

COMMUNICATIONS OVER NONCOHERENT DOUBLY
SELECTIVE CHANNELS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the
Graduate School of The Ohio State University

By

Arun Pachai Kannu, B.E., M.S.

* * * * *

The Ohio State University

2007

Dissertation Committee:

Prof. Philip Schniter, Adviser

Prof. Hesham El Gamal

Prof. Randolph L. Moses

Approved by

Adviser
Graduate Program in
Electrical Engineering

ABSTRACT

Wireless communication systems transferring broadband data in high mobility situations encounter fading channels which are both time *and* frequency selective. In the noncoherent scenario, the time varying impulse response of the doubly selective channel (DSC) is not available at both the transmitter and the receiver. In this dissertation, we consider the problem of communications over such noncoherent doubly selective channels. Our work has two main themes: to find the fundamental limits on the information rates for reliable communication across noncoherent DSC and to develop simple and efficient encoding/decoding techniques to achieve the promised information rates. Towards this end, we consider block transmissions over DSC and utilize complex-exponential (CE) basis expansion model (BEM) to characterize the channel variation within a block.

For noncoherent CE-BEM DSC, we characterize the pre-log factor of the constrained ergodic channel capacity in the high SNR regime, when the channel inputs are continuously distributed. Next, we consider the design of pilot aided transmissions (PAT) for CE-BEM DSC, which embeds known pilot (i.e., training) signals that the receiver uses to estimate the channel. For a given fixed pilot energy, we derive the necessary and sufficient conditions on the pilot/data pattern to attain minimum mean squared error (MMSE), uncover time-frequency duality of MMSE-PAT structures and obtain novel MMSE-PAT patterns. We obtain bounds on the ergodic achievable

rates of MMSE-PAT schemes and perform high signal to noise ratio (SNR) asymptotic analysis which suggests that, a multi-carrier MMSE-PAT achieves higher rates than a single-carrier MMSE-PAT when the channel's delay spread dominates its Doppler spread, and vice versa. We also establish that the pre-log factor of the ergodic rates of *all* the MMSE-PAT patterns are strictly less than that of the constrained channel capacity, for *strictly* doubly selective channels. We also design spectrally efficient PAT schemes whose asymptotic achievable rates have the same pre-log factor as that of the constrained channel capacity. Our results show that there are fundamental differences between singly selective and doubly selective channels and provide insights on how the DSC's delay spread and Doppler spread influence the constrained channel capacity and the PAT design.

We extend the MMSE-PAT design to multi-input multi-output (MIMO) CE-BEM DSC. We establish that the spectral efficiency of MIMO-MMSE-PAT does not necessarily increase even if the number of transmit and receive antennas are increased simultaneously. We also present the optimal number of active antennas which maximizes the spectral efficiency.

dedicated to my *Amma*, *Appa* and my sister *Agalya* ...

ACKNOWLEDGMENTS

First of all, I would like to thank my advisor, Prof. Phil Schniter, for his valuable advice, guidance and help throughout this work. I have benefited a lot from his critical feedback on my research and writing. I thank Prof. Hesham El Gamal for his valuable comments and his Wireless Comm class was a great motivating factor to do quality research. I also thank Profs. Randy Moses and Vadim Utkin for serving in my candidacy and dissertation committees and their constructive comments. I thank all my teachers from kindergarten to graduate school; their commitment to teaching has helped me acquire immense knowledge.

I thank my parents and sister, for their love and support throughout my life. I also thank my uncles, aunts and cousins; they gave me the motivation and inspiration to do well in studies. Time spent together with them has always been so much fun.

I thank all my friends at Ohio State for all the good times and making my stay in Columbus a memorable one. With the deadly combo - Gopala, Raghuna and Kicha - we had so many laughs despite “tough” times at *231*. I also thank my other roommates Rajesh, Vivek and Rao for sharing the burden of cooking and eating what we have cooked! Many thanks to Praveen, Sib, Aditi, Arul for all the enjoyable homework-solving, crossword-solving sessions, lunch break outings, friday night trips to Chipotle (sometimes Udipi, Movies). I thank Kambiz for his efforts to have friday evening IPS-tea-time during which we all had interesting, technical, non-technical,

not-necessarily-politically-correct discussions! I thank Iris, Adam, Sung-Jun, Young-Han, Sugumar, Lifeng, Mehmood for all the friendly conversations, fun-filled excursions to the conferences. I thank Sound and Vignesh for helping me with temp acco. I thank Subhojit, Sukirti for giving company with Tennis. Many thanks to Tate, Rahul, Satya, Anand for joining us in the crazy parties and wild trips to MadMex, HoundDogs. I thank all the newbies in the IPS for making me realize I am getting old, expediting my graduation! I thank Jeri for all her help ranging from conference registration to organizing IPS picnics.

Finally, I thank OSU football team for making all the Fall weekends exciting and entertaining! I will always be a Buckeye fan. **Go Bucks!**

This material is based upon work supported by the National Science Foundation under Grant No. 0237037.

VITA

September 10, 1979 Born - Virudhunagar, India

2001 B.E. Electronics and Communications
Anna University, India

2001 - 2002 Research Associate
AU-KBC Research Center
Anna University, India

Sep 2002 - present Graduate Student
The Ohio State University

PUBLICATIONS

Research Publications

A. P. Kannu and P. Schniter, "On the spectral efficiency of noncoherent doubly selective channels," *Proc. Allerton Conf. on Commun., Control and Computing*, Monticello IL, Sep. 2006.

A. P. Kannu and P. Schniter, "Minimum mean squared error pilot aided transmission for MIMO doubly selective channels," *Proc. Conf. Information Sciences and Systems*, Princeton, NJ, Mar. 2006.

A. P. Kannu and P. Schniter, "Capacity analysis of MMSE-Pilot schemes in doubly selective channels," *Proc. IEEE Workshop SPAWC, New York*, June 2005. (**Student Paper Award**).

A. P. Kannu and P. Schniter, "MSE-Optimal training for linear time-varying channels," *Proc. IEEE Conf. ICASSP*, Philadelphia, Mar. 2005.

A. P. Kannu and P. Schniter, "Reduced-complexity pilot-aided decision-directed tracking of doubly-selective channels," *Proc. Conf. Information Sciences and Systems*, Princeton, NJ, Mar. 2004.

