systems.

# Chapter 3

# Robust Training Sequence Design for Spatially Correlated MIMO Channels



## 3.1 Overview of Channel Estimation

To realize the diversity and spatial multiplexing gain of MIMO system, channel CSI must be obtained accurately. Instead of using time-multiplexed pilots, superimposed training sequence is applied for channel estimation since its higher transmission efficiency. As compared to the SISO channel, the multiple-input MIMO channel contains more unknown coefficients to be estimated, which makes it more difficult. In MIMO system, small angular spread and little spacing between antennas are usually seen due to the small device and poor scattering environment, which cause the spatially correlated MIMO channel. Many algorithms have been discussed in section 2.1.3 to estimate the spatially correlated MIMO channels. However, the effectiveness of these algorithms still hinges on acquiring accurate estimates of the spatial correlation, making these methods somewhat infeasible in real situations.

A approach to the problem is proposed in this chapter by designing a superimposed training sequence that is robust toward spatial correlation uncertainty. The proposed design exploits the affine precoder scheme proposed in [26, 55] to extract the training sequence for channel estimation. System models and derivation of the iterative algorithm are discussed in details. Simulation results shows that the proposed scheme performs extremely well against estimator in [55] which does not take into account the spatial correlation estimate error. Moreover, the RoMMSE estimator also outperforms the RMMSE and LS-RMMSE estimators when the MIMO channel is spatially correlated. Finally, the RoMMSE estimator will also be compared to the RMMSE and LS-RMMSE estimators for uncorrelated MIMO channels in which it is shown that the three estimators perform almost identically and that they are asymptotically equivalent.

## 3.2 Methodology

### 3.2.1 System Model

The system model used in [55] is adopted herein. For the sake of completeness, the model will also be described in the sequel. Consider a spatially correlated flat-fading MIMO channel with $N_t$ transmit and $N_r$ receive antenna, as shown in Figure 3.1. The information-bearing signal vector is denoted as $\mathbf{u}(k) = \begin{bmatrix} u(kN_s) & u(kN_s+1) & \cdots & u(kN_s+N_s-1) \end{bmatrix}^T$, where $k$ is the block index and $N_s$ denotes the block size. Each block of the signal is encoded using a space-time block coder (STBC), which can be used to increase transmit diversity or multiplexing gain [25]. The STBC has $N_t$ number of output vectors, with each vector containing $K \geq N_t$ symbols as full rate STBC is assumed. This can be represented in matrix form as $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_{N_t} \end{bmatrix}^T \in \mathbb{C}^{N_t \times K}$, where $\mathbf{x}_i \in \mathbb{C}^K$, for $i = 1, 2, \ldots, N_t$, denotes the $i^{th}$ output vector. Each vector is then fed into the precoder $\mathbf{P} = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \cdots & \mathbf{p}_K \end{bmatrix}^T \in \mathbb{C}^{K \times (K+L)}$ which adds $L \geq N_t$ redundant symbols to each

block of signal, resulting in the output signal vector $\mathbf{d}_i \in \mathbb{C}^{K+L}$, for $i = 1, 2, \ldots, N_t$. All $N_t$ output of the precoder can be represented in matrix form as

$$\mathbf{D} \triangleq \begin{bmatrix} \mathbf{d}_1^T \\ \mathbf{d}_2^T \\ \vdots \\ \mathbf{d}_{N_t}^T \end{bmatrix} = \mathbf{XP} = \begin{bmatrix} \mathbf{x}_1^T\mathbf{P} \\ \mathbf{x}_2^T\mathbf{P} \\ \vdots \\ \mathbf{x}_{N_t}^T\mathbf{P} \end{bmatrix} \in \mathbb{C}^{N_t \times (K+L)}. \tag{3.1}$$

As seen in the sequel, the precoder is used to assist in the channel estimation [22,54,55] by eliminating the information-bearing signal at the receiver, thus leaving the superimposed training sequence used for channel estimation. It was shown in [26] that the precoder can also be designed to improve symbol detection rate or to minimize mean-squared error between the transmitted and recovered signal [54]. After precoding, the superimposed training sequence vector $\mathbf{c}_i$, for $i = 1, 2, \ldots, N_t$ is added to $\mathbf{d}_i$. Each vector is then serialized before it is transmitted across the flat-fading MIMO channel, represented in matrix form as $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$. Thus, the received signal can be written as

$$\mathbf{Y} = \mathbf{H}(\mathbf{C} + \mathbf{D}) + \boldsymbol{\eta} = \mathbf{HC} + \mathbf{HXP} + \boldsymbol{\eta}, \tag{3.2}$$

where

$$\mathbf{C} \triangleq \begin{bmatrix} \mathbf{c}_1^T \\ \mathbf{c}_2^T \\ \vdots \\ \mathbf{c}_{N_t}^T \end{bmatrix} \in \mathbb{C}^{N_t \times (K+L)}, \text{and } \boldsymbol{\eta} \in \mathbb{C}^{N_r \times (K+L)}$$

are the superimposed training sequence matrix and the additive channel noise matrix, respectively. Notice in (3.2) that the received signal in space lies in the rows of $\mathbf{Y}$. Thus, the rows of the information-bearing portion of the signal, i.e. $\mathbf{x}_i^T\mathbf{P}$, for $i = 1, 2, \ldots, N_t$, belong to the row space of $\mathbf{P}$. Hence, the rows of $\mathbf{HXP}$ also belong to the same subspace. This is different than the conventional model used in [22, 27, 28] where the information-bearing portion of the received signal is embedded inside the range space of the unknown

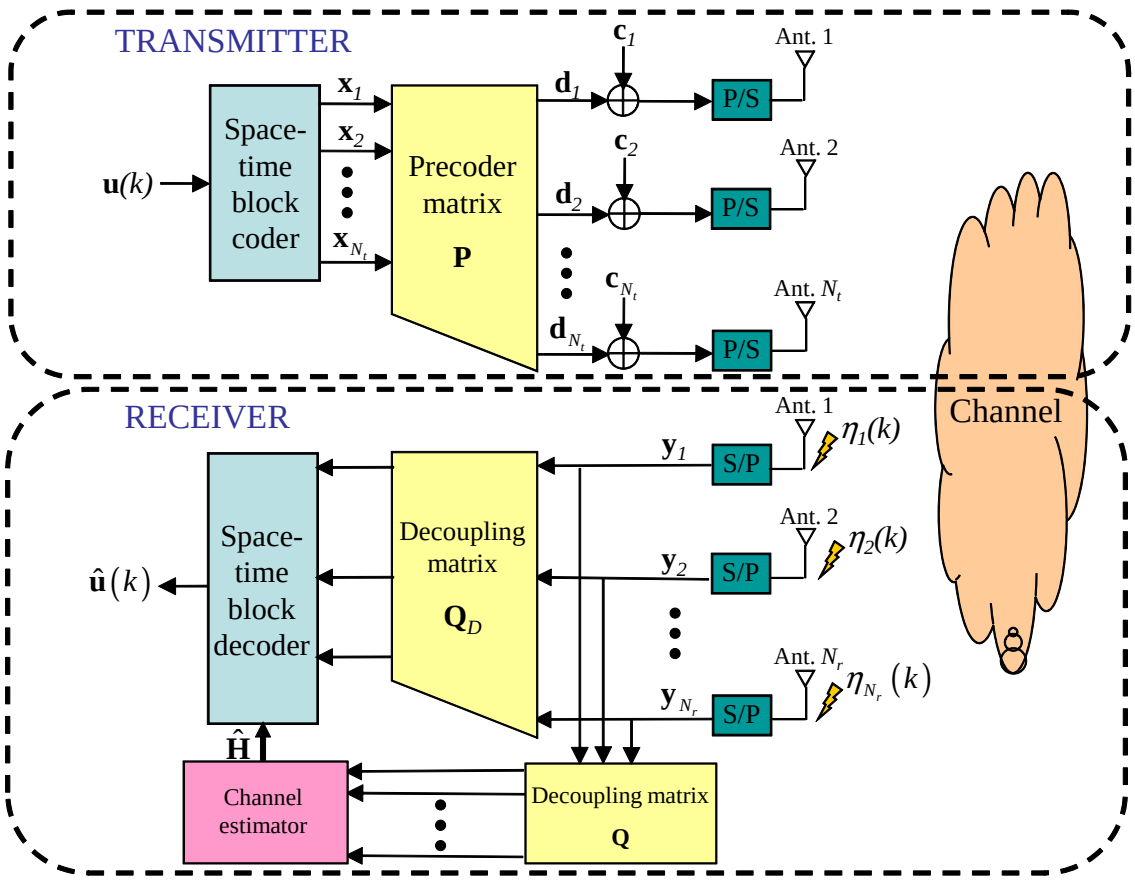



Figure 3.1: Block diagram of MIMO transceiver.

channel matrix, thus making it difficult for channel estimation using SIT sequence. The affine precoding approach adopted herein eases the decoupling of the information-bearing signal and the training sequence because decoupling can now be done by postmultiplying $\mathbf{Y}$ by a decoupling matrix, $\mathbf{Q} = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \dots & \mathbf{q}_{K+L} \end{bmatrix}^T \in \mathbb{C}^{(K+L)\times N_t}$, resulting in

$$\mathbf{YQ} = \mathbf{HCQ} + \mathbf{HXPQ} + \boldsymbol{\eta}\mathbf{Q}. \tag{3.3}$$

Thus, by requiring the columns of $\mathbf{Q}$ to lie in $\mathcal{N}(\mathbf{P})$, i.e. $\mathbf{PQ} = \mathbf{0}_{K\times N_t}$, then (3.3) becomes

$$\mathbf{YQ} = \mathbf{HCQ} + \boldsymbol{\eta}\mathbf{Q}. \tag{3.4}$$

In other words, the training sequence vector $\mathbf{c}_i$, for $i = 1, 2, \dots, N_t$ should lie in the column space of $\mathbf{Q}$. Therefore, the condition that $\mathbf{CP}^H = \mathbf{0}_{N_t\times K}$ guarantees the subspaces spanned by the vectors in $\mathbf{P}$ and $\mathbf{C}$ are complementary [55]. This suggests a simple way

to design $\mathbf{P}$ and $\mathbf{Q}$ is by extracting components off of an orthogonal matrix, i.e.

$$\begin{aligned}\mathbf{P} &= \sqrt{\frac{K+L}{K}}\mathbf{O}\left(1:K,:\right) \in \mathbb{C}^{K\times(K+L)}, \text{ and} \\ \mathbf{Q} &= \left(\mathbf{O}\left((K+1):(K+N_t),:\right)\right)^H \in \mathbb{C}^{(K+L)\times N_t}.\end{aligned}$$

Note that $\mathbf{O}(1:K,:)$ and $\mathbf{O}\left((K+1):(K+N_t),:\right)$ keep only rows 1 to $K$ and rows $(K+1)$ to $(K+N_t)$ of an orthogonal matrix $\mathbf{O} \in \mathbb{C}^{(K+N_t)\times(K+L)}$ [26]. Hence, $\mathbf{Q}^H\mathbf{Q} = \mathbf{I}_{N_t}$ so that noise amplification will not occur in the channel estimation process.

In addition to channel estimation, another decoupling matrix, $\mathbf{Q}_D$, can be designed to maximize symbol detection performance. Such decoupling matrix can be chosen to satisfy the condition $\mathbf{Q}_D = \mathbf{P}^H\left(\mathbf{P}\mathbf{P}^H\right)^{-1}$, where $\mathbf{P}$ is designed such that $\mathbf{C}\mathbf{P}^H = \mathbf{0}_{N_t\times K}$. This ensures the detection process is free of interference from the SIT sequence when $\mathbf{Q}_D$ is postmultiplied to $\mathbf{Y}$. Therefore,

$$\mathbf{P}\mathbf{P}^H = \left(\mathbf{Q}_D^H\mathbf{Q}_D\right)^{-1} = \frac{K+L}{K}\mathbf{I}_K,$$

such that $tr\left(\mathbf{P}\mathbf{P}^H\right) = K+L$. This is to ensure that the average transmitted power of the information-bearing signal is unchanged after precoding.

According to the Kronecker model [25], the channel matrix can be decomposed as

$$\mathbf{H} = \mathbf{\Sigma}_r^{\frac{1}{2}}\mathbf{H}_w\mathbf{\Sigma}_t^{\frac{1}{2}}, \tag{3.5}$$

where $\mathbf{\Sigma}_r^{\frac{1}{2}} \in \mathbb{C}^{N_r\times N_r}$ and $\mathbf{\Sigma}_t^{\frac{1}{2}} \in \mathbb{C}^{N_t\times N_t}$ are the Cholesky factors of the spatial correlation matrix of the receiver and transmitter, respectively. Hence, the overall spatial correlation is $\mathbf{R} = \mathbf{\Sigma}_t \otimes \mathbf{\Sigma}_r$. The entries of $\mathbf{H}_w \in \mathbb{C}^{N_r\times N_t}$ are independent and identically distributed zero-mean complex Gaussian random variables with unit variance. Thus, $E\left[vec(\mathbf{H}_w)vec^H(\mathbf{H}_w)\right] = \mathbf{I}_{N_rN_t}$.

### 3.2.2 MMSE estimator and training sequence design

To derive the proposed RoMMSE estimator, (3.4) is first vectorized to obtain the received signal vector

$$\mathbf{y} = \widetilde{\mathbf{C}}\mathbf{h} + \mathbf{n}, \tag{3.6}$$

where $\mathbf{y} = vec(\mathbf{YQ}) \in \mathbb{C}^{N_r N_t}$ , $\widetilde{\mathbf{C}} = (\mathbf{CQ})^T \otimes \mathbf{I}_{N_r} \in \mathbb{C}^{N_r N_t \times N_r N_t}$, $\mathbf{h} = vec(\mathbf{H}) \in \mathbb{C}^{N_r N_t}$ and $\mathbf{n} = vec(\boldsymbol{\eta}\mathbf{Q}) \in \mathbb{C}^{N_r N_t}$. $E[\mathbf{n}\mathbf{n}^H] = \sigma_n^2 \mathbf{I}_{N_r N_t}$. From the vectorized received signal $\mathbf{y}$, the linear minimum mean-squared error estimator of $\mathbf{h}$ is [29, p.387]

$$\widehat{\mathbf{h}} = \mathbf{R}_{\mathbf{yh}}^H \mathbf{R}_{\mathbf{yy}}^{-1} \mathbf{y} = \mathbf{R}\widetilde{\mathbf{C}}^H (\widetilde{\mathbf{C}}\mathbf{R}\widetilde{\mathbf{C}}^H + \sigma_{\mathbf{nn}}^2 \mathbf{I}_{N_r N_t})^{-1} \mathbf{y}, \tag{3.7}$$

where $\mathbf{R}_{\mathbf{yy}} \triangleq E\left[\mathbf{y}\mathbf{y}^H\right]$, $\mathbf{R}_{\mathbf{yh}} \triangleq E\left[\mathbf{y}\mathbf{h}^H\right]$ and $\mathbf{R} \triangleq E\left[\mathbf{h}\mathbf{h}^H\right]$ are the autocorrelation matrix of the received signal $\mathbf{y}$, the cross-correlation matrix of $\mathbf{y}$ and $\mathbf{h}$, and the spatial correlation matrix of the channel, respectively. All matrices are of size $N_r N_t \times N_r N_t$. Therefore, the optimal MMSE estimate of $\mathbf{h}$ can be obtained by finding the optimal training sequence matrix $\widetilde{\mathbf{C}}$. Note that the mean-squared error matrix between $\mathbf{h}$ and $\widehat{\mathbf{h}}$ is written as [29, p.387]

$$\begin{aligned} \boldsymbol{\xi} &= E[(\mathbf{h} - \widehat{\mathbf{h}})(\mathbf{h} - \widehat{\mathbf{h}})^H] \\ &= \left(\mathbf{R}^{-1} + \widetilde{\mathbf{C}}^H (\sigma_{\mathbf{nn}}^2 \mathbf{I}_{N_r N_t})^{-1} \widetilde{\mathbf{C}}\right)^{-1}. \end{aligned} \tag{3.8}$$

From (3.8), [55] proposed to design the optimal training sequence matrix $\widetilde{\mathbf{C}}$ by minimizing the trace of $\boldsymbol{\xi}$, subject to the power constraint $tr\left(\mathbf{C}\mathbf{C}^H\right) \leq N_t(K+L)\sigma_{\mathbf{cc}}^2 \triangleq P_T$, where $\sigma_{\mathbf{cc}}^2$ is the average power of the training sequence. It was assumed in [55] that the average transmitted power, which includes the power of the information-bearing and the training signals, is normalized as $\sigma_{\mathbf{xx}}^2 + \sigma_{\mathbf{cc}}^2 = 1$, where $\sigma_{\mathbf{xx}}^2$ is the variance of the information-bearing signal. This assumption will also be applied to the proposed RoMMSE estimator. Since $\widetilde{\mathbf{C}} = (\mathbf{CQ})^T \otimes \mathbf{I}_{N_r}$, using the properties that $tr(\mathbf{AB}) = tr(\mathbf{BA})$, $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$, $(\mathbf{A} \otimes \mathbf{B})^H = \mathbf{A}^H \otimes \mathbf{B}^H$, and $tr(\mathbf{A} \otimes \mathbf{B}) = tr(\mathbf{A})tr(\mathbf{B})$, the power constraint

on $\widetilde{\mathbf{C}}$ can be written as

$$\begin{aligned} rcltr\left(\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}^H\right) &= tr\left(\left[(\mathbf{CQ})^T \otimes \mathbf{I}_{N_r}\right]\left[(\mathbf{CQ})^T \otimes \mathbf{I}_{N_r}\right]^H\right) \\ &= tr\left(\left[(\mathbf{CQ})^T \otimes \mathbf{I}_{N_r}\right]\left[(\mathbf{CQ})^* \otimes \mathbf{I}_{N_r}\right]^H\right) \\ &= tr\left(\left[(\mathbf{CQ})^T(\mathbf{CQ})^*\right] \otimes \mathbf{I}_{N_r}\right) \\ &= tr\left((\mathbf{CQ})^T(\mathbf{CQ})^*\right) tr\left(\mathbf{I}_{N_r}\right) \\ &= N_r tr\left(\mathbf{Q}^T\mathbf{C}^T\mathbf{C}^*\mathbf{Q}^*\right) \\ &= N_r tr\left(\mathbf{C}^T\mathbf{C}^*\right) \\ &\leq N_r P_T \triangleq \widetilde{P}_T. \end{aligned} \tag{3.9}$$

The inequality is obtained because $tr\left(\mathbf{C}^T\mathbf{C}^*\right) = tr(\mathbf{C}^H\mathbf{C}) = tr(\mathbf{C}\mathbf{C}^H) = \|\mathbf{C}\|_F^2 \leq P_T$. It is important to note that the performance of the RoMMSE estimator is dependent on the total transmission power, $P_T$, and not the number of redundant vectors, $L$. The latter is however necessary to allow for decoupling of the SIT sequence from the information-bearing signal at the receiver. Figures 3.2 and 3.3 show the MSE performance of the proposed RoMMSE estimator (to be described in the next section) for $2 \times 2$ correlated MIMO channels when $L$ is increased with different and fixed $P_T$, respectively. In the latter case, $\sigma^2_{\mathbf{cc}}$ is decreased while $L$ is increased in order to keep $P_T$ constant. From Figure 3.2, it is clear that MSE performance improves as $L$ increases, while Figure 3.3 shows that such performance improvement is due to the increase in $P_T$, not just $L$, because as $P_T$ is kept constant even while $L$ is increased, there is no change in MSE performance.

### 3.2.3 Proposed training sequence design

It is clear from (3.8) that exact knowledge of $\mathbf{R}$ is required at the receiver in order to obtain an accurate estimate of $\mathbf{h}$ using (3.7). However, in all likelihood, only an estimate of $\mathbf{R}$ can be obtained, for example, using the method proposed in [30]. In order to desensitize the MSE from the estimation error of $\mathbf{R}$, a novel SIT sequence design

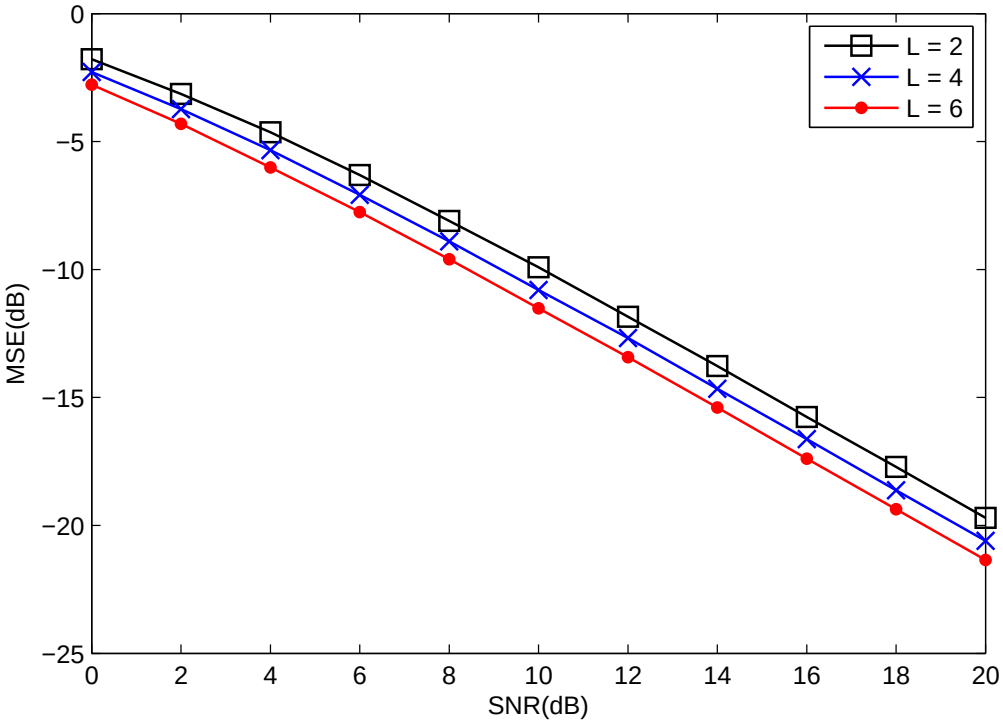



Figure 3.2: MSE vs. SNR performance comparison between different numbers of redundant vectors with different $P_T$ for spatially correlated $2 \times 2$ MIMO system, $\Delta = 5°$, $d_t = 0.5\lambda$ and $d_r = 0.2\lambda$. $\varepsilon = 0.3$.

is proposed herein which incorporates such estimation error. As the spatial correlation matrix is estimated at the receiver before it is fed back to the transmitter using a low-rate control channel, quantization error will also contribute to the error in the spatial correlation estimate, causing the mismatch between the estimated and the actual spatial correlation to be uniformly distributed. Hence, a deterministic approach is proposed herein to bound the error in a norm ball. Applying such a SIT sequence into the MMSE estimator in (3.7) allows the estimator to be more robust against estimation error in the spatial correlation than other MMSE based estimators which do not take such error into account. Even though the rate of change of the channel statistics is slower than that of the channel coefficients, imperfect channel statistics will still adversely affect the channel estimation performance and thus the BER, if not properly accounted for in the

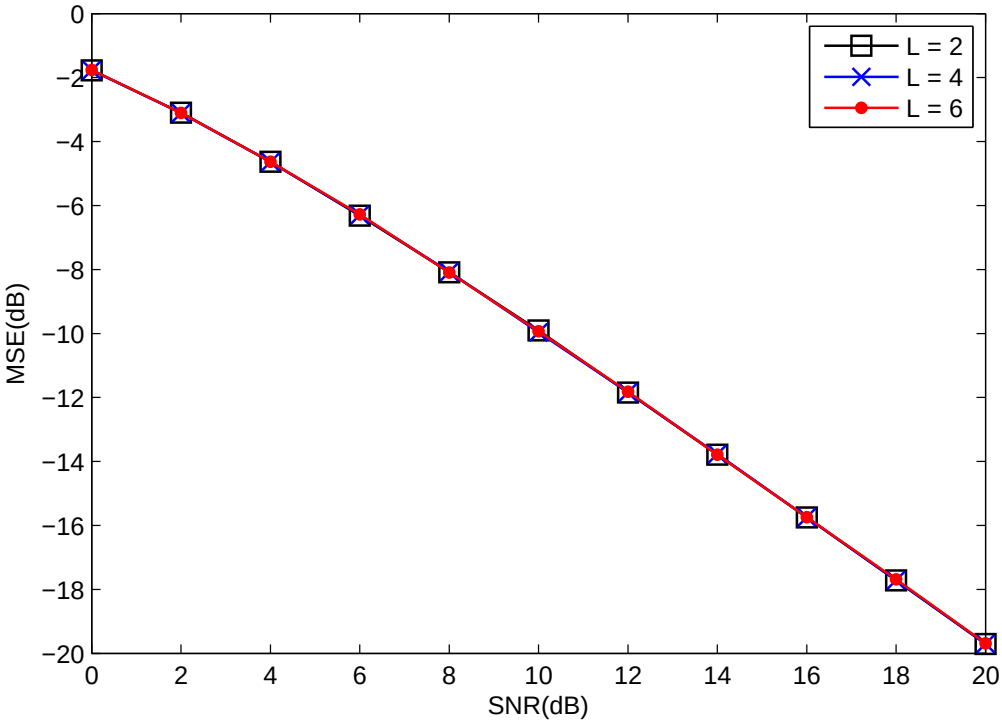



Figure 3.3: MSE vs. SNR performance comparison between different numbers of redundant vectors with fixed $P_T$ is fixed for spatially correlated $2 \times 2$ MIMO system, $\Delta = 5°$, $d_t = 0.5\lambda$ and $d_r = 0.2\lambda$. $\varepsilon = 0.3$.

system design. Moreover, better robust channel estimation can be obtained if the spatial correlation mismatch and channel coefficient mismatch can be separately accounted for as the structure of the spatial correlation mismatch matrix will be different from that of the channel coefficient matrix.

Let

$$\mathbf{R} = \widehat{\mathbf{R}} + \mathbf{E}, \tag{3.10}$$

where $\widehat{\mathbf{R}}$ denotes the estimate of $\mathbf{R}$ and $\mathbf{E}$ is its corresponding spatial correlation mismatch matrix, respectively. In the present scheme, the error power is upper bounded such that $\|\mathbf{E}\|_F \leq \varepsilon$, where $\varepsilon$ is a predefined error power bound. Using this bound with (3.9) and (3.10), the training sequence matrix $\mathbf{C}$ (or its equivalent $\widetilde{\mathbf{C}}$) can be designed by minimizing

the maximum mean-squared error $\boldsymbol{\xi}$, i.e.

$$\min_{\left\|\widetilde{\mathbf{C}}\right\|_F^2 \leq \widetilde{P}_T} \max_{\|\mathbf{E}\|_F \leq \varepsilon} \quad tr\left(\left[\left(\widehat{\mathbf{R}}+\mathbf{E}\right)^{-1}+\widetilde{\mathbf{C}}^H(\sigma_{\mathbf{nn}}^2\mathbf{I}_{N_rN_t})^{-1}\widetilde{\mathbf{C}}\right]^{-1}\right). \tag{3.11}$$

Note that (3.11) is not a convex problem with respect to $\mathbf{E}$ and $\mathbf{C}$. However, the problem can be decomposed into two separate convex optimization problems, one with respect to $\mathbf{E}$ and the other to $\mathbf{C}$. Furthermore, performing SVD on $\widetilde{\mathbf{C}}$, i.e. $\widetilde{\mathbf{C}}=\mathbf{U}_{\widetilde{\mathbf{C}}}\boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}}\mathbf{V}_{\widetilde{\mathbf{C}}}^H$ and using the property $tr\left(\mathbf{AB}\right)=tr(\mathbf{BA})$, the objective function of the maximization problem in (3.11) can be rewritten as

$$\begin{aligned} & tr\left(\left[\left(\widehat{\mathbf{R}}+\mathbf{E}\right)^{-1}+\widetilde{\mathbf{C}}^H\left(\sigma_{\mathbf{nn}}^2\mathbf{I}_{N_rN_t}\right)^{-1}\widetilde{\mathbf{C}}\right]^{-1}\right) \\ = \; & tr\left(\left[\left(\widehat{\mathbf{R}}+\mathbf{E}\right)^{-1}+\sigma_{\mathbf{nn}}^{-2}\mathbf{V}_{\widetilde{\mathbf{C}}}\boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}}^H\boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}}\mathbf{V}_{\widetilde{\mathbf{C}}}^H\right]^{-1}\right) \\ = \; & tr\left(\left[\mathbf{V}_{\widetilde{\mathbf{C}}}^H\left(\widehat{\mathbf{R}}+\mathbf{E}\right)^{-1}\mathbf{V}_{\widetilde{\mathbf{C}}}+\sigma_{\mathbf{nn}}^{-2}\boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}}^H\boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}}\right]^{-1}\right). \end{aligned} \tag{3.12}$$

Next, using the property $tr(\mathbf{A}+\mathbf{B})=tr(\mathbf{A})+tr(\mathbf{B})$ and the matrix inversion lemma $(\mathbf{A}+\mathbf{BCD})^{-1}=\mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1}+\mathbf{DA}^{-1}\mathbf{B})^{-1}\mathbf{DA}^{-1}$, and letting $\mathbf{A}=\sigma_{\mathbf{nn}}^{-2}\boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}}^H\boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}}$, $\mathbf{C}=\mathbf{V}_{\widetilde{\mathbf{C}}}^H\left(\widehat{\mathbf{R}}+\mathbf{E}\right)^{-1}\mathbf{V}_{\widetilde{\mathbf{C}}}$ and $\mathbf{B}=\mathbf{D}=\mathbf{I}_{N_rN_t}$, then (3.12) becomes

$$\begin{aligned} & tr\left(\sigma_{\mathbf{nn}}^2\boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}}^{-1}\boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}}^{-H}-\sigma_{\mathbf{nn}}^2\boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}}^{-1}\boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}}^{-H}\left[\sigma_{\mathbf{nn}}^2\boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}}^{-1}\boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}}^{-H}+\mathbf{V}_{\widetilde{\mathbf{C}}}^H\left(\widehat{\mathbf{R}}+\mathbf{E}\right)\mathbf{V}_{\widetilde{\mathbf{C}}}\right]^{-1}\sigma_{\mathbf{nn}}^2\boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}}^{-1}\boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}}^{-H}\right) \\ = \; & tr\left(\sigma_{\mathbf{nn}}^2\boldsymbol{\Lambda}_{\widetilde{\mathbf{C}}}^{-1}\right)-\sigma_{\mathbf{nn}}^4tr\left(\boldsymbol{\Lambda}_{\widetilde{\mathbf{C}}}^{-1}\left[\sigma_{\mathbf{nn}}^2\boldsymbol{\Lambda}_{\widetilde{\mathbf{C}}}^{-1}+\mathbf{V}_{\widetilde{\mathbf{C}}}^H\left(\widehat{\mathbf{R}}+\mathbf{E}\right)\mathbf{V}_{\widetilde{\mathbf{C}}}\right]^{-1}\boldsymbol{\Lambda}_{\widetilde{\mathbf{C}}}^{-1}\right) \\ = \; & tr\left(\sigma_{\mathbf{nn}}^2\boldsymbol{\Lambda}_{\widetilde{\mathbf{C}}}^{-1}\right)-\sigma_{\mathbf{nn}}^2tr\left(\left[\boldsymbol{\Lambda}_{\widetilde{\mathbf{C}}}+\sigma_{\mathbf{nn}}^{-2}\boldsymbol{\Lambda}_{\widetilde{\mathbf{C}}}\mathbf{V}_{\widetilde{\mathbf{C}}}^H\left(\widehat{\mathbf{R}}+\mathbf{E}\right)\mathbf{V}_{\widetilde{\mathbf{C}}}\boldsymbol{\Lambda}_{\widetilde{\mathbf{C}}}\right]^{-1}\right) \\ = \; & tr\left(\sigma_{\mathbf{nn}}^2\boldsymbol{\Lambda}_{\widetilde{\mathbf{C}}}^{-1}\right)-\sigma_{\mathbf{nn}}^2tr\left(\left[\widetilde{\mathbf{C}}^H\widetilde{\mathbf{C}}+\sigma_{\mathbf{nn}}^{-2}\widetilde{\mathbf{C}}^H\widetilde{\mathbf{C}}\left(\widehat{\mathbf{R}}+\mathbf{E}\right)\widetilde{\mathbf{C}}^H\widetilde{\mathbf{C}}\right]^{-1}\right), \end{aligned} \tag{3.13}$$

where $\boldsymbol{\Lambda}_{\widetilde{\mathbf{C}}}\triangleq\boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}}^H\boldsymbol{\Sigma}_{\widetilde{\mathbf{C}}}$. Since the first term of (3.13) does not depend on $\mathbf{E}$, maximizing (3.11) is equivalent to minimizing the second term in (3.13). Therefore, the maximization problem in (3.11) becomes

$$\min_{\|\mathbf{E}\|_F\leq\varepsilon} tr\left(\left[\widetilde{\mathbf{C}}^H\widetilde{\mathbf{C}}+\sigma_{\mathbf{nn}}^{-2}\widetilde{\mathbf{C}}^H\widetilde{\mathbf{C}}\left(\widehat{\mathbf{R}}+\mathbf{E}\right)\widetilde{\mathbf{C}}^H\widetilde{\mathbf{C}}\right]^{-1}\right), \tag{3.14}$$

and it can be easily solved using convex optimization toolbox such as *cvx* [59] since the $\left(\widehat{\mathbf{R}}+\mathbf{E}\right)^{-1}$ term is eliminated.

Unfortunately, even if (3.14) is substituted into (3.11), it is difficult to find a closed-form solution for $\mathbf{C}$. Therefore, the iterative algorithm in Figure 3.4 is proposed. As seen from the figure, $\widetilde{\mathbf{C}}$ is first initialized to be a full rank matrix satisfying the condition $\mathbf{C}(0)\mathbf{P}^H = \mathbf{0}$. $\mathbf{C}(0)$ is then used in (3.17) (or equivalently (3.14)) to solve for a solution for $\mathbf{E}$. This is then used in (3.18) to solve for $\widetilde{\mathbf{C}}$. This process will be repeated until $\frac{\|\mathbf{E}(n)-\mathbf{E}(n-1)\|^2}{\varepsilon}$ is less than some preset threshold $\alpha$. Note that $\widetilde{\mathbf{C}}$ needs to be initialized to have full rank otherwise the inverse in (3.17) cannot be taken.

Assuming that $\mathbf{C}$ has full row rank. Initializing $\mathbf{C}$ in the algorithm shown in Figure 3.4 to be $\mathbf{C}(0)$, it is obvious that

$$\mathbf{C}(0) = \mathbf{U}_{\mathbf{C}(0)} \left[\mathbf{\Sigma}_{\mathbf{C}(0)} \quad \mathbf{0}_{N_t\times(K+L-N_t)}\right] \mathbf{V}^H_{\mathbf{C}(0)}, \tag{3.15}$$

where $\mathbf{U}_{\mathbf{C}(0)}$, $\mathbf{V}_{\mathbf{C}(0)}$, and $\mathbf{\Sigma}_{\mathbf{C}(0)}$ are the left and singular vector matrix of $\mathbf{C}(0)$, and the invertible portion of the singular value matrix of $\mathbf{C}(0)$, respectively. Hence, to satisfy the condition that $\mathbf{C}\mathbf{P}^H = \mathbf{0}_{N_t\times K}$, it is necessary that $\mathbf{V}_{\mathbf{C}(0)} = \mathbf{U}_{\mathbf{Q}}$, where $\mathbf{U}_{\mathbf{Q}}$ is the eigenvector matrix of $\mathbf{Q}\mathbf{Q}^H$. That is,

$$\begin{aligned}\mathbf{Q}\mathbf{Q}^H &= \mathbf{U}_{\mathbf{Q}}\mathbf{\Lambda}_{\mathbf{Q}}\mathbf{U}^H_{\mathbf{Q}} \\ &= \mathbf{U}_{\mathbf{Q}} \begin{bmatrix} \mathbf{\Lambda}'_{\mathbf{Q}} & \mathbf{0}_{N_t\times(K+L-N_t)} \\ \mathbf{0}_{(K+L-N_t)\times N_t} & \mathbf{0}_{(K+L-N_t)\times(K+L-N_t)} \end{bmatrix} \mathbf{U}^H_{\mathbf{Q}},\end{aligned}$$

where $\mathbf{\Lambda}'_{\mathbf{Q}} \in \mathbb{C}^{N_t\times N_t}$ is a diagonal matrix containing the non-zero eigenvalues of $\mathbf{Q}\mathbf{Q}^H$. Assuming that the diagonal values of $\mathbf{\Lambda}_{\mathbf{Q}}$ are arranged in descending order. Hence, $\widetilde{\mathbf{C}}^2(n) = \left(\mathbf{U}_{\mathbf{C}(n)}\mathbf{\Lambda}_{\mathbf{C}(n)}\mathbf{U}^H_{\mathbf{C}(n)}\right)^* \otimes \mathbf{I}_{N_r}$, for $n = 0, 1, ..., n_0$, where $n$ and $n_0$ denote the iteration index and the iteration time when $\|\mathbf{E}(n) - \mathbf{E}(n-1)\|^2 / \epsilon < \alpha$, respectively, and $\mathbf{\Lambda}_{\mathbf{C}(n)} = \mathbf{\Sigma}_{\mathbf{C}(n)}\mathbf{\Sigma}^*_{\mathbf{C}(n)}$. Thus, $\mathbf{V}_{\mathbf{C}(n)} = \mathbf{V}_{\mathbf{C}(0)} = \mathbf{U}_{\mathbf{Q}}$ for $n = 0, 1, ..., n_0$ and the training sequence when convergence has been reached becomes

$$\mathbf{C}(n_0) = \mathbf{U}_{\mathbf{C}(n_0)} \left[\mathbf{\Sigma}_{\mathbf{C}(n_0)} \quad \mathbf{0}_{N_t\times(K+L-N_t)}\right] \mathbf{U}^H_{\mathbf{Q}}, \tag{3.16}$$

where $\mathbf{U}_{\mathbf{C}(n_0)}$ is the singular vector matrix for $\mathbf{C}(n_0)$ and $\mathbf{\Sigma}_{\mathbf{C}(n_0)} \in \mathbb{C}^{N_t \times N_t}$ is a singular value matrix of $\mathbf{C}(n_0)$ containing all non-zero singular values. This conforms to the structure previously derived in [55].