S. Subramanian, V. P. Reuben, A. P. Kannu and T. Sivanadayan, "Performance of adaptive modulation techniques for Wireless Local Area Networks," *IETE Technical Review*, vol 21, pp 345-350, Sep-Oct. 2004.

A. P. Kannu, T. Sivanadayan and S. Subramanian, "Performance of block adaptive modulation schemes for OFDM based WLAN systems", *Proc. IEEE Personal and Wireless Commun.*, New Delhi, India, Dec. 2002.

A. P. Kannu, V. Rajendran, K. Sundaresan and S. Subramanian, "Multiple access schemes for OFDM based indoor wireless systems", *Proc. National Conf. on Commun.*, Mumbai, India, Jan. 2002.

FIELDS OF STUDY

Major Field: Electrical Engineering

Studies in:

Detection and Estimation theory	Prof. Philip Schniter
Communication and Information theory	Prof. Hesham El Gamal
Control theory	Prof. Vadim Utkin

TABLE OF CONTENTS

	Page
Abstract	ii
Dedication	iv
Acknowledgments	v
Vita	vii
List of Tables	xii
List of Figures	xiii
Chapters:	
1. Introduction	1
1.1 Doubly Selective Channels	2
1.2 Motivation	4
1.3 Contributions and Outline	6
2. System Model & Problem Statement	10
2.1 Doubly Selective Channel Model	10
2.1.1 Block Fading Model	14
2.2 Characterization of Asymptotic Channel Capacity	17
2.3 Pilot Aided Transmission Model	19
2.3.1 PAT Encoder	19
2.3.2 PAT Decoder	21
2.4 PAT Design	22

3.	Asymptotic Channel Capacity	26
3.1	Pre-log Factor	26
3.2	Performance of Gaussian Codes	30
4.	Minimum Mean Squared Error - PAT	34
4.1	MSE Lower Bound	34
4.2	MMSE-PAT Design	40
4.3	Examples of MMSE-PAT	43
4.4	Achievable-Rate Analysis of MMSE-PAT	47
4.5	Pilot/Data Power Allocation	50
4.6	Streaming MMSE-PAT	52
4.7	Numerical Results	57
5.	Spectrally Efficient PAT	63
5.1	Is any MMSE-PAT Spectrally Efficient?	63
5.2	Design of Spectrally Efficient PAT	65
5.3	Comparison of PAT schemes	69
5.3.1	Effective SNR	70
5.3.2	Intersection Point	76
6.	MMSE-PAT for MIMO systems	82
6.1	MIMO System Model	83
6.2	MIMO-MMSE-PAT Design	87
6.2.1	MIMO-MMSE-PAT Examples	92
6.3	Achievable Rates of MIMO-MMSE-PAT	94
6.4	Pilot/Data Power Allocation	99
6.5	Streaming MIMO-MMSE-PAT	100
6.6	Numerical Results	103
7.	Conclusions	107
7.1	Summary of Original Work	107
7.2	Possible Future Research	109

Appendices:

A.	Derivation of channel fading statistics	111
A.1	Discrete-Time Channel Statistics	111
A.2	Basis-Expansion Statistics	112
B.	Proofs of Asymptotic Channel Capacity results	113
B.1	Proof of Theorem 1	113
B.2	Proof of Lemma 1	115
C.	Proofs of MMSE-PAT results	117
C.1	Proof of Theorem 2	117
C.2	Proof of Lemma 2	119
C.3	Design of Optimal Pilot Sequence \mathbf{p}	120
C.4	Proof of Optimality of MMSE-PAT examples	121
C.4.1	TDKD	121
C.4.2	FDKD	122
C.4.3	Time domain Chirps	123
C.4.4	Frequency domain Chirps	124
C.5	Proof of Theorem 3	124
C.6	Proof of Theorem 4	126
D.	Proofs of Spectrally Efficient PAT results	129
D.1	Proof of Theorem 5	129
D.2	Proof of Theorem 6	135
E.	Proofs of MIMO-MMSE-PAT results	137
E.1	Proof of Theorem 7	137
E.2	Proof of Theorem 8	138
E.3	Proof of Theorem 9	139
E.4	Derivation of T_\star and R_\star	140
	Bibliography	142

LIST OF TABLES

Table	Page
1.1 List of abbreviations.	9

LIST OF FIGURES

Figure	Page
1.1 A digital communication system.	2
1.2 Mobility and multipath in wireless channel.	3
3.1 Effect of spreading index on the constrained channel capacity.	29
3.2 Upper bound on the mutual information for Gaussian codes.	31
3.3 Lower bound on the mutual information for Gaussian codes.	33
4.1 Geometric interpretation of MMSE-PAT design requirements.	39
4.2 (a) Structure of TDKD with $N_{\text{delay}} = 5$ and $N_{\text{Dopp}} = 3$. (b) Structure of FDKD with $N_{\text{delay}} = 3$ and $N_{\text{Dopp}} = 5$	45
4.3 Time and bandwidth occupation for several streaming MMSE-PAT systems designed for $N_{\text{delay}} = 3$ and $N_{\text{Dopp}} = 3$. Lightly shaded shows channel input and darkly shaded shows channel output. Note the use of a time-frequency concentrated modulation pulse $\psi(t)$	55
4.4 Bounds on ergodic achievable rates for $N_{\text{delay}} = N_{\text{Dopp}} = 3$	60
4.5 Bounds on ergodic achievable rates for $N_{\text{delay}} = 15$ and $N_{\text{Dopp}} = 3$	61
4.6 Bounds on ergodic achievable rates for $N_{\text{delay}} = 15$ and $N_{\text{Dopp}} = 3$ for (a) non-uniform and (b) uniform decay and Doppler profiles.	62
5.1 Structure of SE-PAT with $N_{\text{delay}} = 5$ and $N_{\text{Dopp}} = 3$	69
5.2 Lower bound on achievable rates for $N = 108$, $N_{\text{delay}} = 9$, $N_{\text{Dopp}} = 3$	71

5.3	Lower bound on achievable rates for $N = 108$, $N_{\text{delay}} = 12$, $N_{\text{Dopp}} = 3$.	72
5.4	Lower bound on achievable rates for $N = 108$, $N_{\text{delay}} = 18$, $N_{\text{Dopp}} = 3$.	73
5.5	Achievable rates of SE-PAT: (a) $N = 135$, $N_{\text{delay}} = 9$ and $N_{\text{Dopp}} = 3$, (b) $N = 135$, $N_{\text{delay}} = 15$ and $N_{\text{Dopp}} = 3$.	77
5.6	Intersection of achievable rates for $N = 126$, $N_{\text{delay}} = 14$, $N_{\text{Dopp}} = 3$.	80
5.7	Intersection of achievable rates for $N = 96$, $N_{\text{delay}} = 12$, $N_{\text{Dopp}} = 3$.	81
6.1	Structure of $T = 2$ MIMO-TDKD for DSC with $N_{\text{delay}} = 5$ and $N_{\text{Dopp}} = 3$.	95
6.2	Time and bandwidth occupation for several streaming $T = 2$ MIMO-MMSE-PAT systems designed for DSC with $N_{\text{delay}} = 3$ and $N_{\text{Dopp}} = 3$. Lightly shaded shows channel input and darkly shaded shows channel output.	102
6.3	Achievable rates for $T = 2$, $R = 2$ MIMO system over a $N_{\text{delay}} = 9$, $N_{\text{Dopp}} = 3$ DSC with uniform decay and Doppler profile.	104
6.4	Achievable rates for $T = 2$, $R = 2$ MIMO system over a $N_{\text{delay}} = 9$, $N_{\text{Dopp}} = 3$ DSC with exponential decay and Jakes Doppler profile.	105
6.5	Achievable rates of SCP-MIMO-FDKD with $T \in \{1, 2, 3\}$ transmit antennas and $R = 2$ receive antennas.	106
D.1	Elements of the set $\{e_{(k,m)}, k \in \mathcal{N}_{\text{delay}}, m \in \mathcal{N}_{\text{Dopp}}\}$ arranged in a grid	130
E.1	Pre-log factor with number of active transmit antennas t .	141

CHAPTER 1

INTRODUCTION

Transferring information reliably across a noisy channel is the fundamental problem in communication theory. Shannon's seminal paper [1] showed that (almost) error-free digital communication systems can be designed even across a noisy channel. Key components in a digital communication system [2] are illustrated in Fig. 1.1. At the transmitter, an encoder maps the information bits into a "suitable" signal which is transmitted across the communication medium. The communication medium, commonly referred as the channel, usually distorts the signal based on its propagation characteristics. The receiver observes the distorted version of the transmit signal and decodes the information bits. Because of the distortion introduced by channel, there is a possibility of decoding error, in which case, the decoded bits do not coincide with the information bits the transmitter intended to transfer. For reliable communication, one would like to keep the probability of decoding error very small. Two important questions in designing a communication system are:

1. What is the maximum rate of information that can be reliably communicated?
2. How to design the encoder and decoder so that the information can be reliably transferred?

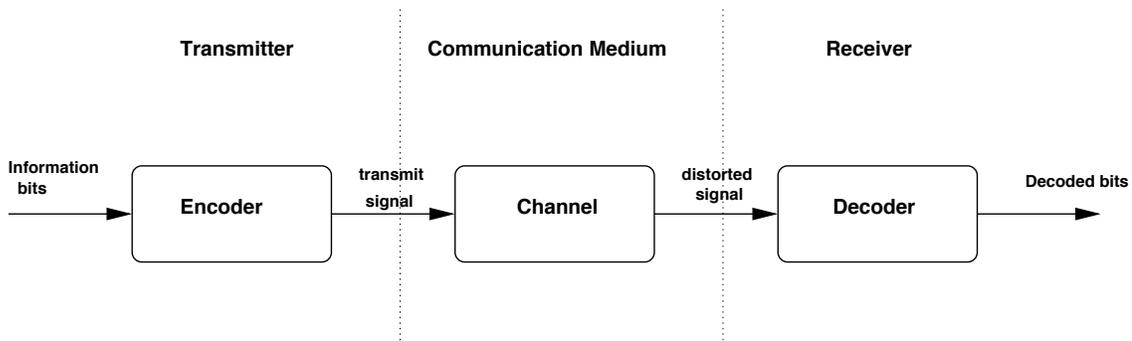


Figure 1.1: A digital communication system.

The answers to these questions depend heavily on the nature of the channel. Intuitively speaking, if the distortion introduced by the channel is very high, then the reliable information rate will be correspondingly less. Shannon introduced many new concepts and tools to study the communication problem [1] and found answers to the above questions for a class of *memoryless channels*. Since then, several researchers have analyzed various other channels using the tools from Shannon's original paper. The maximum information rate which can be reliably communicated with arbitrarily small decoding error across a channel is usually referred as its *capacity*.

1.1 Doubly Selective Channels

The rapidly evolving global information structure includes broadband wireless communication as a key component. In wireless communications (for e.g., cell-phones, satellite communication), the information bearing signal is transmitted across free space. Two principal factors which influence the distortion of the signal transmitted across the wireless (free space) medium are multipath fading and mobility [3],

as illustrated in Fig. 1.2¹. Multipath is the phenomenon in which the transmitted signal arrives at the receiver via multiple propagation paths at different delays. Because of the multipath fading, the multiple signals arriving at the receiver may add constructively or destructively at the receiver resulting in wide variations in the signal strength. Mobility is the phenomenon in which the relative positions of the different objects in the environment including the transmitter and the receiver change with time, causing the nature of channel distortion to vary with time. Future wireless links are expected to provide high data rate transfer of multimedia services in high mobility situations.

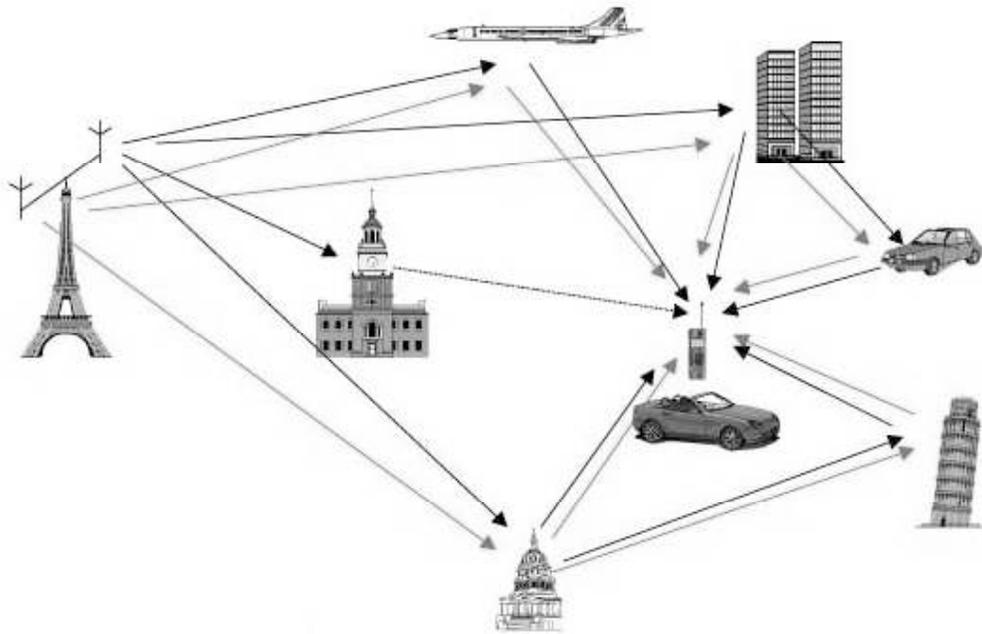


Figure 1.2: Mobility and multipath in wireless channel.