### 3.2.4 Convergence analysis

**Theorem 1** *The iteration depicted in Figure 3.4 will always converge to the global optimal solution given that* $\mathbf{C}$ *is initialized as a matrix with full rank, where the constraint* $\mathbf{C}(0)\mathbf{P}^H = \mathbf{0}_{N_t \times K}$ *is satisfied.*

**Proof.** Define the convex sets $\mathcal{E} = \left\{ \left( \widetilde{\mathbf{C}}, \mathbf{E} \right) | \|\mathbf{E}\|_F \leq \varepsilon \right\}$ and $\mathcal{C} = \left\{ \left( \widetilde{\mathbf{C}}, \mathbf{E} \right) \left| \left\| \widetilde{\mathbf{C}} \right\|_F^2 \leq \widetilde{P}_T \right. \right\}$ containing elements that are 2-tuples.[1] Expressing the objective function in (3.11) as $f\left( \widetilde{\mathbf{C}}, \mathbf{E} \right)$. Given $\widetilde{\mathbf{C}}$, it is clear that $\max f\left( \widetilde{\mathbf{C}}, \mathbf{E} \right)$ with respect to the $\mathbf{E}$ is a non-expansive operator, i.e.

$$\begin{aligned} 0 &= \left\| \max_{\|\mathbf{E}_1\|_F \leq \varepsilon} f\left( \widetilde{\mathbf{C}}, \mathbf{E}_1 \right) - \max_{\|\mathbf{E}_2\|_F \leq \varepsilon} f\left( \widetilde{\mathbf{C}}, \mathbf{E}_2 \right) \right\| \\ &\leq \|\mathbf{E}_1 - \mathbf{E}_2\| . \end{aligned} \tag{3.19}$$

Similarly, given $\mathbf{E}$, $\min f\left( \widetilde{\mathbf{C}}, \mathbf{E} \right)$ with respect to the $\widetilde{\mathbf{C}}$ is also a non-expansive operator, i.e.

$$\begin{aligned} 0 &= \left\| \min_{\|\widetilde{\mathbf{C}}_1\|_F^2 \leq \widetilde{P}_T} f\left( \widetilde{\mathbf{C}}_1, \mathbf{E} \right) - \min_{\|\widetilde{\mathbf{C}}_2\|_F^2 \leq \widetilde{P}_T} f\left( \widetilde{\mathbf{C}}_2, \mathbf{E} \right) \right\| \\ &\leq \left\| \widetilde{\mathbf{C}}_1 - \widetilde{\mathbf{C}}_2 \right\| . \end{aligned} \tag{3.20}$$

Moreover, the solutions in (3.19) and (3.20) will always belong to either $\mathcal{E}$ or $\mathcal{C}$. Then according to the theory of alternating projections [32], the algorithm depicted in Figure 3.4 will always converge, given appropriate initial conditions. Since the two sets are convex, there is a unique point of intersection and thus the solution obtained in Figure 3.4 will always be the global optimal solution.

Note that it is possible for $\mathbf{C}(n)$ to lose rank when the SNR is low and when $\varepsilon$ is sufficiently small (e.g. SNR $= 0$ dB and $\varepsilon = 0.1$). To prevent this from occurring, $\widetilde{\mathbf{C}}^2(n)$ is diagonally loaded, i.e. $\widetilde{\mathbf{C}}^2(n) = \widetilde{\mathbf{C}}^2(n) + \rho \mathbf{I}_{N_t N_r}$, where $\rho$ is a small value compared to $\left\| \widetilde{\mathbf{C}}^2(n) \right\|_F$, e.g. $\rho = 0.01 \left\| \widetilde{\mathbf{C}}^2(n) \right\|_F$.

[1] Both sets are convex because their respective constraints form a norm ball.

Define:

$\mathbf{C}(n)$ : training sequence matrix at the $n^{th}$ iteration

$\mathbf{E}(n)$ : error mismatch matrix at the $n^{th}$ iteration

$\widetilde{\mathbf{C}}(n) \triangleq (\mathbf{C}(n)\mathbf{Q})^T \otimes \mathbf{I}_{N_r}$

$\widetilde{\mathbf{C}}^2(n) \triangleq \widetilde{\mathbf{C}}^H(n)\widetilde{\mathbf{C}}(n)$

Algorithm:

initialize $\mathbf{C}(0)$ to any full rank matrix, s.t.

$\mathbf{C}(0)\mathbf{P}^H = \mathbf{0}_{N_t \times K}$

initialize $\mathbf{E}(1) = \mathbf{0}_{N_r N_t \times N_r N_t}$

initialize $\alpha$ to the value of a given threshold

initialize $\varepsilon$ to the value of a given error bound

initialize $\rho$ to a given value for the diagonal loading

$n = 1$

DO:

if $rank\left(\widetilde{\mathbf{C}}^2(n)\right) < N_t N_r$

$\widetilde{\mathbf{C}}^2(n) = \widetilde{\mathbf{C}}^2(n) + \rho \mathbf{I}_{N_r N_t}$

endif

$$\mathbf{E}_{opt}(n) = \arg \min_{\substack{\|\mathbf{E}(n)\|_F \le \varepsilon \\ tr\left(\widehat{\mathbf{R}}+\mathbf{E}(n)\right)=N_t N_r}} tr\left(\left[\widetilde{\mathbf{C}}^2(n-1) + \sigma_{\mathbf{nn}}^{-2}\widetilde{\mathbf{C}}^2(n-1)\left(\widehat{\mathbf{R}} + \mathbf{E}(n)\right)\widetilde{\mathbf{C}}^2(n-1)\right]^{-1}\right) \quad (3.17)$$

$$\widetilde{\mathbf{C}}_{opt}(n) = \arg \min_{\|\widetilde{\mathbf{C}}(n)\|_F^2 \le \widetilde{P}_T} tr\left(\left[\left(\widehat{\mathbf{R}} + \mathbf{E}_{opt}(n)\right)^{-1} + \sigma_{\mathbf{nn}}^{-2}\widetilde{\mathbf{C}}^H(n)\widetilde{\mathbf{C}}(n)\right]^{-1}\right) \quad (3.18)$$

$\widetilde{\mathbf{C}}(n) = \widetilde{\mathbf{C}}_{opt}(n)$

$n = n + 1$

$\mathbf{E}(n) = \mathbf{E}_{opt}(n)$

Until $\frac{\|\mathbf{E}(n)-\mathbf{E}(n-1)\|^2}{\varepsilon} < \alpha$

Figure 3.4: Algorithm pseudocode for training sequence design.

### 3.2.5 Power Allocation

Since the use of the SIT sequence requires reducing the power of the information-bearing signal, this will have an adverse effect on the recovery of the information-bearing signal. A suboptimal power allocation scheme for the training sequence was derived in [55] in which the effective SNR

$$\mathrm{SNR}_{eff} = \frac{E\left[\left\|\widehat{\mathbf{H}}\mathbf{X}\right\|^2\right]}{E\left[\left\|\widetilde{\mathbf{H}}\mathbf{X} + \mathbf{N}\mathbf{Q}_D\right\|^2\right]} \tag{3.21}$$

was maximized. A similar power allocation scheme can similarly be derived with the spatial correlation mismatch matrix, $\mathbf{E}$, taken into account. The method is suboptimal due to the fact that the numerator in (3.21) can be written as

$$\begin{aligned} E\left[\left\|\widehat{\mathbf{H}}\mathbf{X}\right\|^2\right] &= tr\left(E\left[\widehat{\mathbf{H}}\mathbf{X}\mathbf{X}^H\widehat{\mathbf{H}}^H\right]\right) \\ &= K\sigma_x^2 tr\left(E\left[\mathbf{H}\mathbf{H}^H\right] + E\left[\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^H\right]\right) \\ &= K\sigma_x^2\left(tr\left(\widehat{\mathbf{R}} + \mathbf{E}\right) + \epsilon\right) \\ &= K\sigma_{\mathbf{xx}}^2\left(N_t N_r + \epsilon\right), \end{aligned} \tag{3.22}$$

where it has been assumed that $tr\left(E\left[\mathbf{H}\mathbf{H}^H\right]\right) = tr\left(\widehat{\mathbf{R}} + \mathbf{E}\right) = N_t N_r$, with $\epsilon = tr\left(E\left[\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^H\right]\right)$ denoting the mean-squared error of the channel. Note that the received SNR is defined as $\mathrm{SNR} = -10\log_{10}\sigma_{\mathbf{nn}}^2$ under the assumption that the power of the received signal is normalized to 1. Thus, $\mathrm{trace}(\widehat{\mathbf{R}} + \mathbf{E}) = N_t N_r$. In addition, it should be noted that there is an error in the expression for $E\left[\left\|\widehat{\mathbf{H}}\mathbf{X}\right\|^2\right]$ in [55] in which $\epsilon$ was preceded by a minus sign, even if it should be preceded by a plus sign instead, as indicated in (3.22). Hence, $\mathrm{SNR}_{eff}$ becomes

$$\mathrm{SNR}_{eff} = \frac{\sigma_{\mathbf{xx}}^2\left(N_t N_r + \epsilon\right)}{\sigma_{\mathbf{xx}}^2\epsilon + \gamma}, \tag{3.23}$$

where $\gamma = N_r\sigma_{\mathbf{nn}}^2\frac{K}{K+L}$. Using the property that if $tr\left(\mathbf{A}\right) > tr\left(\mathbf{B}\right)$, then $tr\left(\mathbf{B}^{-1}\right) > tr\left(\mathbf{A}^{-1}\right)$, given that $\mathbf{A}$ and $\mathbf{B}$ are positive definite matrices. It then follows that $\epsilon$ is

upper bounded by $\frac{N_r N_t}{N_r+N_t}\frac{\sigma^2_{\mathbf{nn}}}{\sigma^2_{\mathbf{cc}}} = \beta\frac{\sigma^2_{\mathbf{nn}}}{\sigma^2_{\mathbf{cc}}}$. Substituting this upper bound into (3.22), the effective SNR is then lower bounded as

$$\mathrm{SNR}_{eff}\left(\sigma^2_{\mathbf{cc}}\right) \geq \frac{\left(1-\sigma^2_{\mathbf{cc}}\right)\left[\left(N_t N_r+\right)\sigma^2_{\mathbf{cc}}+\beta\sigma^2_{\mathbf{nn}}\right]}{\beta\sigma^2_{\mathbf{nn}}-\beta\sigma^2_{\mathbf{nn}}\sigma^2_{\mathbf{cc}}+\gamma\sigma^2_{\mathbf{cc}}}. \tag{3.24}$$

The maximum of the effective SNR can then be achieved by maximizing the lower bound in (3.24), which can be accomplished by differentiating the bound with respect to $\sigma^2_{\mathbf{cc}}$, setting the result to zero, and solving for $\sigma^2_{\mathbf{cc}}$. This results in the suboptimal power allocation for the SIT sequence

$$\sigma^2_{\mathbf{cc},subopt} = \frac{\delta\beta\sigma^2_{\mathbf{nn}}-\sqrt{\delta\gamma\beta\sigma^2_{\mathbf{nn}}\left(\delta-\gamma+\beta\sigma^2_{\mathbf{nn}}\right)}}{\delta(\beta\sigma^2_{\mathbf{nn}})}, \tag{3.25}$$

where $\delta = N_t N_r$. $\sigma^2_{\mathbf{cc},subopt}$ in (3.25) is similar to the expression derived in [55] except for the sign error as previously indicated. The difference is due to the sign error in (3.22). However, the power allocation expression above is derived directly with inclusion of the spatial correlation mismatch, thus generalizing the result previously reported in [55].

## 3.3 Simulation Results

Monte-Carlo simulations were used to demonstrate the robustness of the proposed scheme. Channels used in all simulations are assumed to be quasi-static block Rayleigh fading and spatially correlated, unless otherwise specified. The one-ring model [31] is used to generate entries of the Cholesky factors of the spatial transmit and receive correlation matrices

$$\Sigma_t(m,n) \approx J_0\big(\Delta\frac{2\pi}{\lambda}d_t|m-n|\big), \tag{3.26}$$

and

$$\Sigma_r(i,j) \approx J_0\big(\frac{2\pi}{\lambda}d_r|i-j|\big), \tag{3.27}$$

where $d_t$ and $d_r$ are the spacing between transmit and receive antennas, respectively. $\Delta$ denotes the angular spread, $\lambda$ denotes the carrier wavelength, and $J_0$ is the $0^{th}$-order Bessel

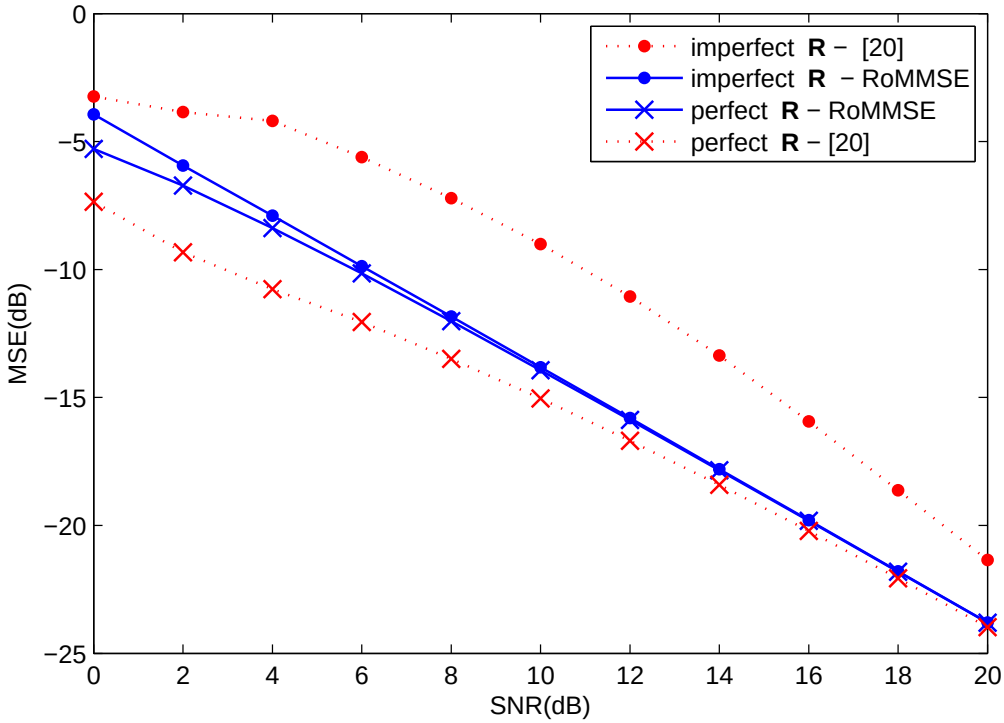



Figure 3.5: MSE vs. SNR performance comparison between RoMMSE and [55] for spatially correlated $2 \times 2$ MIMO channel. $\Delta = 5°$, $d_t = 0.5\lambda$ and $d_r = 0.2\lambda$. $\varepsilon = 0.3$.

function of the first kind. The power allocation scheme in (3.25) for the training sequence is adopted. QPSK and Alamouti STBC are used for modulation of the information-bearing signals. In all simulations, the threshold for the iteration algorithm is $\alpha = 10^{-6}$. With this value for $\alpha$, the proposed algorithm required only at most 6 iterations before convergence in all the simulations.

A $2 \times 2$ spatially correlated MIMO system with $\Delta = 5°$, $d_t = 0.5\lambda$ and $d_r = 0.2\lambda$ is considered in Figure 3.5. The data block size $K$ is 60, and $L = N_t = 2$. When the correlation matrix $\mathbf{R}$ is estimated perfectly, i.e. $\mathbf{R} = \widehat{\mathbf{R}}$, the sequence design in [55] outperforms the proposed RoMMSE algorithm with $\varepsilon = 0.3$. This is the case since the sequence design in [55] is MMSE optimal when perfect knowledge of $\mathbf{R}$ is available. However, when $\mathbf{R}$ is not estimated accurately, i.e. $\mathbf{R} = \hat{\mathbf{R}} + \mathbf{E}$, the proposed RoMMSE estimator outperforms the estimator in [55] by as much as 8 dB.

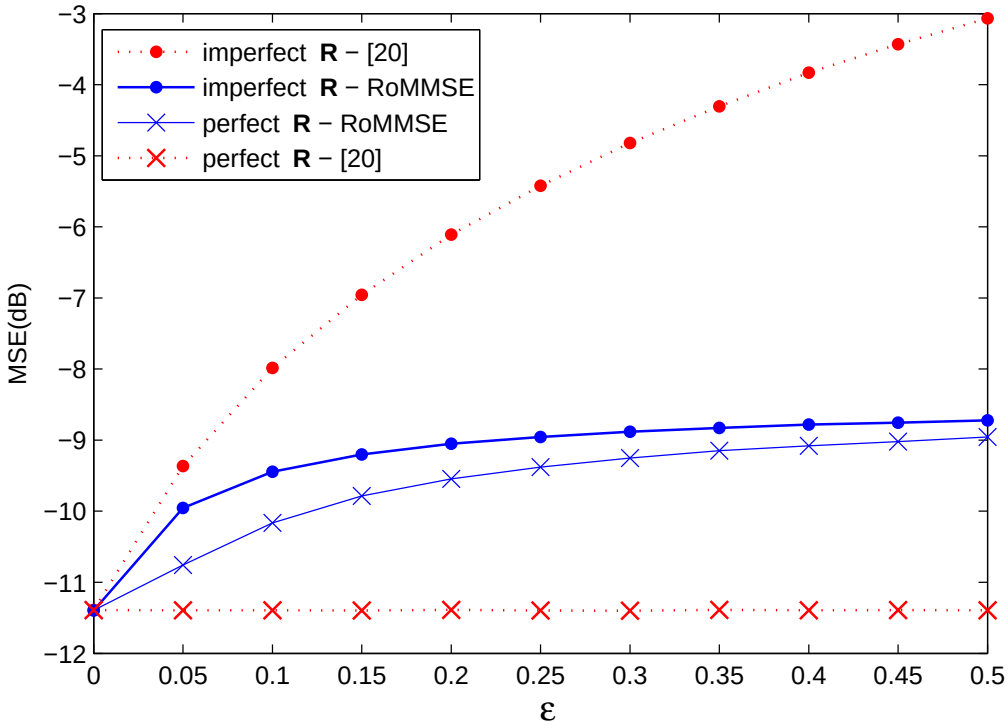



Figure 3.6: MSE vs. $\varepsilon$ performance comparison between RoMMSE and [55] for spatially correlated $2 \times 2$ MIMO channel. $\Delta = 5°$, $d_t = 0.5\lambda$ and $d_r = 0.2\lambda$. SNR = 5 dB.

Figure 3.6 compares the MSE performance of the proposed scheme with that of [55] when the spatial correlation matrix error power is varied. The channel parameters in this figure is identical to those in Figure 3.5 with SNR = 5 dB. For the case of "imperfect $\mathbf{R}$", i.e. $\mathbf{E} \neq \mathbf{0}_{N_r N_t \times N_r N_t}$, the results for [55] is obtained by solving (3.8) with $\mathbf{R} = \widehat{\mathbf{R}}$. In the case of "perfect $\mathbf{R}$", i.e. $\mathbf{E} = \mathbf{0}_{N_r N_t \times N_r N_t}$, the exact matrix channel correlation matrix is used to design the training sequence for both algorithms. Observed from the figure that the algorithm in [55] outperforms the proposed scheme when an accurate spatial correlation matrix is available for estimation. However, when $\widehat{\mathbf{R}} \neq \mathbf{R}$, then the proposed scheme outperforms [55]. Moreover, as the estimation error $\varepsilon$ increases, the MSE of the RoMMSE estimator rises only gradually while the MSE increases unbounded for [55].