¹This figure is courtesy of Prof. Y. Liu.

In digital communication systems, for most of the channels, the discrete information bearing symbols are modulated (multiplied, in a sense) with a continuous pulse shape and transmitted across the channel [2]. In most cases, the pulse shapes are localized in time and frequency so that transmission of each discrete symbol consumes a small tile in the time-frequency plane. Across wireless channels, when the multipath delay spread is comparable to the symbol duration, the signal encounters severe intersymbol interference and the channel's response becomes frequency-selective. The rate of variation of the channel response across time, due to mobility in the environment, is characterized by the Doppler spread. If the channel response varies significantly in the signaling duration, it becomes time-selective. Channels whose response are both time and frequency selective are commonly referred as doubly-selective channels.

Doubly selective channels are typically modeled as linear time varying filters. The received sequence $y[n]$ (after match filtering and sampling) is related to the transmit sequence $x[n]$ (in the complex baseband) by

$$y[n] = \sum_l h[n; l]x[n - l] + v[n], \quad (1.1)$$

where $h[n; l]$ denotes the channel's impulse response at time instant n and $v[n]$ is additive noise term. The impulse response coefficients $h[n; l]$ are typically modeled as samples of a random process.

1.2 Motivation

When the channel's fading coefficients $h[n; l]$ are known apriori at the receiver, it is referred as coherent scenario. But, in most of the practical scenarios, the channel fading coefficients are not known apriori at the receiver. This situation is commonly referred as noncoherent case. In wireless communications, several coherent coding

and decoding techniques have been developed to combat fading for the case when the receiver and/or transmitter know these fading coefficients [3, 4]. In the noncoherent case, it is common to employ pilot aided transmission (PAT), whereby the transmitter embeds known pilot (i.e., training) signals which the receiver can use to estimate the channel. The estimated channel coefficients can then be used along with coherent decoding techniques. Cavers [5] authored one of the first analytical studies of PAT. Since then, there has been a growing interest in PAT design. [6] provides comprehensive overview of PAT design problems addressed in the literature. The *achievable rates* of the PAT schemes are in general less than the channel capacity, since the power and bandwidth spent in the transmission of training signals compromises the power and bandwidth available for transmitting the information bearing symbols. Achievable rates of a scheme usually refers to the rates which can be reliably communicated if one employs that particular encoding/decoding scheme. Using information-theoretic tools, some authors have worked on quantifying the achievable rates of PAT schemes for certain fading models, for e.g., [7].

In this dissertation, we consider the problem of communications over noncoherent doubly selective channels (DSC). One motivation to study DSC is that future wireless systems transferring broadband data in high mobility situations encounter fading channels which are both time selective and frequency selective. There are not many studies in the literature on DSC although there exist many studies about the special cases of DSC such as flat fading, time *or* frequency selective channels. Another motivation to analyze DSC is that the results obtained can be easily applied to the special cases of DSC. First, we find the fundamental limits on the information rates for reliable communication across noncoherent DSC. Next, we develop simple and efficient

encoding/decoding techniques to achieve the promised information rates. Since pilot aided transmissions are simple and widely used in the noncoherent scenarios, we focus on the PAT design for DSC.

1.3 Contributions and Outline

In Chapter 2, we present our single-input single-output (SISO) block transmission model and derive complex-exponential (CE) basis expansion model (BEM) to characterize the channel variation within a block, which is reasonably accurate and tractable in the design and analysis of communication systems. We also present the statement of problems addressed in the dissertation.

In Chapter 3, we characterize the pre-log factor of the constrained ergodic channel capacity of CE-BEM DSC in the high signal to noise ratio (SNR) regime, when the channel inputs are continuously distributed. We also present numerical results on the rates achieved by i.i.d. Gaussian codes.

In Chapter 4, we design PAT schemes for CE-BEM DSC based on mean squared error (MSE) criterion. Deriving the necessary and sufficient conditions on the pilot/data pattern to attain minimal MSE (MMSE) for a given pilot energy, we uncover time-frequency duality of MMSE-PAT structures and obtain novel MMSE-PAT patterns. We obtain bounds on the ergodic achievable rates of MMSE-PAT schemes and perform high-SNR asymptotic analysis which suggests that, the channel's spreading parameters should be taken into account when choosing among MMSE-PAT schemes. Specifically, we establish that a multi-carrier MMSE-PAT achieves higher rates than a single-carrier MMSE-PAT when the channel's delay spread dominates its Doppler spread, and vice versa.

In Chapter 5 we consider the design of PAT schemes whose asymptotic achievable rates have the same pre-log factor as that of constrained channel capacity. Referring these PAT schemes as spectrally efficient (SE) PAT, we present their design conditions and a novel pattern. Further analyzing the structure of MMSE-PAT schemes, we also establish that they are spectrally inefficient over strictly doubly selective channels. We also present numerical comparison of the rates achieved by SE-PAT and MMSE-PAT schemes.

In Chapter 6, we extend the MMSE-PAT design to multi-input multi-output (MIMO) CE-BEM DSC. We establish that the spectral efficiency of MIMO-MMSE-PAT does not necessarily increase even if the number of transmit and receive antennas are increased simultaneously. We also present the optimal number of active antennas which maximizes the spectral efficiency.

In Chapter 7, we conclude by summarizing our original work and indicating future research possibilities.

In order to enhance the flow of the dissertation, detailed mathematical derivations and proofs are deferred to appendices.

Notation

Matrices (column vectors) are denoted by upper-case (lower-case) bold-face letters. Hermitian is denoted by $(\cdot)^H$, transpose by $(\cdot)^T$, and conjugate by $(\cdot)^*$. The determinant and Frobenius norm are denoted by $\det(\cdot)$ and $\|\cdot\|_F$, respectively. The expectation, trace, delta, Kronecker product, modulo- N and integer ceiling operations are denoted by $E\{\cdot\}$, $\text{tr}\{\cdot\}$, $\delta\{\cdot\}$, \otimes , $\langle \cdot \rangle_N$ and $\lceil \cdot \rceil$, respectively. The null space and column space of a matrix are denoted by $\text{null}(\cdot)$ and $\text{col}(\cdot)$, respectively, and the

dimension of a vector space is denoted by $\dim(\cdot)$. The operation $[\cdot]_{n,m}$ extracts the $(n, m)^{th}$ element of a matrix, where the row/column indices n, m begin with 0, while $\text{diag}(\cdot)$ denotes a diagonal matrix constructed from the vector-valued argument. The $N \times N$ unitary discrete Fourier transform (DFT) and identity matrices are denoted by \mathbf{F}_N and \mathbf{I}_N , respectively, and appropriately dimensioned identity and all-zero matrices are denoted by \mathbf{I} and $\mathbf{0}$, respectively. The union, intersection, and set-minus operators are denoted by \cup , \cap , and \setminus , respectively, while the empty set is denoted by \emptyset . $I(\cdot)$ and $h(\cdot)$ denote mutual information and differential entropy respectively. Finally, the sets of integers, reals, positive reals, and complex numbers are denoted by \mathbb{Z} , \mathbb{R} , \mathbb{R}^+ , and \mathbb{C} , respectively.

AWGN	Additive White Gaussian Noise
BEM	Basis Expansion Model
CE	Complex Exponential
CP	Cyclic Prefix
CSI	Channel State Information
CWGN	Circular White Gaussian Noise
DFT	Discrete Fourier Transform
DOF	Degrees of Freedom
DSC	Doubly Selective Channels
FDKD	Frequency Domain Kronecker Delta
i.i.d.	independent and identically distributed
LMMSE	Linear Minimum Mean Squared Error
MIMO	Multi Input Multi Output
MMSE	Minimum Mean Squared Error
MSE	Mean Squared Error
OFDM	Orthogonal Frequency Division Multiplexing
PAT	Pilot Aided Transmissions
SCM	Single Carrier Modulation
SCP	Streaming Cyclic Prefix
SE	Spectrally Efficient
SISO	Single Input Single Output
SNR	Signal to Noise Ratio
SSC	Singly Selective Channels
SZP	Streaming Zero Prefix
TDKD	Time Domain Kronecker Delta
ZP	Zero Prefix

Table 1.1: List of abbreviations.

CHAPTER 2

SYSTEM MODEL & PROBLEM STATEMENT

In this chapter, we present our doubly selective channel model and PAT system model for single antenna systems, which provide the mathematical setup for our analysis in Chapters 3-5. We also present detailed discussions on the problems addressed in our work and their relations to the previous work in the literature.

2.1 Doubly Selective Channel Model

Starting with a continuous time fading channel model, using pulse shaping techniques we obtain a discrete time baseband equivalent doubly selective channel model. Using basis expansion model, we approximate the time variation of the channel and analyze its accuracy.

Consider a baseband-equivalent wireless multipath channel which can be modeled as a linear time-variant distortion plus an additive noise [8]:

$$y(t) = \int h(t; \tau)x(t - \tau)d\tau + v(t). \quad (2.1)$$

Say that, over the small time period $\mathcal{T}_{\text{small}}$, the path lengths vary by at most a few wavelengths, so that path gains and delays can be assumed to remain constant. Over this small time period, it is reasonable to model $h(t; \tau)$ a stationary random process

for which

$$\mathbb{E}\{h(t; \tau)h^*(t - t_o; \tau - \tau_o)\} = R_{\text{lag;delay}}(t_o; \tau)\delta(\tau_o). \quad (2.2)$$

Property (2.2) is commonly known as wide-sense stationary uncorrelated scattering [2]. If we define

$$R_{\text{Dopp;delay}}(f; \tau) = \int R_{\text{lag;delay}}(t; \tau)e^{-j2\pi ft} dt, \quad (2.3)$$

then the practical assumptions of finite path-length differences and finite rates of path-length variation imply that

$$R_{\text{Dopp;delay}}(f; \tau) \approx 0 \quad \text{for} \quad \begin{cases} f \notin [-\mathcal{B}_{\text{Dopp}}, \mathcal{B}_{\text{Dopp}}] \\ \tau \notin [0, \mathcal{T}_{\text{delay}}] \end{cases} \quad (2.4)$$

In other words, the channel has a causal delay spread of $\mathcal{T}_{\text{delay}}$ seconds and a single-sided Doppler spread of $\mathcal{B}_{\text{Dopp}}$ Hz. From $R_{\text{Dopp;delay}}(f; \tau)$, one can define the delay power profile $P_{\text{delay}}(\tau) = \int R_{\text{Dopp;delay}}(f; \tau)df$ and the Doppler power profile $P_{\text{Dopp}}(f) = \int R_{\text{Dopp;delay}}(f; \tau)d\tau$.

We consider a communication system in which the information bearing discrete symbols are parsed into transmission blocks of length N . For each transmission block, the baseband-equivalent modulation of the discrete symbols of that block $\{x[n]\}_{n=0}^{N-1}$, is accomplished by $x(t) = \sum_n x[n]\psi(t - n\mathcal{T}_s)$, where \mathcal{T}_s is the sampling interval in seconds and where $\psi(t)$ is a unit-energy pulse that is (approximately) zero-valued outside the time interval $[-\frac{\mathcal{T}_s}{2}, \frac{\mathcal{T}_s}{2})$ seconds and the frequency interval $[-\frac{1}{2\mathcal{T}_s}, \frac{1}{2\mathcal{T}_s})$ Hz. We assume that the block transmission duration $\mathcal{T}_{\text{burst}} \approx N\mathcal{T}_s$ is less than the small scale fading duration $\mathcal{T}_{\text{small}}$. Baseband-equivalent demodulation is accomplished by generating the received samples $y[n] = \int y(t)\psi^*(t - n\mathcal{T}_s)dt$ for $n \in \mathbb{Z}$.