Figures 3.7 and 3.8 illustrate the same performance comparison as Figures 3.5 and 3.6, but for $4 \times 4$ MIMO systems. The angular spread $\Delta$ is set to be 15° and antenna spacing $d_t$

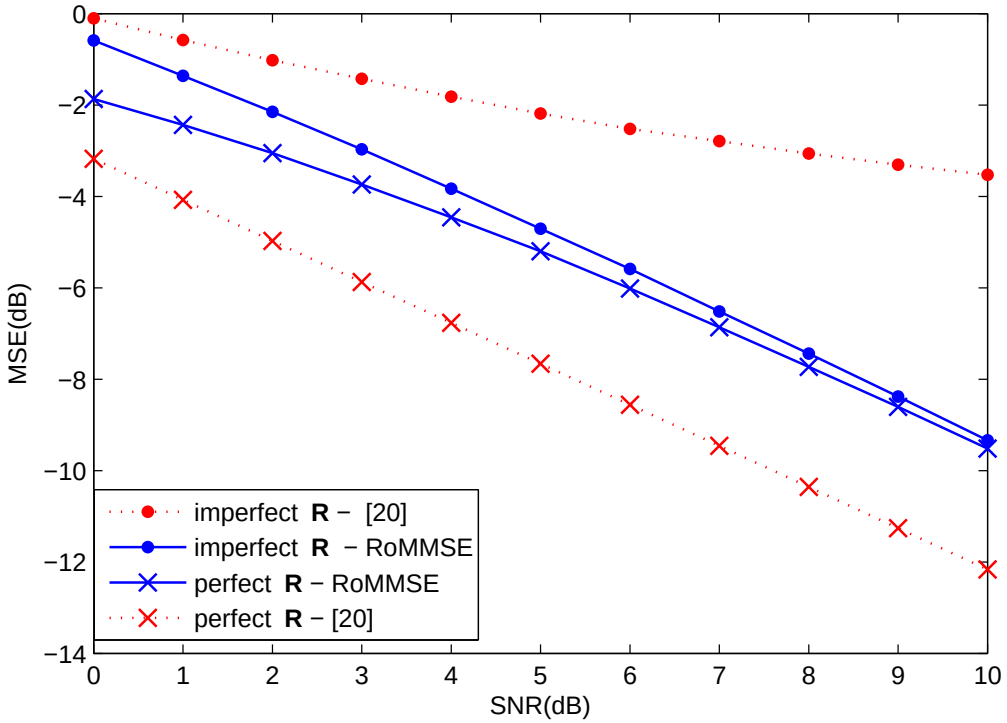



Figure 3.7: MSE vs. SNR performance comparison between RoMMSE and [55] for spatially correlated $4 \times 4$ MIMO channel. $\Delta = 15°$, $d_t = 0.5\lambda$ and $d_r = 0.2\lambda$. $\varepsilon = 0.3$.

and $d_r$ are $0.5\lambda$ and $0.2\lambda$, respectively. $K = 60$ and $L = N_t = 4$. $\varepsilon = 0.3$ is used in Figure 3.7 while SNR = 5 dB is used for Figure 3.8. From Figure 3.7, the performance of both algorithms for the $4 \times 4$ system follows the same pattern as that of the $2 \times 2$. Specifically, the proposed RoMMSE estimator outperforms the estimator in [55] by as much as 9 dB when the MSE $= -1$ dB. Also, unlike the algorithm in [55], the RoMMSE estimator performance does not flatten out as the SNR increases. This is because the inaccuracy in $\mathbf{R}$ has been taken into account during the channel estimation process. However, since there are more parameters to be estimated in the $4 \times 4$ system compared to the $2 \times 2$, there is a performance degradation not only in terms of the absolute MSE, but also the rate of decrease of the MSE has also diminished.

Besides MSE performance, Figures 3.9 and 3.10 compare the BER performance of the RoMMSE algorithm and that of [55] when the estimate of spatial correlation is imperfect

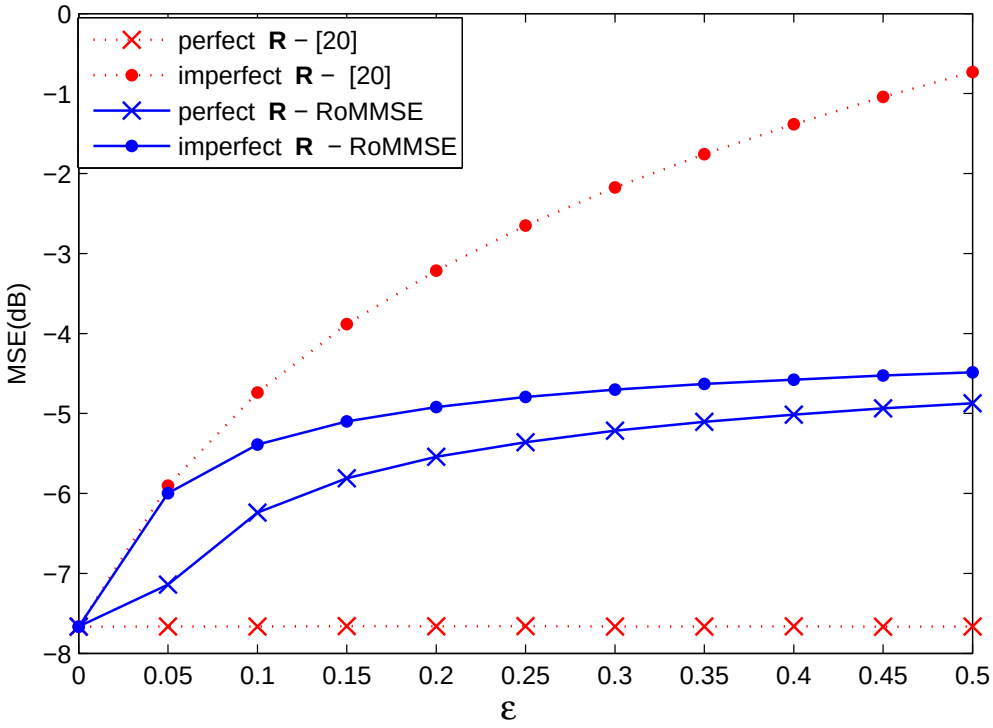



Figure 3.8: MSE vs. $\varepsilon$ performance comparison between RoMMSE and [55] for spatially correlated $4 \times 4$ MIMO channel. $\Delta = 15°$, $d_t = 0.5\lambda$ and $d_r = 0.2\lambda$. SNR = 5 dB.

and when it is perfect, respectively. $2 \times 2$ MIMO systems are used. From Figure 3.9, it can be seen that the RoMMSE algorithm outperforms the algorithm in [55] by 2 dB when the SNR is low. However, when the spatial correlation has been perfectly estimated, the RoMMSE algorithm and the algorithm in [55] render identical performance.

Figures 3.11 and 3.12 compare the MSE performance for spatially correlated and uncorrelated MIMO channels of the RoMMSE estimator to the RMMSE and LS-RMMSE estimators in [23]. The spatial correlation in Figure 3.11 is created by letting $\Delta = 5°$, $d_t = 0.5\lambda$, and $d_r = 0.2\lambda$. The RMMSE uses diagonal loading to derive an MMSE estimator that requires only knowledge of $tr(\mathbf{R})$ instead of $\mathbf{R}$ to estimate the MIMO channel. LS-RMMSE further relaxes the requirement in RMMSE by using LS method to derive an MMSE estimator that no longer requires knowledge of $tr(\mathbf{R})$. Instead, only knowledge about the Frobenius norm of the received signal matrix is required. As seen in Figure

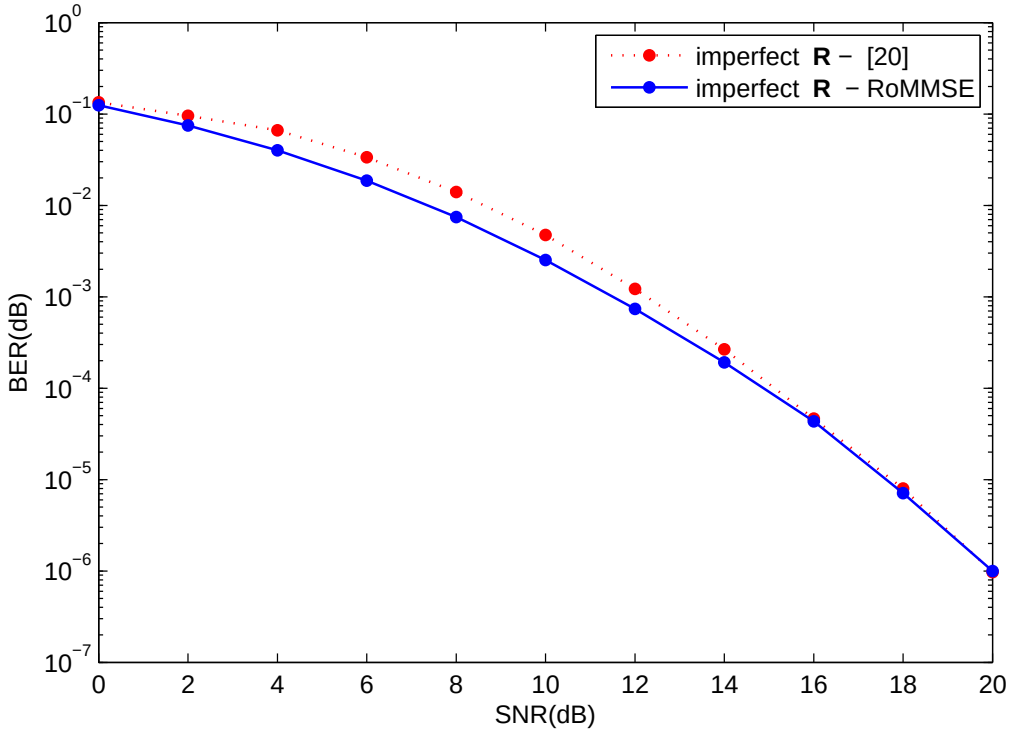



Figure 3.9: BER vs. SNR performance comparison between RoMMSE and [55] for spatially correlated $2 \times 2$ MIMO channel. $\Delta = 5°$, $d_t = 0.5\lambda$ and $d_r = 0.2\lambda$. $\varepsilon = 0.3$.

3.11, when spatial correlation exists, the proposed RoMMSE algorithm outperforms the RMMSE and LS-RMMSE algorithms in low SNR by 4 dB when $\varepsilon = 0.05$, but only about 2 dB when $\epsilon = 0.2$. This shows that the upper error power bound cannot be too high, otherwise performance of the proposed scheme will degrade. The reason for this behavior is because as $\varepsilon$ increases, $\mathbf{E}$ obtained from the iterative algorithm will decorrelate the spatial correlation more, thus adversely affecting the performance of the proposed scheme. This can be explained as follows. The RoMMSE estimator strives to minimize the worst case MSE as seen in (3.11). The worst case MSE can be attained by increasing the number of parameters that needs to be estimated, the maximum being $N_r N_t$. In other words, the present method attempts to increase the degrees of freedom in the correlated MIMO channel by reducing the spatial correlation. As $\varepsilon$ increases, $\|\mathbf{E}\|_F$ also increases, which allows $\mathbf{E}$ more freedom to zero out the off-diagonal elements of $\mathbf{R}$, therefore lessening the

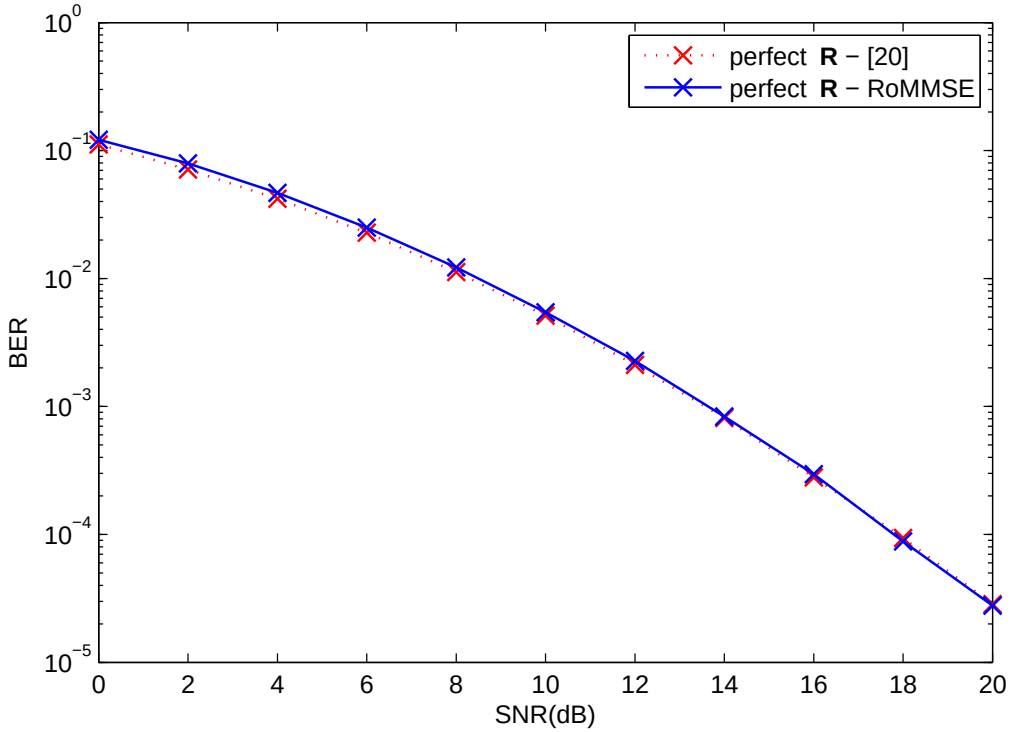



Figure 3.10: BER vs. SNR performance comparison between RoMMSE and [55] for spatially correlated $2 \times 2$ MIMO channel. $\Delta = 5°$, $d_t = 0.5\lambda$ and $d_r = 0.2\lambda$. $\varepsilon = 0.3$.

spatial correlation. Complete decorrelation of $\mathbf{R}$ is attained as $\varepsilon \to \infty$. In addition, it has been observed that when the threshold $\alpha$ is met, $\mathbf{E}$ and $\widetilde{\mathbf{C}}^2$ share the same eigenvector matrix as $\widehat{\mathbf{R}}$. This has been proven analytically in Appendix 3.4. When there is no spatial correlation mismatch, it was shown in [14] that the transmitted signal corresponds to transmitting in specific eigenmodes of the spatial correlation, which determines which particular eigenmode of the channel will be estimated. Furthermore, the power on each eigenmode is determined by waterfilling solution based on some optimization criteria, such as minimum MSE and maximum conditional mutual information. When the SNR is low, it was found that all the power will be allocated to the strongest eigenmode. However, when the SNR is high, power is evenly distributed among all eigenmodes. When spatial correlation mismatch has been accounted, it can be seen from the simulations that no matter if the system is operating under low or high SNR, the mismatch matrix $\mathbf{E}$ not

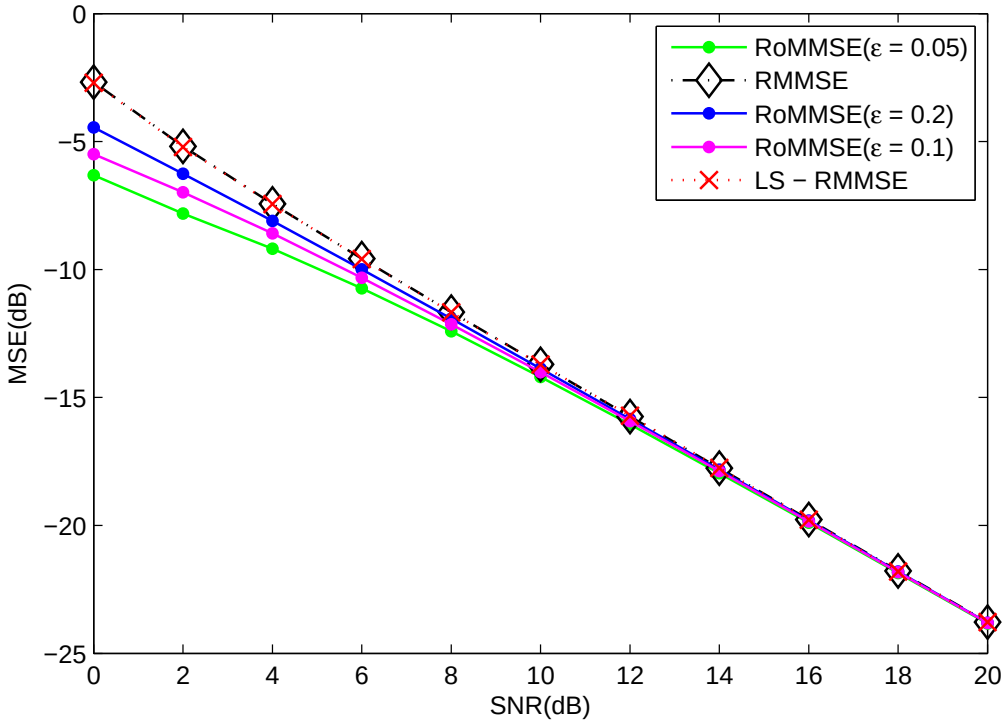



Figure 3.11: MSE vs. SNR performance comparison between RoMMSE, LS-RMMSE and RMMSE [23] for spatially correlated $2 \times 2$ MIMO system. $\Delta = 5°$, $d_t = 0.5\lambda$, $d_r = 0.2\lambda$.

only decorrelates the channel, but it also equalizes all the diagonal value of $\mathbf{R}$ such that $tr\left(\widehat{\mathbf{R}} + \mathbf{E}\right) = N_r N_t$ given that $\varepsilon$ is sufficiently large. Note that this is also true even when $tr\left(\widehat{\mathbf{R}} + \mathbf{E}\right) = N_r N_t$ is not a constraint in (3.17). Hence, the robust training sequence evenly distributes power across all eigenmodes. This is summarizes in Table 3.1. This can easily be explained due to the fact that the worst-case mismatch matrix $\mathbf{E}_w$ can be obtained only when $\widehat{\mathbf{R}}$, $\mathbf{E}$, and $\widetilde{\mathbf{C}}^2$ are all diagonalized and that $\mathbf{E}$ and $\widetilde{\mathbf{C}}^2$ share the same eigenvectors as $\widehat{\mathbf{R}}$ (see Appendix 3.4). With the constraint that $tr\left(\widehat{\mathbf{R}} + \mathbf{E}\right) = N_r N_t$, this forces $\mathbf{E}$ to diagonalize $\widehat{\mathbf{R}}$ and equalizes the diagonal values of $\widehat{\mathbf{R}}$ such that the constraint is satisfied. Hence, the mismatch matrix $\mathbf{E}$ will evenly distribute power across all the eigenmodes of $\mathbf{R}$. If $\varepsilon$ is not sufficient large, that $\mathbf{E}$ will not have enough degrees of freedom to diagonalize and equalize the diagonal values of $\widehat{\mathbf{R}}$.

Performance of the RoMMSE estimator is also compared to that of the RMMSE and

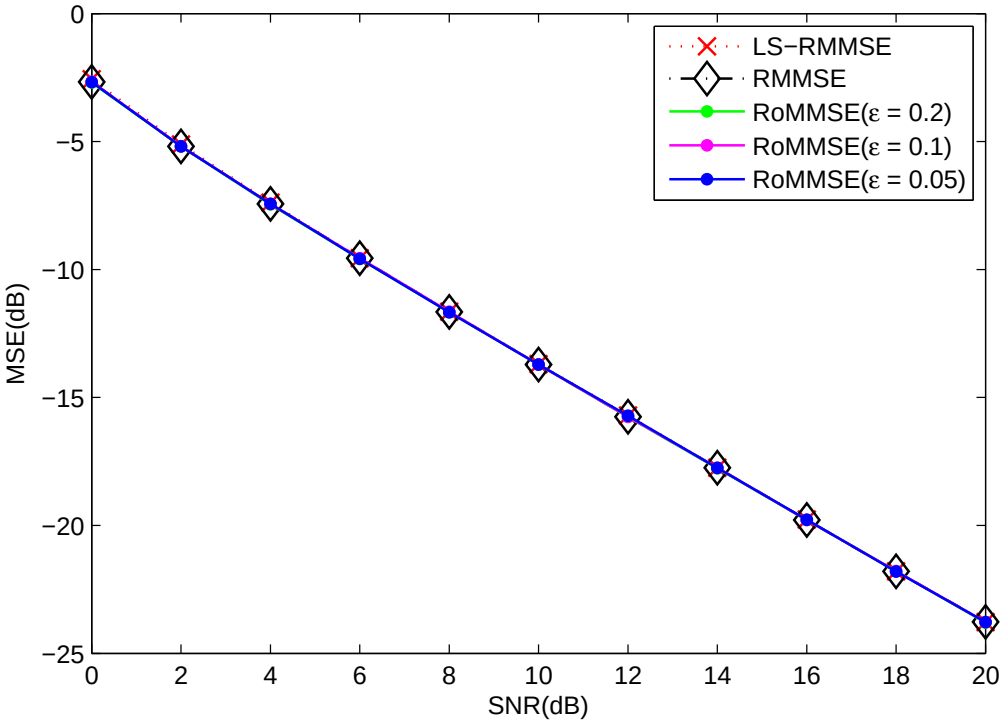



Figure 3.12: MSE vs. SNR performance comparison between RoMMSE, LS-RMMSE and RMMSE [23] for spatially uncorrelated $2 \times 2$ MIMO system, i.e. $\mathbf{R} = \mathbf{I}_{N_r N_t}$. $\varepsilon = 0.3$.

LS-RMMSE when the MIMO channel is spatially uncorrelated. Figure 3.12 indicates that in this situation all three estimators render similar MSE performance, which suggests that all three estimators are identical. This is indeed the case as it is proven in Appendix 3.5.

Data detection performance for RMMSE and RoMMSE algorithms are also compared in the case of spatially correlated and uncorrelated channels. Since the RMMSE algorithm is proposed in time-multiplexed pilots scheme, both MSE and BER are compared using

Table 3.1: Number of eigenmodes used during channel estimation for spatially correlated MIMO channel.

| | Low SNR | High SNR |
|---|---|---|
| $\mathbf{E} = \mathbf{0}_{N_r N_t \times N_r N_t}$ | one | all |
| $\mathbf{E} \neq \mathbf{0}_{N_r N_t \times N_r N_t}$ | all | all |

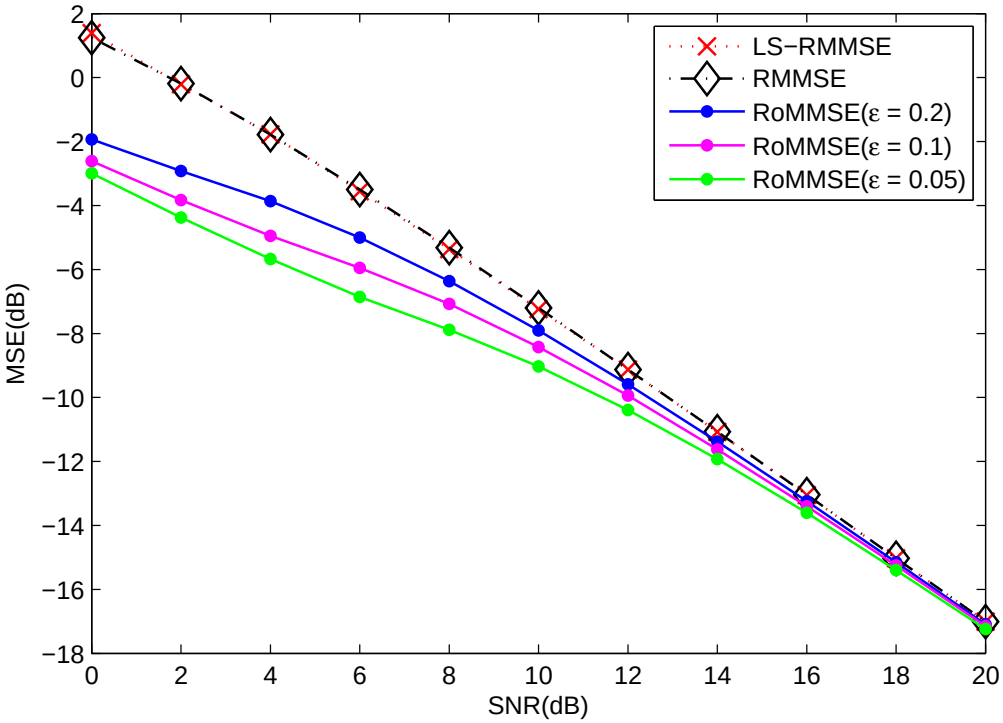



Figure 3.13: MSE vs. SNR performance comparison using time-multiplexed pilots between RoMMSE, LS-RMMSE and RMMSE [23] for spatially correlated $2 \times 2$ MIMO system. $\Delta = 5°$, $d_t = 0.5\lambda$, $d_r = 0.2\lambda$.

time-multiplexed pilots. Channel estimation performance using time-multiplexed pilots is shown in Figure 3.13, where the RoMMSE algorithm outperforms the RMMSE and LS-RMMSE algorithms, similar to the performance shown in Figure 3.11. Notice that the estimation performance of the RoMMSE algorithm in Figure 3.13 is worse than that shown in Figure 3.11. This is because the power of the pilot is less than that of the SIT sequence. Next, the BER performance comparison is shown in Figure 3.14. With $\varepsilon = 0.05$, RoMMSE outperforms the RMMSE by 2.5 dB in the low SNR region.

The BER performance of the RoMMSE algorithm vs. different values of $\varepsilon$ is shown in Figure 3.15. Notice that lower BER is obtained with increasing $\varepsilon$. As previously explained, this is because the mismatch matrix has more freedom to decorrelate the spatial correlation matrix as $\varepsilon$ increases, which enhances the spatial diversity of the system, thus

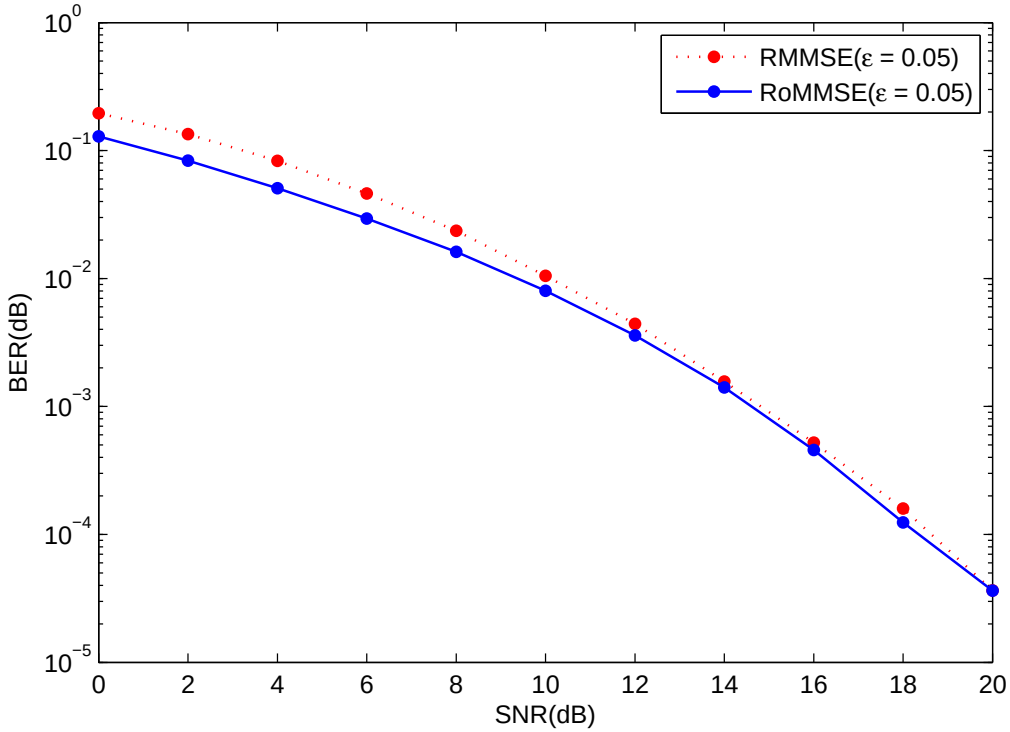



Figure 3.14: BER vs. SNR performance comparison using time-multiplexed pilots between RoMMSE and RMMSE [23] for spatially correlated $2 \times 2$ MIMO channel. $\Delta = 5°$, $d_t = 0.5\lambda$ and $d_r = 0.2\lambda$. $\varepsilon = 0.05$.

improving the BER performance.

Table 3.2 shows the number of iterations needed before the algorithm in Figure 3.4 converges under different initial conditions for $\mathbf{C}(0)$. It shows that the proposed iterative algorithm always, on the average, converges faster if $\mathbf{C}(0)$ is initialized to be an orthogonal matrix than when it is initialized to be a random matrix. Even though the table only shows performance when SNR is 5 dB, this convergence behavior has been observed for all the SNR values that have been tested. This speed up is due to the fact that the worst-case MSE is achieved if the argument inside the trace operator in (3.17) forms a diagonal matrix. Hence, if $\mathbf{C}(0)$ is initialized to be an orthogonal matrix, $\widetilde{\mathbf{C}}(n)$, for $n > 0$, will be closer to the optimal solution than when $\mathbf{C}(0)$ is initialized to be a random matrix as it is already equaled to a diagonal matrix of the form $\alpha \mathbf{I}_{N_r N_t}$, for $\alpha$ being equal to an

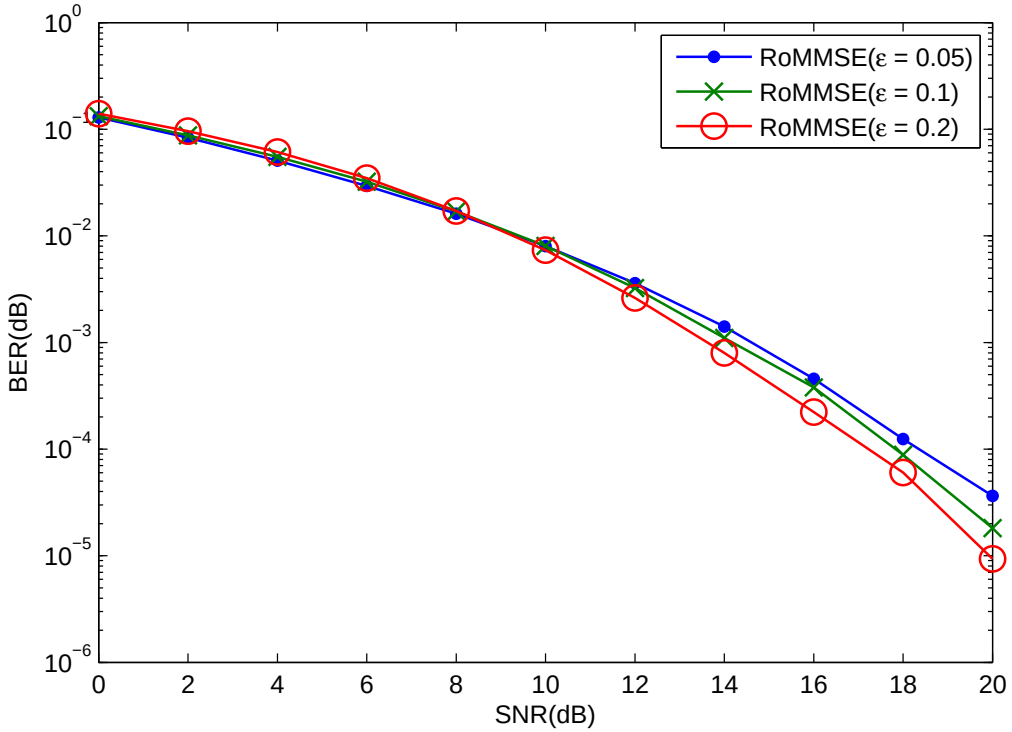



Figure 3.15: BER vs. SNR performance comparison using time-multiplexed pilots of RoMMSE with different $\epsilon$ for spatially correlated $2\times2$ MIMO channel. $\Delta = 5°$, $d_t = 0.5\lambda$ and $d_r = 0.2\lambda$.

arbitrary constant.

## 3.4 Decorrelation of $\widehat{\mathbf{R}}$

Inserting (3.14) into (3.11), and noting that the SVDs of $\widetilde{\mathbf{C}}^H\widetilde{\mathbf{C}}$ and $\widehat{\mathbf{R}}$ are $\mathbf{V}_{\widetilde{\mathbf{C}}^2}\mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2}\mathbf{V}_{\widetilde{\mathbf{C}}^2}^H$ and $\mathbf{U}_{\hat{\mathbf{R}}}\mathbf{\Lambda}_{\hat{\mathbf{R}}}\mathbf{U}_{\hat{\mathbf{R}}}^H$, respectively. The sequence design problem in (3.11) is equivalent to solving (3.17) and (3.18) iteratively. Using the SVD, (3.17) can be rewritten as

$$\min_{\|\mathbf{E}\|_F\le\varepsilon} tr\left(\left[\mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2} + \mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2}\mathbf{V}_{\widetilde{\mathbf{C}}^2}^H\left(\mathbf{U}_{\hat{\mathbf{R}}}\mathbf{\Lambda}_{\hat{\mathbf{R}}}\mathbf{U}_{\hat{\mathbf{R}}}^H + \mathbf{E}\right)\mathbf{V}_{\widetilde{\mathbf{C}}^2}\mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2}\right]^{-1}\right).$$

Defining $\mathbf{A} \triangleq \mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2} + \mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2}\mathbf{V}_{\widetilde{\mathbf{C}}^2}^H\left(\mathbf{U}_{\hat{\mathbf{R}}}\mathbf{\Lambda}_{\hat{\mathbf{R}}}\mathbf{U}_{\hat{\mathbf{R}}}^H + \mathbf{E}\right)\mathbf{V}_{\widetilde{\mathbf{C}}^2}\mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2}$. The objective function can then be written as $f(\boldsymbol{\lambda}(\mathbf{A})) = \sum_i \frac{1}{\lambda_i(\mathbf{A})}$, where $\boldsymbol{\lambda}(\mathbf{A})$ denotes a vector composed of

Table 3.2: Average number of iterations required for convergence for the proposed RoMMSE estimator. $d_t = 0.5\lambda$, $d_r = 0.2\lambda$, $\varepsilon = 0.3$ and SNR = 5 dB, angular spread = 15° ($4 \times 4$), 5° ($2 \times 2$).

| $N_t \times N_r$ | $\mathbf{C}(0)$ | Average number of iterations |
|---|---|---|
| $4 \times 4$ | randomly generated | 5.742 |
| $4 \times 4$ | orthogonal | 4.000 |
| $2 \times 2$ | randomly generated | 3.524 |
| $2 \times 2$ | orthogonal | 3.000 |

eigenvalues of $\mathbf{A}$. Since $\phi(\lambda_i(\mathbf{A})) = \frac{1}{\lambda_i(\mathbf{A})}$ is a convex function, $f(\boldsymbol{\lambda}(\mathbf{A}))$ is Schur-convex [14]. Moreover, $\mathbf{A}$ is a symmetric matrix. Therefore, $f(\boldsymbol{\lambda}(\mathbf{A}))$ majorizes $f(\mathbf{d}(\mathbf{A}))$, i.e. $f(\boldsymbol{\lambda}(\mathbf{A})) \geq f(\mathbf{d}(\mathbf{A}))$ [14], where $\mathbf{d}(\mathbf{A})$ denotes a vector which is composed of diagonal values of $\mathbf{A}$. Since the equality will hold when $\mathbf{A}$ is a diagonal matrix, then the worst case mismatch error will be $\mathbf{E}_w = \mathbf{V}_{\widetilde{\mathbf{C}}} \boldsymbol{\Lambda}_{\mathbf{E}} \mathbf{V}_{\widetilde{\mathbf{C}}}^H$, which ensures that the lower bound of the MSE is reached. This implies that $\mathbf{V}_{\widetilde{\mathbf{C}}^2}^H \mathbf{U}_{\widehat{\mathbf{R}}}$ has to be a diagonal matrix and that $\mathbf{E}$ shares the same eigenvectors as $\widehat{\mathbf{R}}$. The first condition can thus be achieved if $\widehat{\mathbf{R}}$ and $\widetilde{\mathbf{C}}^2$ also share the same eigenvector matrix.

## 3.5 Comparison of RoMMSE and RMMSE estimators

When the MIMO system is spatially uncorrelated, i.e. $\mathbf{R} = \mathbf{I}_{N_r N_t}$, the channel estimate from the RMMSE channel estimator in [23] becomes

$$\begin{aligned}\widehat{\mathbf{h}}_{RMMSE} &= \widetilde{\mathbf{C}}^H \left( \widetilde{\mathbf{C}} \widetilde{\mathbf{C}}^H + \frac{\sigma_{\mathbf{nn}}^2 N_r N_t}{tr\left(\widehat{\mathbf{R}}\right)} \mathbf{I}_{N_r N_t} \right)^{-1} \mathbf{y} \\ &= \widetilde{\mathbf{C}}^H \left( \widetilde{\mathbf{C}} \widetilde{\mathbf{C}}^H + \sigma_{\mathbf{nn}}^2 \mathbf{I}_{N_r N_t} \right)^{-1} \mathbf{y}.\end{aligned}$$

It is assumed orthogonal sequences are employed for the RMMSE channel estimator [23], i.e. $\mathbf{C}^H\mathbf{C} = \frac{P_T}{N_t}\mathbf{I}$. The estimate of the RoMMSE estimator is written as

$$\begin{aligned}\widehat{\mathbf{h}}_{RoMMSE} &= \left(\widehat{\mathbf{R}} + \mathbf{E}_w\right)\widetilde{\mathbf{C}}^H\left(\widetilde{\mathbf{C}}\left(\widehat{\mathbf{R}} + \mathbf{E}_w\right)\widetilde{\mathbf{C}}^H + \sigma^2_{\mathbf{nn}}\mathbf{I}_{N_rN_t}\right)^{-1}\mathbf{y} \\ &= \left(\mathbf{I}_{N_rN_t} + \mathbf{E}_w\right)\widetilde{\mathbf{C}}^H\left(\widetilde{\mathbf{C}}(\mathbf{I}_{N_rN_t} + \mathbf{E}_w)\widetilde{\mathbf{C}}^H + \sigma^2_{\mathbf{nn}}\mathbf{I}_{N_rN_t}\right)^{-1}\mathbf{y},\end{aligned}$$

where $\mathbf{E}_w$ is the worst case error of the estimated spatial correlation. Let $\mathbf{V}_{\widetilde{\mathbf{C}}^2}$ and $\mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2}$ to denote the eigenvector and eigenvalue matrix of $\widetilde{\mathbf{C}}^2$, respectively. From (3.13), (3.17) is equivalent to

$$\begin{aligned}&\min_{\|\mathbf{E}\|_F \le \varepsilon} tr\left(\left[\mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2} + \mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2}\mathbf{V}^H_{\widetilde{\mathbf{C}}^2}\left(\mathbf{I}_{N_rN_t} + \mathbf{E}\right)\mathbf{V}_{\widetilde{\mathbf{C}}^2}\mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2}\right]^{-1}\right) \\ = &\min_{\|\mathbf{E}\|_F \le \varepsilon} tr\left(\left[\mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2} + \mathbf{\Lambda}^2_{\widetilde{\mathbf{C}}^2} + \mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2}\mathbf{V}^H_{\widetilde{\mathbf{C}}^2}\mathbf{E}\mathbf{V}_{\widetilde{\mathbf{C}}^2}\mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2}\right]^{-1}\right). \qquad (3.28)\end{aligned}$$

Defining $\mathbf{A} \triangleq \mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2} + \mathbf{\Lambda}^2_{\widetilde{\mathbf{C}}^2} + \mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2}\mathbf{V}^H_{\widetilde{\mathbf{C}}^2}\mathbf{E}\mathbf{V}_{\widetilde{\mathbf{C}}^2}\mathbf{\Lambda}_{\widetilde{\mathbf{C}}^2}$. The objective function can then be written as $f(\boldsymbol{\lambda}(\mathbf{A})) = \sum_i \frac{1}{\lambda_i(\mathbf{A})}$, where $\boldsymbol{\lambda}(\mathbf{A})$ denotes a vector composed of eigenvalues of $\mathbf{A}$. Since $\phi(\lambda_i(\mathbf{A})) = \frac{1}{\lambda_i(\mathbf{A})}$ is a convex function, $f(\boldsymbol{\lambda}(\mathbf{A}))$ is Schur-convex [14]. Moreover, $\mathbf{A}$ is a symmetric matrix. Therefore, $f(\boldsymbol{\lambda}(\mathbf{A}))$ majorizes $f(\mathbf{d}(\mathbf{A}))$, i.e. $f(\boldsymbol{\lambda}(\mathbf{A})) \ge f(\mathbf{d}(\mathbf{A}))$ [14], where $\mathbf{d}(\mathbf{A})$ denotes a vector which is composed of diagonal values of $\mathbf{A}$. Since the equality will hold when $\mathbf{A}$ is a diagonal matrix, then the worst case mismatch error will be $\mathbf{E}_w = \mathbf{V}_{\widetilde{\mathbf{C}}}\mathbf{\Lambda}_{\mathbf{E}}\mathbf{V}^H_{\widetilde{\mathbf{C}}}$, which ensures that the lower bound of the MSE is reached. Since $\widehat{\mathbf{R}} = \mathbf{I}_{N_rN_t}$, from Section 3.3, $\mathbf{E}_w$ will have to be a diagonal matrix (or a linear combination of one) as not to minimize the degrees of freedom in the MIMO channel. This implies that either $\mathbf{V}_{\widetilde{\mathbf{C}}^2}$ or $\mathbf{\Lambda}_{\mathbf{E}}$ is an identity matrix. However, since $\widetilde{\mathbf{C}}^2$ is not necessarily a diagonal matrix, it is not necessary for $\mathbf{V}_{\widetilde{\mathbf{C}}^2}$ to be an identity matrix. This implies that $\mathbf{\Lambda}_{\mathbf{E}}$ must be either an identity or an all zero matrix. Since the constraint $tr(\widehat{\mathbf{R}} + \mathbf{E}) = N_tN_r$ must be satisfied, therefore, $\mathbf{E}_w$ must be an all zero matrix, i.e. $\mathbf{E}_w = \mathbf{0}_{N_rN_t \times N_rN_t}$. Substituting $\mathbf{E}_w = \mathbf{0}_{N_rN_t \times N_rN_t}$ into (3.11) and solving for $\mathbf{C}$, $\mathbf{C}$ becomes an orthogonal matrix. That is, the optimal SIT sequence is an orthogonal sequence, which agrees with the conclusion in [23] that the optimal training sequence for spatially uncorrelated MIMO channel is an

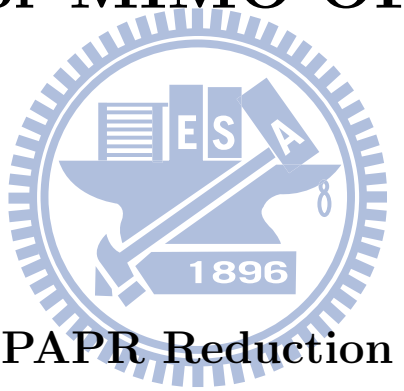
orthogonal sequence. Therefore, $\widehat{\mathbf{h}}_{RoMMSE}$ becomes

$$\widehat{\mathbf{h}}_{RoMMSE} = \widetilde{\mathbf{C}}^H \left( \widetilde{\mathbf{C}} \widetilde{\mathbf{C}}^H + \sigma_{\mathbf{nn}}^2 \mathbf{I}_{N_r N_t} \right)^{-1} \mathbf{y},$$

which implies that the estimation performance of the proposed RoMMSE estimator and the RMMSE estimator is identical when the MIMO channel is spatially uncorrelated, thus agreeing with the simulation results in Section 3.3.