The relationship between the discrete-time transmitted and received sequences is

$$y[n] = \sum_l h[n; l]x[n - l] + v[n], \quad (2.5)$$

where it is straightforward to show that $v[n] = \int v(t)\psi^*(t - n\mathcal{T}_s)dt$ and

$$h[n; l] = \int \int \psi^*(t)h(t + n\mathcal{T}_s; \tau + l\mathcal{T}_s)\psi(t - \tau)dtd\tau. \quad (2.6)$$

The time support of $\psi(t)$ implies that $h[n; l]$ is a locally-averaged version of $h(t; \tau)$ around the point $(t; \tau) = (n\mathcal{T}_s; l\mathcal{T}_s)$. Furthermore, it is shown in Appendix A.1 that, when $R_{\text{lag;delay}}(t_o; \tau)$ varies slowly with respect to t_o and τ variations on the order of \mathcal{T}_s seconds, then

$$\mathbb{E}\{h[n; l]h^*[n - p; l - q]\} \approx R_{\text{lag;delay}}(p\mathcal{T}_s; l\mathcal{T}_s) \delta[q]C_\psi, \quad (2.7)$$

where $C_\psi = \int |\int \psi(t)\psi^*(t - \tau)dt|^2 d\tau$. In this case, the finite delay-spread of $R(t_o, \tau)$ implies that $h[n; l] \approx 0$ for $l \notin [0, \dots, N_{\text{delay}} - 1]$, where $N_{\text{delay}} = \mathcal{T}_{\text{delay}}/\mathcal{T}_s$. If the discrete delay spread of the channel $N_{\text{delay}} > 1$, then the channel outputs $\{y[n]\}_0^{N_{\text{delay}}-1}$ depend on the past transmitted symbols $\{x[n]\}_{-N_{\text{delay}}+1}^{-1}$. To keep the block fading model more general, we allow these symbols to be arbitrary including the possibility of being zero.

Now, we obtain the basis expansion approximation. For one block fading duration, the channel response is characterized by the channel coefficients $h[n; l]$, $n \in \{0, \dots, N - 1\}$, $l \in \{0, \dots, N_{\text{delay}} - 1\}$. These channel coefficients can be parameterized using the basis expansion

$$h[n; l] = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \lambda[k; l] e^{j\frac{2\pi}{N}nk}, \quad (2.8)$$

where it is shown in Appendix A.2 that

$$\begin{aligned} & \mathbb{E}\{\lambda[k; l]\lambda^*[k-p; l-q]\} \\ & \approx C_\psi \delta[q]\delta[p] \int R_{\text{Dopp};\text{delay}}(f' + \frac{k}{N\mathcal{T}_s}; l\mathcal{T}_s) \left(\frac{\sin(\pi f' \mathcal{T}_s N)}{\sin(\pi f' \mathcal{T}_s)} \right)^2 df' \end{aligned} \quad (2.9)$$

$$\stackrel{N \rightarrow \infty}{\equiv} C_\psi \delta[q]\delta[p] R_{\text{Dopp};\text{delay}}(\frac{k}{N\mathcal{T}_s}; l\mathcal{T}_s). \quad (2.10)$$

The approximation in (2.9) follows from the use of (2.7). Note that, under the mild assumptions of (2.9), the basis coefficients are uncorrelated over both indices. Note furthermore that, under the additional assumption of large N , $\lambda[k; l] \approx 0$ for $k \notin \{-\lfloor \mathcal{B}_{\text{Dopp}} \mathcal{T}_s N \rfloor, \dots, \lfloor \mathcal{B}_{\text{Dopp}} \mathcal{T}_s N \rfloor\}$ as a consequence of the limited support of $R_{\text{Dopp};\text{delay}}(f; \cdot)$. In this case, it suffices to parameterize the channel (over the block interval) as

$$h[n; l] \approx \frac{1}{\sqrt{N}} \sum_{k=-\lfloor N_{\text{Dopp}}-1 \rfloor/2}^{\lfloor N_{\text{Dopp}}-1 \rfloor/2} \lambda[k; l] e^{j \frac{2\pi}{N} nk}. \quad (2.11)$$

where $N_{\text{Dopp}} = 2\lfloor \mathcal{B}_{\text{Dopp}} \mathcal{T}_s N \rfloor + 1$ and where $\{\lambda[k; l]\}$ are uncorrelated coefficients whose variances are assigned by uniformly sampling $R_{\text{Dopp};\text{delay}}(f; \tau)$. We refer N_{Dopp} as the discrete Doppler spread of the channel and the “normalized” product

$$\gamma = \frac{N_{\text{delay}} N_{\text{Dopp}}}{N}, \quad (2.12)$$

as the channel’s spreading index. Our derivation of the complex-exponential basis expansion model (CE-BEM) resembles that of [9] for time-selective channels. But with mild restrictions on $R_{\text{lag};\text{delay}}(t_o; \tau)$, we are able to establish that BEM coefficients are uncorrelated.

The CE-BEM (2.11) has been widely used to model time-varying communication channels (e.g., [9–13]) and can be interpreted as an N_{Dopp} -term truncated Fourier-series approximation of each of the N_{delay} coefficient trajectories $\{h[0; l], \dots, h[N -$

$1; l\}}_{l=0}^{N_{\text{delay}}-1}$. The application of truncated Fourier series can be motivated by the bandlimited nature of coefficient trajectories that results from finite mobile velocities. Specifically, path lengths which vary by at most v_{max} meters per second imply a maximum single-sided Doppler spread of $\mathcal{B}_{\text{Dopp}} = 2v_{\text{max}}f_c/c$ Hz, where f_c denotes the carrier frequency [10] and where c denotes the speed of light. Since the use of $N_{\text{Dopp}} = 2\lceil\mathcal{B}_{\text{Dopp}}\mathcal{T}_s N\rceil + 1$ terms in the Fourier series yields a reasonably accurate approximation to each trajectory, we assume this value of N_{Dopp} throughout. We allow CE-BEM coefficients with possibly unequal variances in order to model arbitrary delay profiles and Doppler spectra.

We emphasize that, like all models, this CE-BEM *approximates* the DSC; it does not yield a “perfect” description. Alternative BEMs have been proposed that, in some cases, yield “better” approximations (e.g., the polynomial [14], Slepian [15], and oversampled-CE [16] BEMs). We adopt the CE-BEM (2.11) because it is reasonably accurate and yields a tractable analysis with insightful results.