# Chapter 4

# Joint Sequence Design for Robust Channel Estimation and PAPR Reduction for MIMO-OFDM Systems



## 4.1 Overview of PAPR Reduction

To combat the unfavorable effects brought about by wideband channels, OFDM is a promising candidate that can easily remove intersymbol interference without inducing great penalty in computational complexity. Unfortunately, due to the use of the IFFT at the transmitter, the amplitude of the transmitted signal is no longer constrained, thus incurring a high peak-to-average power ratio. The problem is worsened when the OFDM systems are combined with MIMO as more RF chains are required for transmission where different antennas may exhibit varying large degrees of PAPR. It is well known that non-linearity in high power amplifier causes distortion in the transmit signal. Such distortion can lead to undesirable spectral regrowth, thus interfering with signals in the neighboring

subcarriers. Large amount of distortion can also cause in-band self-interference, which increases bit error rate. Hence, it is customary for power amplifiers to operate with a certain power backoff, which is defined as the ratio of maximum saturation output power to lower average output power [34]. However, such backoff schemes lowers the efficiency of the power amplifier and increases overall power consumption.

Miscellaneous approaches have been proposed to reduce the PAPR including coding, TR, TI and multiple signal representation such as PTS and SLM and interleaving. In chapter 2.2.3 , advantages and disadvantages of the existing techniques are discussed in several criterion such as distortion, power increase, data rate loss and implementation complexity. Since distortion is the main contribution for the BER degradation, a distortionless technique is the first concern. For the techniques like PTS, SLM and interleaving, although they are distortionless, side information is needed. Side information should be estimated correctly, otherwise, BER performance gets worsened.

Herein, a superimposed sequence design using per-tone affine precoding technique is proposed to reduce the peak-to-average power ratio (PAPR) for MIMO-OFDM systems and estimate MIMO-OFDM channels even if spatial correlation uncertainty exists. The proposed technique can easily trade-off between BER and PAPR reduction performance. Moreover, it does not require side information to be transmitted for the removal of the sequence at the receiver, and the transmit redundancy can be as small as 1 symbol/subcarrier. The superimposed sequence is designed using linear programming and has a computational complexity of $O\left(N\log(N)\right)$. Simulation results have shown that the proposed technique, which shall be called superimposed sequence for PAPR reduction, or SIPR, outperforms methods such as tone reservation in terms of PAPR reduction. The chapter is organized as follows. The system model and a detailed description of SIPR are given in Section 4.2, followed by the simulation results in Section 4.3.

## 4.2 Methodology

### 4.2.1 System Model

Consider an affine precoded MIMO-OFDM system with $N_t$ transmit and $N_r$ receive antennas as shown in Figure 4.1. $N_f$-point DFT and IDFT are adopted. $\mathbf{u}(k) = \begin{bmatrix} u(kN_s) & u(kN_s+1) & \cdots & u(kN_s+N_s-1) \end{bmatrix}^T$ is the information-bearing signal vector, with $k$ denoting the block index and $N_s$ denoting the block size. Without loss of generality, $N_t = 2$ is assumed such that $2 \times 2$ Alamouti STBC is used. The STBC encodes two adjacent OFDM symbols at a time for each subcarrier for a total of $K \geq N_t$ symbols. The output signal from the $i^{th}$ subcarrier at the $\ell^{th}$ antenna can then be grouped together to form $\mathbf{x}_{i,\ell} \in \mathbb{C}^K$, for $i = 1, 2, \ldots, N_f$. Defining $\mathbf{X}^i \triangleq \begin{bmatrix} \mathbf{x}_{i,1} & \mathbf{x}_{i,2} & \cdots & \mathbf{x}_{i,N_t} \end{bmatrix}^T \in \mathbb{C}^{N_t \times K}$. The information-bearing signal can then be written as an $N_f N_t \times N_f K$ complex-valued matrix

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}^1 & & & \\ & \mathbf{X}^2 & & \\ & & \ddots & \\ & & & \mathbf{X}^{N_f} \end{bmatrix} \tag{4.1}$$

$\mathbf{X}$ is then premultiplied by the precoder matrix $\mathbf{P} = \begin{bmatrix} \mathbf{P}_1^T & \mathbf{P}_2^T & \cdots & \mathbf{P}_{N_f}^T \end{bmatrix}^T \in \mathbb{C}^{N_f K \times (K+L)}$ to form

$$\mathbf{G} = \mathbf{XP} = \begin{bmatrix} \mathbf{X}^1\mathbf{P}_1 \\ \mathbf{X}^2\mathbf{P}_2 \\ \vdots \\ \mathbf{X}^{N_f}\mathbf{P}_{N_f} \end{bmatrix} \in \mathbb{C}^{N_f N_t \times (K+L)}, \tag{4.2}$$

where $\mathbf{P}_i = \begin{bmatrix} \mathbf{p}_{i,1} & \mathbf{p}_{i,2} & \cdots & \mathbf{p}_{i,K} \end{bmatrix}^T \in \mathbb{C}^{K \times (K+L)}$ is the per-tone affine precoder matrix which adds $L \geq N_t$ redundant symbols to the signal at the $i^{th}$ subcarrier. $\mathbf{p}_{i,j} \in \mathbb{C}^{K+L}$ denotes the precoder vector for $j = 1, 2, \ldots K$. The premultiplication of the $i^{th}$ per-tone

affine precoder to the data matrix $\mathbf{X}^i$ in (4.2) runs in contrast to conventional precoding techniques where the precoding matrix is postmultiplied to the data matrix. It is clear that $\mathbf{x}_{i,\ell}^T\mathbf{P}_i$ belongs to the range space of $\mathbf{P}_i^T$, which as will be explained in the sequel, eases the removal of the superimposed sequences as the sequences and data belongs to different orthogonal subspaces. As explained in [55, 56], removal of the sequences will be more difficult if the precoding matrix is postmultiplied to the data matrix as the information-bearing signal belongs to the range space of the channel, which cannot guarantee to be orthogonal to the spaces occupied by the sequences. For ease of derivation, the precoded data shall be presented as

$$\mathbf{D}_\ell = \mathbf{G}\left(\ell : N_t : \ell + (N_f - 1)N_t, :\right). \tag{4.3}$$

Defining the $N_tN_f \times (K+L)$ complex-valued aggregate precoded data matrix $\mathbf{D} = \begin{bmatrix} \mathbf{D}_1^T & \mathbf{D}_2^T & \cdots & \mathbf{D}_{N_t}^T \end{bmatrix}^T$, the $N_fN_t \times (K+L)$ complex-valued superimposed training (SIT) sequence matrix, $\mathbf{C}_1 = \begin{bmatrix} \mathbf{C}_{1,1}^T & \mathbf{C}_{1,2}^T & \cdots & \mathbf{C}_{1,N_t}^T \end{bmatrix}^T$, as proposed in [56], and the complex-valued SIPR matrix $\mathbf{C}_2 = \begin{bmatrix} \mathbf{C}_{2,1}^T & \mathbf{C}_{2,2}^T & \cdots & \mathbf{C}_{2,N_t}^T \end{bmatrix}^T \in \mathbb{C}^{N_fN_t\times(K+L)}$ of the same size, where $\mathbf{C}_{1,\ell} = \begin{bmatrix} \mathbf{c}_{1,1,\ell} & \cdots & \mathbf{c}_{1,N_f,\ell} \end{bmatrix}^T \in \mathbb{C}^{N_f\times(K+L)}$ and $\mathbf{C}_{2,\ell} = \begin{bmatrix} \mathbf{c}_{2,1,\ell} & \cdots & \mathbf{c}_{2,N_f,\ell} \end{bmatrix}^T \in \mathbb{C}^{N_f\times(K+L)}$ with $\mathbf{c}_{1,i,\ell}, \mathbf{c}_{2,i,\ell} \in \mathbb{C}^{K+L}$ denoting the SIT and SIPR vector for the $\ell^{th}$ antenna at the $i^{th}$ subcarrier, respectively. The superimposed sequence matrices will then be added to the precoded data before the IFFT resulting in

$$\mathbf{U} = \mathbf{C} + \mathbf{D} = \mathbf{C}_1 + \mathbf{C}_2 + \mathbf{D} \in \mathbb{C}^{N_tN_f\times(K+L)}, \tag{4.4}$$

as illustrated in Figure 4.1.

Combining IFFT, cyclic prefix (CP) insertion, CP removal, and FFT with the frequency-

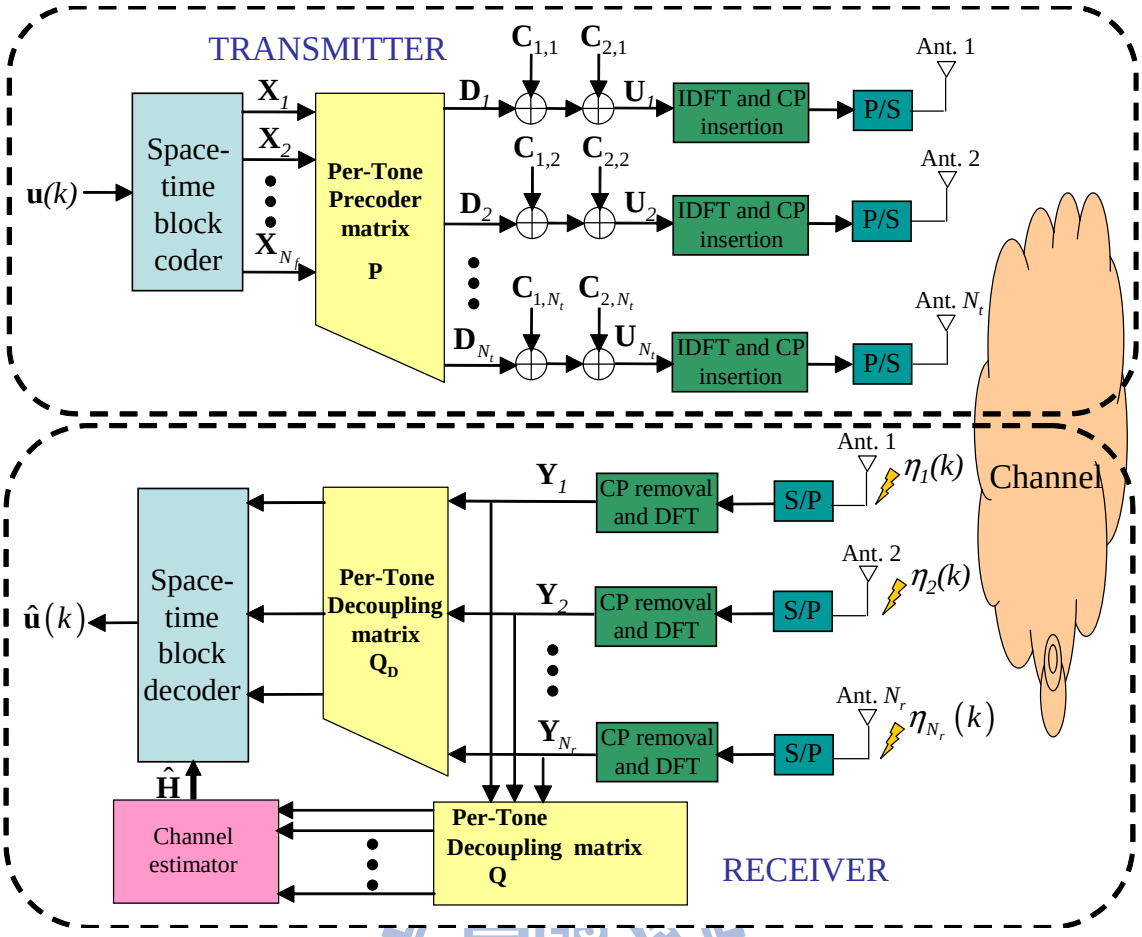



Figure 4.1: Block diagram of MIMO-OFDM transceiver with affine precoder.

selective MIMO channel results in an equivalent channel matrix

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} & \cdots & \mathbf{H}_{1N_t} \\ \mathbf{H}_{21} & \ddots & & \mathbf{H}_{2N_t} \\ \vdots & & \ddots & \vdots \\ \mathbf{H}_{N_r1} & \cdots & \cdots & \mathbf{H}_{N_rN_t} \end{bmatrix} \in \mathbb{C}^{N_fN_r \times N_fN_r}, \tag{4.5}$$

where $\mathbf{H}_{m\ell} \in \mathbb{C}^{N_f \times N_f}$ is a diagonal matrix containing the Fourier coefficients of the channel between the $m^{th}$ receive and $\ell^{th}$ transmit antenna. Thus, the receive data after CP removal can be written as

$$\begin{aligned} \mathbf{Y} &= \mathbf{HU} + \mathbf{W} \\ &= \mathbf{H}\left(\mathbf{C}_1 + \mathbf{C}_2 + \mathbf{D}\right) + \mathbf{W}, \end{aligned} \tag{4.6}$$

where $\mathbf{W}$ denotes the channel noise matrix of appropriate size.

### 4.2.2 Proposed Sequence Design

It is clear from (4.6) that recovery of the data-bearing signal embedded inside $\mathbf{D}$ requires the removal of $\mathbf{C}_1$, $\mathbf{C}_2$ and $\mathbf{W}$, which can only be carried out after channel estimation has been performed using $\mathbf{C}_1$. However, $\mathbf{C}_2$ will have to be removed during the channel estimation and the detection phase as $\mathbf{C}_2$ is only exploited by the transmitter.

Defining

$$\mathbf{C}_1^i \triangleq \mathbf{C}_1\left(i : N_f : i + (N_t - 1)N_f, :\right)$$

$$\mathbf{C}_2^i \triangleq \mathbf{C}_2\left(i : N_f : i + (N_t - 1)N_f, :\right)$$

$$\mathbf{Y}^i \triangleq \mathbf{Y}\left(i : N_f : i + (N_r - 1)N_f. :\right)$$

for ease for derivation, where $^i$ denotes the index for subcarrier. In previous techniques [55, 56], the decoupling of the SIT sequence $\mathbf{C}_1^i$ can be achieved by premultiplying $\mathbf{Y}^i$ by the per-tone decoder $\mathbf{Q}_i = \begin{bmatrix} \mathbf{q}_{i,1} & \mathbf{q}_{i,2} & \cdots & \mathbf{q}_{i,K+L} \end{bmatrix}^T \in \mathbb{C}^{(K+L)\times N_t}$, for $i = 1, 2 \ldots N_f$, where $\mathbf{Q}_i \in \mathcal{N}\left(\mathbf{P}_i\right)$, resulting in

$$\begin{aligned} \mathbf{Y}^i\mathbf{Q}_i &= \mathbf{H}^i\left(\mathbf{X}^i\mathbf{P}_i + \mathbf{C}_1^i + \mathbf{C}_2^i\right)\mathbf{Q}_i + \mathbf{W}\mathbf{Q}_i, \\ &= \mathbf{H}^i\left(\mathbf{C}_1^i + \mathbf{C}_2^i\right)\mathbf{Q}_i + \mathbf{W}\mathbf{Q}_i \end{aligned}$$

where $\mathbf{H}^i \in \mathbb{C}^{N_r\times N_t}$ is the matrix containing the Fourier coefficients at the $i^{th}$ subcarrier for all transmit and receive antennas. In other words, the SIT vector $c_{1,i,\ell}$ should lie in the column space of $\mathbf{Q}_i$. Therefore, the condition that $\mathbf{C}_1^i\mathbf{P}_i^H = \mathbf{0}_{N_fN_t\times K}$ guarantees the subspaces spanned by the vectors in $\mathbf{P}_i$ and $\mathbf{C}_1^i$ are complementary. In [56], the SVD of $\mathbf{C}_1^i$ is written as

$$\mathbf{C}_1^i = \mathbf{U}_{\mathbf{C}_1^i}\begin{bmatrix}\mathbf{\Sigma}_{\mathbf{C}_1^i} & \mathbf{0}_{N_t\times(K+L-N_t)}\end{bmatrix}\mathbf{V}_{\mathbf{C}_1^i}^H, \tag{4.7}$$

where $\mathbf{U}_{\mathbf{C}_1^i}$, $\mathbf{V}_{\mathbf{C}_1^i}$, and $\mathbf{\Sigma}_{\mathbf{C}_1^i}$ are the left and singular vector matrix of $\mathbf{C}_1^i$, and the invertible

portion of the singular value matrix of $\mathbf{C}_1^i$, respectively. Notice that $\mathbf{C}\mathbf{P}^H = \mathbf{0}_{N_t \times K}$ if $\mathbf{V}_{\mathbf{C}_1^i} = \mathbf{U}_{\mathbf{Q}_i}$, where $\mathbf{U}_{\mathbf{Q}_i}$ is the eigenvector matrix of $\mathbf{Q}_i\mathbf{Q}_i^H$.