We restrict our focus to channels for which $\gamma < 1$, known as “underspread” channels [17], so that $N_{\text{delay}}N_{\text{Dopp}}$, the number of independent channel parameters per block, is less than the block length N . The underspread assumption is standard for radio-frequency channels [3] and for most underwater acoustic channels [18].

2.1.1 Block Fading Model

Now, we summarize our discrete doubly selective block fading model. The channel output is given by

$$y[n] = \sqrt{\rho} \sum_{l=0}^{N_{\text{delay}}-1} h[n; l]x[n-l] + v[n], n \in \{0, \dots, N-1\}, \quad (2.13)$$

Now, we define the diagonal matrix

$$\mathbf{X}_i = \begin{pmatrix} x[i] & & & \\ & x[i+1] & & \\ & & \ddots & \\ & & & x[i+N-1] \end{pmatrix} \quad (2.16)$$

and

$$\mathbf{X} = (\mathbf{X}_0 \ \mathbf{X}_{-1} \ \cdots \ \mathbf{X}_{-N_{\text{delay}}+1}). \quad (2.17)$$

Collecting the channel coefficients of the block $h[n; l]$, $n \in \{0, \dots, N-1\}$, $l \in \{0, \dots, N_{\text{delay}}-1\}$ in a vector $\mathbf{h} = [\mathbf{h}_0^\top, \dots, \mathbf{h}_{N_{\text{delay}}-1}^\top]^\top$, with $\mathbf{h}_l = [h[0; l], \dots, h[N-1; l]]^\top$, (2.15) can also be written as

$$\mathbf{y} = \sqrt{\rho} \mathbf{X} \mathbf{h} + \mathbf{v}. \quad (2.18)$$

Sometimes, in our analysis, we find it convenient to use (2.18). Also, defining $\boldsymbol{\lambda}_l = [\lambda[-\frac{N_{\text{Dopp}}-1}{2}; l], \dots, \lambda[\frac{N_{\text{Dopp}}-1}{2}; l]]^\top$, $\boldsymbol{\lambda} = [\boldsymbol{\lambda}_0^\top, \dots, \boldsymbol{\lambda}_{N_{\text{delay}}-1}^\top]^\top$, we have

$$\mathbf{h} = \mathbf{U} \boldsymbol{\lambda}, \quad (2.19)$$

where $\mathbf{U} = \mathbf{I}_{N_{\text{delay}}} \otimes \bar{\mathbf{F}}$ with $\bar{\mathbf{F}} \in \mathbb{C}^{N \times N_{\text{Dopp}}}$ given element-wise as

$$[\bar{\mathbf{F}}]_{n,m} = \frac{1}{\sqrt{N}} e^{j \frac{2\pi}{N} n(m - (N_{\text{Dopp}}-1)/2)}. \quad (2.20)$$

From our assumptions on BEM coefficients, we have $\mathbf{R}_\lambda = E\{\boldsymbol{\lambda} \boldsymbol{\lambda}^H\}$ is diagonal and positive definite. We consider an energy preserving channel which satisfies $\text{tr}\{\mathbf{R}_\lambda\} = N$. Clearly, (2.19) gives Karhunen-Loeve expansion for the channel vector \mathbf{h} . Combining (2.18) and (2.19), we have another representation of input-output equation

$$\mathbf{y} = \sqrt{\rho} \mathbf{X} \mathbf{U} \boldsymbol{\lambda} + \mathbf{v}, \quad (2.21)$$

which will also be frequently used in our analysis.

2.2 Characterization of Asymptotic Channel Capacity

Recently, many researchers have been working on problems related to characterization of the capacity of wireless multipath channels under the practical assumption that neither the transmitter nor the receiver has channel state information (CSI). The problem becomes especially interesting in the high SNR regime, where the impact of channel uncertainty on communication performance is most pronounced. The high-SNR capacity of the noncoherent MIMO Gaussian flat-fading channel was characterized by Marzetta and Hochwald [19] and Zheng and Tse [7] using the *block-fading* approximation, whereby the channel coefficients are assumed to remain constant over a block of N symbols and change independently from block to block. Later, Vikalo et al. [20] characterized the high-SNR capacity of the noncoherent Gaussian frequency-selective block-fading SISO channel under the assumption that the discrete block length N exceeds the discrete channel delay spread N_{delay} . Liang and Veeravalli [9] characterized the high-SNR capacity of SISO Gaussian time-selective block-fading channel, where, within the block, the channel coefficients vary according to a finite-term Fourier series with $N_{\text{Dopp}} \leq N$ expansion coefficients with full rank covariance matrix, but change independently from block to block. For these flat, frequency-selective, and time-selective noncoherent block-fading Gaussian channels, the aforementioned works have shown that the capacity $\mathcal{C}(\rho)$ obeys $\lim_{\rho \rightarrow \infty} \mathcal{C}(\rho)/\log(\rho) = \eta$, where the *prelog factor* η decreases with the degree of channel uncertainty. For example, $\eta = \frac{N-1}{N}$ in the SISO flat-fading case, $\eta = \frac{N-N_{\text{delay}}}{N}$ in the SISO frequency-selective case,² $\eta = \frac{N-N_{\text{Dopp}}}{N}$ in the SISO time-selective case.

²Assuming uncorrelated inter-symbol interference coefficients.

Many authors have studied the capacity of noncoherent fading channels [21–27] in the low-SNR regime. In most fading scenarios, it has been established that the noncoherent capacity obeys $\lim_{\rho \rightarrow 0} \mathcal{C}(\rho)/\rho = 1$. Interestingly, the above result holds for both additive white noise (AWGN) channel capacity and coherent fading channel capacity. So, there is no significant penalty in the noncoherent channel capacity in the low-SNR regime. However, the signal achieving noncoherent capacity is very “peaky” [22, 24–27]. In general, the behavior of the channel capacity and optimal signaling techniques in high-SNR regime and low-SNR regime are very different [3, 7, 9, 22, 23].

In our work, we consider the problem of characterizing the high-SNR capacity of underspread SISO doubly selective block fading channel described in Section 2.1.1, which uses a finite-length impulse response whose N_{delay} Gaussian coefficients vary according to a N_{Dopp} -term Fourier series within the block but change independently from block to block. We show that the prelog factor of the constrained asymptotic channel capacity with continuous inputs is $\frac{N - N_{\text{delay}} N_{\text{Dopp}}}{N}$. Detailed discussion on the results on the asymptotic channel capacity is given in Chapter 3.