In the detection phase, the information-bearing signal can be recovered by premultiplying $\mathbf{Y}$ by the decoding matrix $\mathbf{Q}_{D_i} = \mathbf{P}_i^H \left(\mathbf{P}_i\mathbf{P}_i^H\right)^{-1} \in \mathbb{C}^{(K+L)\times K}$, i.e.

$$\begin{aligned}\mathbf{Y}^i\mathbf{Q}_{D_i} &= \mathbf{H}^i \left(\mathbf{X}^i\mathbf{P}_i + \mathbf{C}_1^i + \mathbf{C}_2^i\right)\mathbf{Q}_{D_i} + \mathbf{W}\mathbf{Q}_{D_i}, \\ &= \mathbf{H}^i \left(\mathbf{X}^i + \mathbf{C}_2^i\mathbf{Q}_{D_i}\right) + \mathbf{W}\mathbf{Q}_{D_i}.\end{aligned}$$

Hence, a simple way to design $\mathbf{P}_i$ and $\mathbf{Q}_i$ is by extracting components off of an orthogonal matrix $\mathbf{O}_i \in \mathbb{C}^{(K+L)\times(K+L)}$, i.e.

$$\begin{aligned}\mathbf{P}_i &= \sqrt{\frac{K+L}{K}}\mathbf{O}_i\left(1:1:K,:\right) \in \mathbb{C}^{K\times(K+L)}, \\ \mathbf{Q}_i &= \mathbf{O}_i^H\left((K+1):1:(K+N_t),:\right) \in \mathbb{C}^{(K+L)\times N_t}.\end{aligned}$$

Unfortunately, this is insufficient for the purpose herein as $\mathbf{C}_2^i$ remains. However, suppose

$$\mathbf{C}_2^i = \mathbf{B}_i\mathbf{M}_i, \tag{4.8}$$

where $\mathbf{B}_i = \begin{bmatrix} \mathbf{b}_{i,1} & \mathbf{b}_{i,2} & \cdots & \mathbf{b}_{i,N_t} \end{bmatrix}^T \in \mathbb{C}^{N_t\times(L-N_t)}$ contains arbitrary scalars, and

$$\mathbf{M}_i = \mathbf{O}_i\left((K+N_t+1):1:(K+L),:\right) \tag{4.9}$$

denotes the matrix extracted from the remaining rows of the orthogonal matrix $\mathbf{O}_i$ after $\mathbf{P}_i$ and $\mathbf{Q}_i$ have been extracted. That is, $\mathbf{C}_2^i$ is a linear combination of the remaining rows of the orthogonal matrix $\mathbf{O}_i$ after $\mathbf{P}_i$ and $\mathbf{Q}_i$ have been extracted. Therefore, the row vectors of $\mathbf{D}^i$, $\mathbf{C}_1^i$, and $\mathbf{C}_2^i$ are guarantee to be orthogonal, as illustrated in Figure 4.2. Then, $\mathbf{C}_2^i$ will satisfy the constraints

$$\begin{aligned}\mathbf{C}_2^i\mathbf{Q}_i &= \mathbf{0}_{N_r\times N_t}, \\ \mathbf{C}_2^i\mathbf{Q}_{D_i} &= \mathbf{0}_{N_r\times K},\end{aligned} \tag{4.10}$$

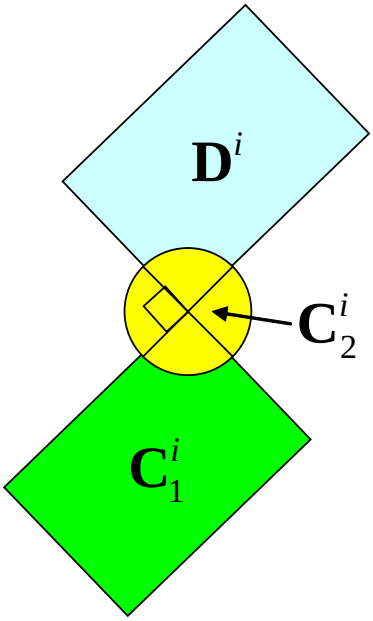



Figure 4.2: Three-dimensional illustration of the orthogonality between the row space of $\mathbf{D}^i$, $\mathbf{C}_1^i$, and $\mathbf{C}_2^i$. Note that the row space of $\mathbf{C}_2^i$ is coming out of the page.

thus allowing the SIPR matrix $\mathbf{C}_2^i$ to be completely eliminated by $\mathbf{Q}_i$ and $\mathbf{Q}_{D_i}$ such that it will not interfere with either the channel estimation nor the detection phase. This clearly can be done without transmission of any side information from the transmitter to the receiver, therefore, does not further impact the spectral efficiency. Note that the number of superimposed subcarriers for $\mathbf{C}_2^i$ can be arbitrary chosen. For example, if there are $N_f$ subcarriers, $\mathbf{C}_2$ is only added to the equally spaced $\frac{N_f}{\Delta_f}$ subcarriers, where $\Delta_f$ is the an arbitrarily chosen subcarrier spacing. For the subcarriers which contain $\mathbf{C}_2$, $L > N_t$, else $L \geq N_t$. Thus, $K$ and $L$ can be different for each subcarrier as long as the sum of $K$ and $L$ for each subcarrier is identical. This is done so that it minimizes the impact on the spectral efficiency.

To account for uncertainty in the spatial correlation matrix, $\mathbf{C}_1$ can be designed using the technique proposed in [56] in order to improve MSE performance over that of [55]. The design of $\mathbf{C}_2$ requires the minimization of [53]

$$PAPR_{MIMO} = \max_{\ell=1,2...N_t} \frac{\max |s_\ell(t)|^2}{E(|s_\ell(t)|^2)}, \tag{4.11}$$

where $s_\ell(t)$ denotes the signal in time domain at the $\ell^{th}$ antenna. From (4.4), the output from the IFFT becomes

$$\mathbf{S} = \mathbf{F}^H(\mathbf{C}_1 + \mathbf{C}_2 + \mathbf{D}),$$

where

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_{N_f} & & & \\ & \mathbf{F}_{N_f} & & \\ & & \ddots & \\ & & & \mathbf{F}_{N_f} \end{bmatrix} \in \mathbb{C}^{N_f N_t \times N_f N_t}.$$

In other words, $\mathbf{F}$ is a block diagonal matrix composed of $N_t$ number of $N_f$-point DFT matrices $\mathbf{F}_{N_f}$. From (4.8), $\mathbf{C}_2$ can be expressed as

$$\mathbf{C}_2 = \mathbf{BM}, \tag{4.12}$$

where

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}_{1,1}^T & & & & \\ & \mathbf{b}_{2,1}^T & & & \\ & & \ddots & & \\ & & & \mathbf{b}_{N_f,1}^T & \\ \mathbf{b}_{1,2}^T & & & & \\ & \mathbf{b}_{2,2}^T & & & \\ & & \ddots & & \\ & & & \mathbf{b}_{N_f,2}^T & \\ \vdots & & & \vdots & \\ \mathbf{b}_{1,N_t}^T & & & & \\ & \mathbf{b}_{2,N_t}^T & & & \\ & & \ddots & & \\ & & & \mathbf{b}_{N_f,N_t}^T & \end{bmatrix} \in \mathbb{C}^{N_f N_t \times N_f (L-N_t)}$$

and $\mathbf{M} = \begin{bmatrix} \mathbf{M}_1^T & \mathbf{M}_2^T & \cdots & \mathbf{M}_{N_f}^T \end{bmatrix}^T$. $\mathbf{C}_2$ is then parameterized by $\mathbf{B}$, which is composed of

all the $\mathbf{b}_{i,l}$. Using the vectorization property $vec\,(\mathbf{AEB}) = (\mathbf{B}^T \otimes \mathbf{A})vec(\mathbf{E})$ and (4.12), the transmitted block matrix $\mathbf{S}$ can then be written

$$\mathbf{s} = vec\left(\mathbf{F}^H\left(\mathbf{C}_1 + \mathbf{BM} + \mathbf{D}\right)\right)$$

The sequence in $\mathbf{C}_2$ can then be designed by solving

$$\begin{aligned} &\min_{\|\mathbf{C}_2\|_F \le P_2} \|\mathbf{s}\|_\infty \\ =\ &\min_{\|\mathbf{C}_2\|_F \le P_2} \left\|\left(\mathbf{I}\otimes\mathbf{F}^H\right) vec\left(\mathbf{C}_1 + \mathbf{C}_2 + \mathbf{D}\right)\right\|_\infty \\ =\ &\min_{\|\mathbf{B}\|_F \le P_2} \left\|\left(\mathbf{I}\otimes\mathbf{F}^H\right) vec\left(\mathbf{C}_1 + \mathbf{BM} + \mathbf{D}\right)\right\|_\infty \end{aligned} \tag{4.13}$$

where $P_2$ is the power allocated to $\mathbf{C}_2$. It is known that $\mathbf{C}_2$ can be stacked using $\mathbf{C}_2^i$, so

$$\begin{aligned} \|\mathbf{C}_2\|_F^2 &= \sum_i tr(\mathbf{C}_2^i \mathbf{C}_2^{iH}), \\ &= \sum_i tr(\mathbf{B}_i \mathbf{M}_i \mathbf{M}_i^H \mathbf{B}_i^H), \\ &= \sum_i tr(\mathbf{B}_i \mathbf{B}_i^H), \\ &= tr(\mathbf{B}\mathbf{B}^H), \\ &= \|\mathbf{B}\|_F^2, \end{aligned}$$

which makes the equality in (4.13) hold and the designs of $\mathbf{C}_2$ and $\mathbf{B}$ be equivalent. $\mathbf{B}$ thus has less coefficients to be designed compared to what $\mathbf{C}_2$ has. Note that the total signal power is normalized, i.e. $\sigma_{\mathbf{D}}^2 + \sigma_{\mathbf{C}_1}^2 + \sigma_{\mathbf{C}_2}^2 = 1$, where $\sigma_{\mathbf{D}}^2$, $\sigma_{\mathbf{C}_1}^2$ and $\sigma_{\mathbf{C}_2}^2$ denote the variance of $\mathbf{D}$, $\mathbf{C}_1$ and $\mathbf{C}_2$, respectively. It is assumed that the information-bearing signal, and superimposed sequences all have zero-mean and are statistically independent from each other. Thus, $P_2 = \frac{\sigma_{\mathbf{C}_2}^2 N_t N_f (K+L)}{\Delta_f}$. Note that (4.13) can be recasted as a linear programming problem [58]

$$\begin{aligned} &\min && t \\ -t\mathbf{1}_{N_s\times 1} &\preceq \left(\mathbf{I}\otimes\mathbf{F}^H\right) vec\left(\mathbf{C}_1 + \mathbf{BM} + \mathbf{D}\right) \preceq t\mathbf{1}_{Ns\times 1} \\ &\|\mathbf{B}\|_F \le P_2 \end{aligned}$$

, where $N_s = N_f N_t(K + L)$ denotes the size of the vector $\mathbf{s}$. And the computational complexity is similar to that of [48] so that the computatonal complexity is $O(N_s \log(N_s))$.

### 4.2.3 Power Allocation

Thus far, no optimal method is found to optimally allocate the power between $\mathbf{D}$, $\mathbf{C}_1$, and $\mathbf{C}_2$. However, the suboptimal power allocation algorithm in [57] can be employed to allocate power between $\mathbf{D}$ and $\mathbf{C}_1 + \mathbf{C}_2$. From the results in [57], since the BER and MSE performance are not linearly related, the MSE performance can be traded off in order to increase the PAPR reduction performance without significantly impacting the BER. The proper amount of tradeoff may be computed using a greedy algorithm, which is currently under investigation.

## 4.3 Simulation Results

Complementary cumulative distribution function (CCDF) of the PAPR is used for performance comparison. A MIMO system with $N_t = N_r = 2$ is considered. $N_f = 128$ and $K = 10$ are used throughout all the simulations. $\sigma^2_{\mathbf{C}_1}$ is equal to 0.1 for the proposed SIPR scheme unless otherwise specified. QPSK and Alamouti STBC are used to modulate the information-bearing signals.

In Figure 4.3 shows the CCDF performance of the proposed SIPR scheme with different frequency spacings. The result is compared to a system without the addition of $\mathbf{C}_2$. The number of the redundant vectors $L$ equals to 4, and $\sigma^2_{\mathbf{C}_2} = 0.2$. When the value of $\Delta_f$ decreases, the number of superimposed subcarriers increases, which allows $\sigma_{\mathbf{C}_2}$ to have more design freedom in the frequency domain to lower the PAPR. When CCDF is equal to $10^{-3}$ and $\Delta_f = 4$, the proposed technique outperforms the one without $\mathbf{C}_2$ by 3 dB. When $\Delta_f$ equals to 8 and 16, the SIPR algorithm outperforms the system without $\mathbf{C}_2$ by

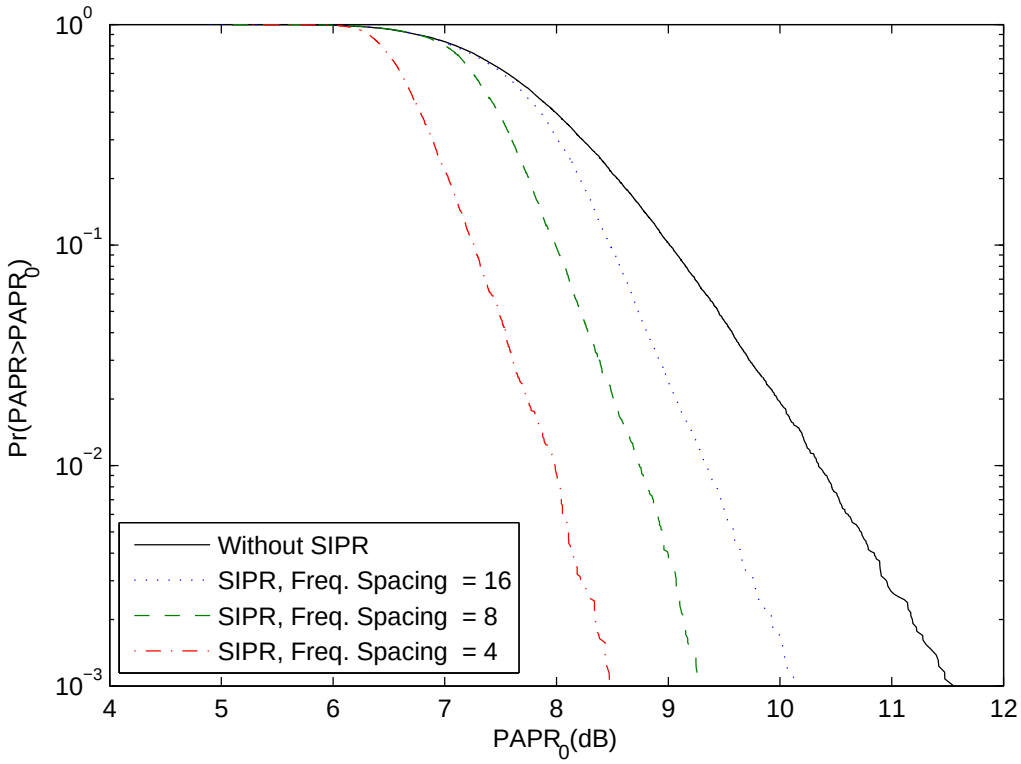



Figure 4.3: CCDF comparison of proposed SIPR method with different frequency spacings, and to a system without $\mathbf{C}_2$. $L = 4$ and $\sigma^2_{\mathbf{C}_2} = 0.2$.

about 2.2 dB and 1.2 dB, respectively.

Performance comparison of the SIPR method with different number of redundant vectors and to a system without $\mathbf{C}_2$ is shown in Figure 4.4. Frequency spacing is set to be 4 and $\sigma^2_{\mathbf{C}_2}$ is equal to 0.2. When $L = 3$ and CCDF $= 10^{-3}$, the PAPR performance of the SIPR method outperforms the system without $\mathbf{C}_2$ by 2.5 dB. Furthermore, as the number of redundant vectors increases, the PAPR performance improves. This is expected because as $L$ increases, the number of elements in $M_{A,i}$ also increases, thereby increasing the design freedom of $\mathbf{C}_2$. When $L = 5$, the performance gain of the SIPR scheme over that of the system without $\mathbf{C}_2$ becomes 3.2 dB when CCDF is equal to $10^{-3}$.

Next, the PAPR reduction performance of the SIPR scheme is examined by varying $\sigma^2_{\mathbf{C}_2}$. Figure 4.5 shows the performance comparison of the SIPR method with different

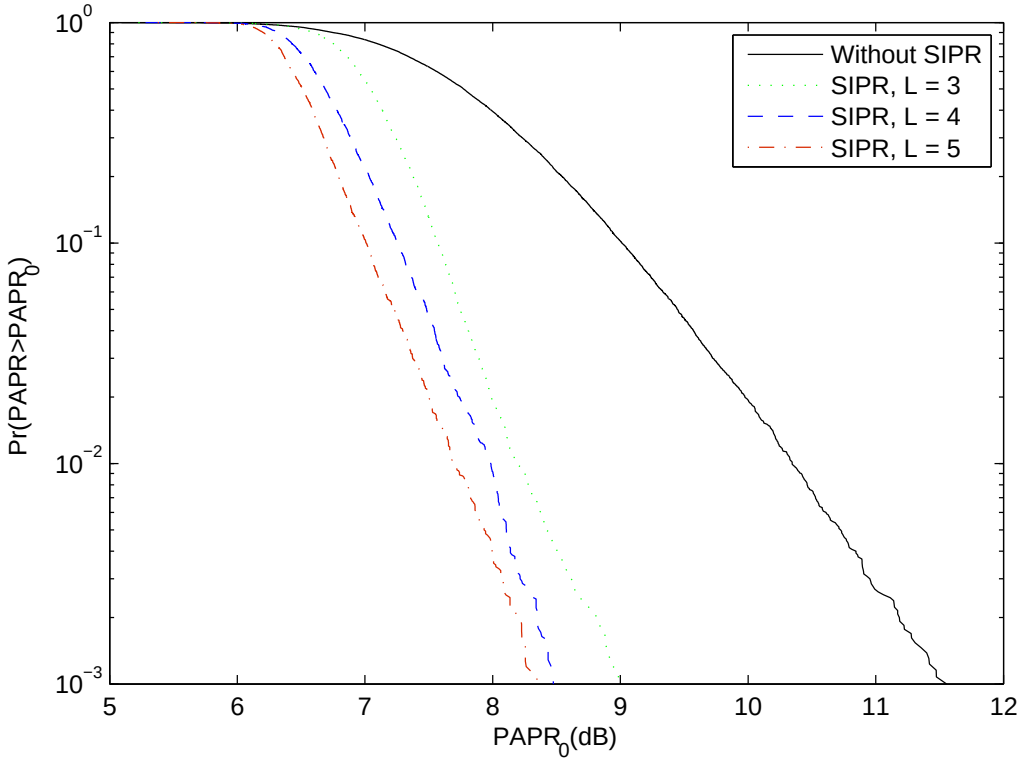



Figure 4.4: CCDF comparison of proposed SIPR method with different number of redundant vectors $L$. $\Delta_f = 4$ and $\sigma^2_{\mathbf{C}_2} = 0.2$.

values of $\sigma^2_{\mathbf{C}_2}$ and a system without $\mathbf{C}_2$. From the figure, it is clear that the performance of the SIPR method improves as $\sigma^2_{\mathbf{C}_2}$ increases. However, $\sigma^2_{\mathbf{C}_2}$ cannot be increased indefinitely as that will decrease the power allocated to $\mathbf{C}_1$ and $\mathbf{D}$, thus, degrading the channel estimation and data detection performance. As alluded in Section 4.2.3, the optimal (or suboptimal) amount of power to be allocated for PAPR reduction is currently under investigation.

Performance comparison of tone reservation and the proposed SIPR method is shown in Figure 4.6. To make the comparison fair in the sense of total number of blocks transmitted, $K$ and $L$ are set to be 12 and 2 for the TR method. The total power for $\frac{N_f}{\Delta_f}$ reserved subcarriers of all the $N_t$ transmit antenna is upbounded by the normalized power $\frac{N_f N_t}{\Delta_f}$ in $N_t$ symbols. When CCDF is equal to $10^{-3}$, the SIPR method outperforms the TR method

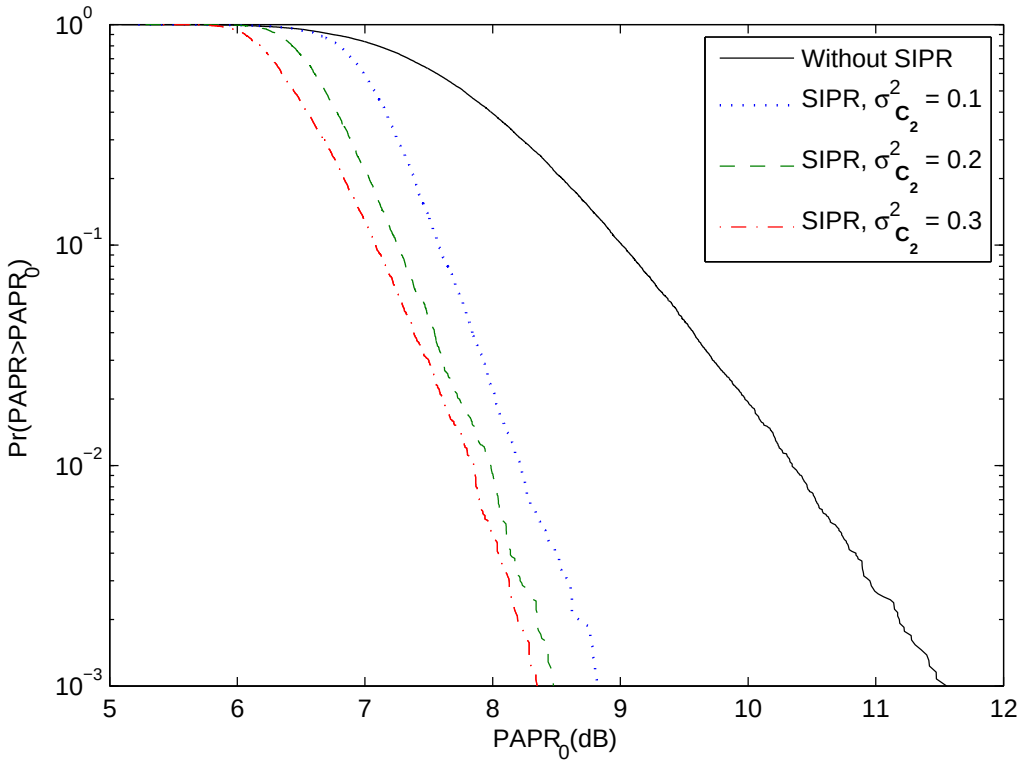



Figure 4.5: CCDF comparison of proposed SIPR method with different $\sigma^2_{\mathbf{C}_2}$. $L = 4$, and $\Delta_f = 4$.

in PAPR reduction by 0.7 dB. For TR method, since the subcarriers are reserved only for PAPR reduction, it can significantly decrease the throughput if large PAPR reduction is desired. In the SIPR method, if $\mathbf{C}_1$ is not required for channel estimation, then from (4.9), the necessary condition for the existence of $\mathbf{C}_2$ is for $L \geq 1$ (given that $N_t = 1$). Thus, the transmit redundancy can be as small as 1 symbol/subcarrier, which does not significantly reduce the spectral efficiency. To gain further insight into the impact of the transmit redundancy, define the transmit efficiency as

$$\rho \triangleq \frac{\kappa}{\tau}, \tag{4.14}$$

where $\kappa$ denotes the number of the transmitted symbols used for channel estimation and data detection, and $\tau$ denotes the total transmitted symbols. In Figure 4.6, $\rho$ is equal to 75% for the TR method because the number of reserved subcarriers used is 25% of the

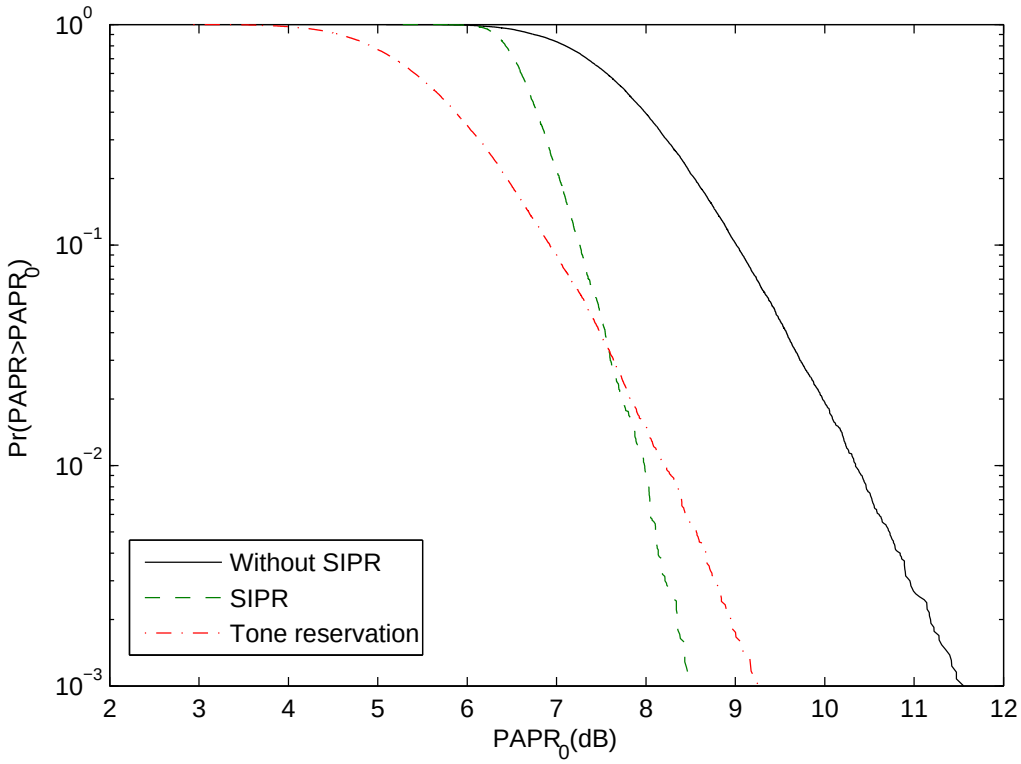



Figure 4.6: CCDF comparison of proposed method and tone reservation, equal power for redundancy , frequency spacing is set to be 4.

total number of subcarriers. However,

$$\rho = \frac{\frac{N_f}{\Delta_f}K_1 + \left(N_f - \frac{N_f}{\Delta_f}\right)K_2}{K_2 N_f} = \frac{(32)(12) + (96)(14)}{(14)(128)} = 96.4\%$$

for the proposed SIPR scheme, where $K_1 = 12$ is the number of symbols per subcarrier which are transmitted when $\mathbf{C}_2$ is employed on that particular subcarrier, and $K_2 = 14$ is the number of symbols per subcarrier which are transmitted when $\mathbf{C}_2$ is not used. These values match the values which have been used for the TR method in the current simulation. It can be seen that the proposed scheme outperforms the tone reservation not only in CCDF performance comparison but also the transmit efficiency.

# Chapter 5

# Conclusions and Future Works

## 5.1 Conclusions

A robust superimposed training sequence design algorithm for spatially correlated MIMO channel estimation has been proposed. The algorithm has shown to be robust against error in the spatial correlation estimate. When the robust training sequence is inserted into the MMSE estimator, a robust MMSE, or RoMMSE, estimator is derived. Also, the corrected sub-optimal power allocation, which is sub-optimal in maximizing the effective SNR, of the superimposed training sequence is also shown in this thesis. Simulation results have shown that the proposed RoMMSE estimator not only outperforms the optimal MMSE estimator in [55] when error in the spatial correlation exists, but it also outperforms other robust designs, such as RMMSE and LS-RMMSE [23]. Channel estimation error affects the data detection performance directly. Simulations have also shown that the proposed design outperforms other techniques mentioned in the sense of data detection.

A joint sequence design is proposed for robust channel estimation and PAPR reduction for MIMO-OFDM systems. The proposed SIPR method is able to accurately estimate MIMO-OFDM channels even with error in the spatial correlation estimate, and also significantly lower the PAPR without significantly impacting the spectral efficiency. When

channel estimation using superimposed training sequence is not required, the transmit redundancy has been shown to be as small as 1 symbol/subcarrier. The computational complexity is shown to be $O\left(N_f N_t \log\left(N_f N_t\right)\right)$. Simulation results show that the proposed scheme is able to outperform the tone reservation method in certain SNR range and has a significantly higher transmit efficiency.

## 5.2 Future Works

In the thesis, RoMMSE estimator only concerns the worst case mismatch of the spatial correlation in SU-MIMO systems. Since worst case mismatch does not occur frequently, probabilistic constraint of the mismatch can also be considered. Therefore, it is more close to the practical situation. RoMMSE estimator is proposed in the SU-MIMO system. In MU-MIMO, there might be some interference from other unwanted users. Therefore, mismatch of the interference correlation can also be considered in the training sequence design. Design of interference correlation estimator is also a possible direction, which can further enhance the performance of training sequence design.

Since power allocation for the superimposed sequence, which is for lowering the PAPR, is arbitrary designed, there is still room to find a optimal power allocation. The optimal power allocation could not only make the signal operating in the linear region of the power amplifier but also enhance the channel estimation and also data detection.

# Bibliography

[1] S.M. Alamouti, "A simple transmit diversity technique for wireless communications". *IEEE Journal on Selected Areas in Communications* ,vol.16(8), pp. 1451-1458 , Oct.1998.

[2] M. Dong and L. Tong, "Optimal Design and Placement of Pilot Symbols for Channel Estimation," *IEEE Trans. on Signal Processing*, vol. 50, pp. 3055V3069, December 2002.

[3] G.J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, Autumn 1996.

[4] G.J. Foschini and M.J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6(3) p. 311-335, Mar. 1998.

[5] I.E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. on Telecommunications*, vol. 10(6), pp. 585-595, Nov./Dec. 1999.

[6] C. Budianu and L. Tong, "Channel estimation for space-time orthogonal block codes," *IEEE Trans. on Signal Processing*, vol. 50(10), pp. 2515-2528, Oct. 2002.

[7] B. Hassibi and B.M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. on Information Theory*, vol. 49(4), pp. 2515-2528, Apr. 2003.

[8] B.L. Hughes, "Differential space-time modulation," *IEEE Trans. on Information Theory*, vol. 46(7), pp. 2567-2578, Nov. 2000.

[9] B.M. Hochwald and T.L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. on Information Theory*, vol. 46(2), pp. 543-564, Mar. 2000.

[10] V. Tarokh and H. Jafarkhani, "A differential detection scheme for transmit diversity," *IEEE Journal on Selected Areas in Communications*, vol. 18(7), pp. 1169-1174, Jul. 2000.

[11] H. Jafarkhani and V. Tarokh, "Multiple transmit antenna differential detection from generalized orthogonal designs," *IEEE Trans. on Information Theory*, vol. 47(6), pp. 2626-2631, Sep. 2001.

[12] M. Enescu, T. Roman and V. Koivunen, "Channel estimation and tracking in spatially correlated MIMO OFDM systems," *Proc. of the IEEE Workshop on Statistical Signal Processing*, pp. 347-250, Sep. 2003.

[13] M. Kiessling, J. Speidel and Y. Chen, "MIMO channel estimation in correlated fading environments," *Proc. of the IEEE Vehicular Technology Conference - Fall*, vol. 2, pp. 1187-1191, Oct. 2003.

[14] J.H. Kotecha and A.M. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. on Signal Processing*, vol. 52(2), pp. 546-557, Feb. 2004.

[15] N. Czink *et al.*, "Improved MMSE estimation of correlated MIMO channels using a structured correlation estimator," *Proc. of the IEEE Workshop on Signal Processing Advances in Wireless Communications*, pp. 595-599, Jun. 2005.

[16] B. Farhang-Boroujeny, "Pilot-based channel identification: proposal for semi-blind identification of communication channels," *IEEE Electronics Letters*, vol. 31(13), pp. 1044-1046, Jun. 1995.

[17] G.T. Zhou, M. Viberg and T. McKelvey, "Superimposed periodic pilots for blind channel estimation," *Proc. of the 35th Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 653-657, Nov. 2001.

[18] J.K. Tugnait and W. Luo, "On channel estimation using superimposed training and first-order statistics," *IEEE Communications Letters*, vol. 7(9), pp. 413-415, Sep. 2003.

[19] A.G. Orozco-Lugo, M.M. Lara and D.C. McLernon, "Channel estimation using implicit training," *IEEE Trans. on Signal Processing*, vol. 52(1), pp. 240-254, Jan. 2004.

[20] V. Nguyen *et al.*, "Optimal superimposed training design for spatially correlated fading MIMO channels," *IEEE Trans. on Wireless Communications*, vol. 7(8), pp. 3206-3217, Aug. 2008.

[21] D.H. Pham and J.H. Manton, "Orthogonal superimposed training on linear precoding: a new affine precoder design," *Proc. of the IEEE Workshop on Signal Processing Advances in Wireless Communications*, pp. 445-449, Jun. 2005.

[22] A. Vosoughi and A. Scaglione, "Everything you always wanted to know about training: guidelines derived using affine precoding framework and the CRB," *IEEE Trans. on Signal Processing*, vol. 54(3), pp. 940-954, Mar. 2006.

[23] M. Biguesh and A.B. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE Trans. on Signal Processing*, vol. 54(3), pp. 884-893, Mar. 2006.

[24] Y.-C. Chen and Y.T. Su, "MIMO channel estimation in spatially correlated environments," *Proc. of the IEEE Intl. Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 1, pp. 498-502, Sep. 2004.

[25] B. Vucetic and J. Yuan, "Space-Time Coding," *John Wiley and Sons*, 2003.

[26] N.N. Tran *et al.*, "Orthogonal affine precoding and decoding for channel estimation and source detection in MIMO Frequency-Selective Fading Channels," *IEEE Trans. on Signal Processing*, vol. 57(3), pp. 1151-1162, Mar. 2009.

[27] S. Ohno and G.B. Giannakis, "Optimal training and redundant precoding for block transmissions with application to wireless OFDM," *IEEE Trans. on Communications*, vol. 50(12), pp. 2113-2123, Dec. 2002.

[28] X. Ma, L. Yang and G.B. Giannakis, "Optimal training for MIMO frequency selective fading channels," *IEEE Trans. on Wireless Communications*, vol. 4(2), pp. 453-466, Mar. 2005.

[29] S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall, 1993.

[30] H. Miao, and M.J. Juntti, "Space-time channel estimation and performance analysis for wireless MIMO-OFDM systems with spatial correlation," *IEEE Trans. on Vehicular Technology*, vol. 54(6), pp. 2003-2016, Nov. 2005.

[31] D. Shiu *et al.*, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. on Communications*, vol. 48(3), pp. 502-513 , Mar. 2002.

[32] D.C. Youla and H. Webb, "Image restoration by the method of convex projections: part 1 - theory," *IEEE Trans. on Medical Imaging*, vol. MI-1(2), pp. 81-94, Oct. 1982.

[33] W. Weichselberger, M. Herdin, H. Özcelik, and E. Bonek, "A stochastic MIMO channel model with joint correlation of both link ends," *IEEE Trans. on Wireless Communications*, vol. 5, no. 1, pp. 90V99, 2006.

[34] F. Danilo-Lemoine *et al.*, "Power backoff reduction techniques for generalized multicarrier waveforms," *EURASIP Journal on Wireless Communications and Networking*, vol. 2008, pp. 1-13, Jan. 2008.

[35] A. Bahai, M. Singh, A. Goldsmith, and B. Saltzberg, "A new approach for evaluating clipping distortion in multicarrier systems," *IEEE Journal on Selected Areas in Communications*, vol. 20(5), pp. 1037-1046, Jun. 2002.

[36] X. Li and L.J. Cimini, Jr., "Effects of clipping and filtering on the performance of OFDM," *IEEE Communications Letters*, vol. 2(5), pp. 131-133, May 1998.

[37] T.A. Wilkinson and A.E. Jones, "Minimisation of the peak to mean envelope power ratio of multicarrier transmission schemes by block coding," *Proc. of the IEEE Vehicular Technology Conf.*, vol. 2, pp. 825-829, 1995.

[38] S.H. Müller and J.B. Huber, "OFDM with reduced peak-to-average power ratio by optimum combination of partial transmit sequences," *Electronics Letters*, vol. 33(5), pp. 368-369, Feb. 1997.

[39] A.D.S. Jayalath and C. Tellambura, "Adaptive PTS approach for reduction of peak-to-average power ratio of OFDM signal," *Electronics Letters*, vol. 36(14), pp. 1225-1228, Jul. 2000.

[40] S.H. Han and J.H. Lee, "PAPR reduction of OFDM signals using a reduced complexity PTS technique," *IEEE Signal Processing Letters*, vol. 11(11), pp. 887-890, Nov. 2004.

[41] R.W. Baüml, R.F.H. Fisher, and J.B. Huber, "Reducing the peak-to-average power ratio of multicarrier modulation by selected mapping," *Electronics Letters*, vol. 32(22), pp. 2056-2057, Oct. 1996.

[42] H. Breiling, S.H. Müller-Weinfurtner, and J.B. Huber, "SLM peak-power reduction without explicit side information," *IEEE Communications Letters*, vol. 5(6), pp. 239-241, Jun. 2001.

[43] P. Van Eetvelt, G. Wade, and M. Tomlinson, "Peak to average power reduction for OFDM schemes by selective scrambling," *Electronics Letters*, vol. 32(21), pp. 1963-1964, Oct. 1996.

[44] G.R. Hill, M. Faulkner, and J. Singh, "Reducing the peak-to-average power ratio in OFDM by cyclically shifting partial transmit sequences," *Electronics Letters*, vol. 36(6), pp. 560-561, Mar. 2000.

[45] A.D.S. Jayalath and C. Tellambura, "Reducing the peak-to-average power ratio of orthogonal frequency division multiplexing signal through bit or symbol interleaving," *Electronics Letters*, vol. 36(13), pp. 1161-1163, Jun. 2000.

[46] A.D.S. Jayalath and C. Tellambura, "SLM and PTS peak-power reduction of OFDM signals without side information," *IEEE Trans. on Wireless Communications*, vol. 4(5), pp. 2006-2013, Sep. 2005.

[47] J. Tellado and J. M. Cioffi, "Peak power reduction for multicarrier transmission, *Proc. of the IEEE GlobeCom Communications Theory MiniConf*, pp. 219-224, 1998.

[48] J. Tellado and J.M. Cioffi, "Efficient algorithms for reducing PAR in multicarrier systems," *Proc. of the IEEE Intl. Sym. on Information Theory*, pp. 191, Aug. 1998.

[49] B.S. Krongold and D.L. Jones, "PAR reduction in OFDM via active constellation extension," *IEEE Trans. on Broadcasting*, vol. 49(3), pp. 258-268, Sep. 2003.

[50] M. Tan, Z. Latinović, and Y. Bar-Ness, "STBC MIMO-OFDM peak-to-average power ratio reduction by cross-antenna rotation and inversion," *IEEE Communications Letters*, vol. 9(7), pp. 592-594, Jul. 2005.

[51] S.H. Han and J.H. Lee, "An overview of peak-to-average power ratio reduction techniques for multicarrier transmission, " *IEEE Wireless Communications*, vol. 12(2), pp. 56-65, Apr. 2005.

[52] T. Jiang and Y. Wu " An overview: peak-to-average power ratio reduction techniques for OFDM signals," *IEEE Trans. on Broadcasting*, vol. 54(2), pp.257-268, Jun. 2008.

[53] R.F.H. Fischer and M. Hoch "Peak-to-average power ratio reduction in MIMO OFDM," *IEEE Intl. Conf. on Communications*, Jun. 2007.

[54] D.H. Pham and J.H. Manton, "Orthogonal superimposed training on linear precoding: a new affine precoder design," *Proc. of the IEEE Workshop on Signal Processing Advances in Wireless Communications*, pp. 445-449, Jun. 2005.

[55] V. Nguyen *et al.*, "Optimal superimposed training design for spatially correlated fading MIMO channels," *IEEE Trans. on Wireless Communications*, vol. 7(8), pp. 3206-3217, Aug. 2008.

[56] C.-T. Chiang and C.C. Fung, "Robust training sequence design for spatially correlated MIMO channel estimation using affine precoder," *Proc. of the Intl. Conf. on Communications*, May 2010.

[57] C.-T. Chiang and C.C. Fung, "Robust training sequence design for spatially correlated MIMO channel estimation," *submitted to the IEEE Trans. on Vehicular Technology*, May 2010.

[58] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[59] M. Grant and S. Boyd, "cvx User's Guide", Feb. 2009.