Our work relies on the block-fading assumption, which can be justified in systems that employ block-interleaving or frequency-hopping. Other investigations have circumvented the block-fading assumption through the use of time-selective channel models whose coefficients vary from symbol to symbol in a stationary manner. For these stationary models, it is necessary to make a distinction between non-regular (e.g., bandlimited) fading processes and regular (e.g., Gauss-Markov) fading processes; while non-regular fading channels have been shown to behave similarly to time-selective block-fading channels, regular fading channels behave quite differently [28].

Though similar results are expected for doubly-selective stationary channel models, the details lie outside the scope of our work.

2.3 Pilot Aided Transmission Model

In this section, we give the details on the encoding and decoding techniques employed in PAT schemes analyzed in our work. Since one of the main advantages of PAT schemes is the utilization of communication techniques developed for coherent channels, we consider Gaussian codes and minimum distance decoding. The mathematical details are given below.

2.3.1 PAT Encoder

We describe the general form of the encoder of the PAT schemes studied in our work. Regarding the prefix portion, we restrict ourselves to cyclic-prefixed (CP) or zero-prefixed (ZP) block transmissions so that,

$$[x[-N_{\text{delay}} + 1], \dots, x[-1]] = \begin{cases} \mathbf{0} & \text{ZP} \\ [x[N - N_{\text{delay}} + 1], \dots, x[N - 1]] & \text{CP.} \end{cases} \quad (2.22)$$

Since for both the CP and ZP schemes, the vector $\hat{\mathbf{x}} = [x[0], \dots, x[N - 1]]^\top$ determines completely the entire transmit vector \mathbf{x} , we focus our attention on the structure of $\hat{\mathbf{x}}$. The input-vector $\hat{\mathbf{x}}$ is generated as

$$\hat{\mathbf{x}} = \mathbf{p} + \mathbf{d}, \quad (2.23)$$

where \mathbf{p} is a deterministic pilot vector and \mathbf{d} is zero-mean data. The data vector $\mathbf{d} = [d[0], \dots, d[N - 1]]^\top$ is constructed via linear precoding as

$$\mathbf{d} = \mathbf{B}\mathbf{s}, \quad (2.24)$$

where $\mathbf{s} = [s[0], \dots, s[N_s - 1]]^\top$ is the information-bearing symbol vector and $\mathbf{B} \in \mathbb{C}^{N \times N_s}$ is an arbitrary full-rank precoding matrix. Thus, our transmission strategy coincides with the general case of affine precoding [29], i.e.,

$$\hat{\mathbf{x}} = \mathbf{p} + \mathbf{B}\mathbf{s}. \quad (2.25)$$

We refer to $N_s = \text{rank}(\mathbf{B})$ as the “data dimension” of the PAT scheme. Our generic transmission model allows us to analyze PAT for many modulation schemes under a common framework. Notice that, for single carrier modulation (SCM) and orthogonal frequency division multiplexing (OFDM) systems, the columns of \mathbf{B} are chosen from those of \mathbf{I}_N and \mathbf{F}_N , respectively. Defining $E_p = \|\mathbf{p}\|^2$ and $E_d = \mathbb{E}\{\|\mathbf{d}\|^2\}$, we require that $E_p \geq 0$, $E_d > 0$, and the total energy $E_{\text{tot}} = E_p + E_d \leq N$. Denoting the CP or ZP mapper by $\mathbf{M} \in \mathbb{C}^{(N+N_{\text{delay}}-1) \times N}$ so that

$$\mathbf{x} = \mathbf{M}\hat{\mathbf{x}}, \quad (2.26)$$

the DSC output (2.15) for the PAT model can be written as

$$\mathbf{y} = \sqrt{\rho}\hat{\mathbf{H}}\{\mathbf{p} + \mathbf{B}\mathbf{s}\} + \mathbf{v}, \quad (2.27)$$

where we have

$$\hat{\mathbf{H}} = \mathbf{H}\mathbf{M}. \quad (2.28)$$

Constructing the transmit matrix \mathbf{X} in the same manner as in (2.18), the pilot and data components in \mathbf{X} are given by $\mathbf{P} = \mathbb{E}\{\mathbf{X}\}$ and $\mathbf{D} = \mathbf{X} - \mathbf{P}$. Using (2.19), it follows that

$$\mathbf{y} = \sqrt{\rho}\mathbf{P}\mathbf{U}\boldsymbol{\lambda} + \sqrt{\rho}\mathbf{D}\mathbf{U}\boldsymbol{\lambda} + \mathbf{v}. \quad (2.29)$$

Since $\boldsymbol{\lambda}$ captures all the degrees of randomness in \mathbf{h} , estimating $\boldsymbol{\lambda}$ is equivalent to estimating \mathbf{h} . We use the equation (2.29) in deriving the channel estimator and in the design of pilot/data patterns.

2.3.2 PAT Decoder

Our PAT decoder has two parts; a channel estimator part followed by a data detector part. The channel estimator obtains the linear MMSE (LMMSE) estimate of \mathbf{h} from the observation \mathbf{y} , using the knowledge of \mathbf{p} and the second order statistics of \mathbf{h} , \mathbf{v} and \mathbf{s} . Specifically, denoting $\mathbf{R}_y = \mathbb{E}\{\mathbf{y}\mathbf{y}^H\}$ and $\mathbf{R}_{y,h} = \mathbb{E}\{\mathbf{y}\mathbf{h}^H\}$, the channel estimate is obtained as

$$\hat{\mathbf{h}} = \mathbf{R}_{y,h}^H \mathbf{R}_y^{-1} \mathbf{y}. \quad (2.30)$$

Let $\hat{\mathbf{H}}$ denote the matrix constructed from $\hat{\mathbf{h}}$ in the same manner as \mathbf{H} . The data detector is a “mismatched” weighted minimum distance decoder which treats the channel estimate $\hat{\mathbf{H}}$ as though it is the true channel [30]. Specifically, $\hat{\mathbf{s}} = \operatorname{argmin}_{\mathbf{s}} \|\mathbf{Q}\{\mathbf{y} - \sqrt{\rho}\hat{\mathbf{H}}(\mathbf{p} + \mathbf{B}\mathbf{s})\}\|$. The choice of the weighting matrix \mathbf{Q} is arbitrary, at the moment. Our PAT decoder has a simple structure compared to a decoder which does channel estimation and data detection jointly. Also, note that the weighted minimum distance decoder is the maximum-likelihood decoder for coherent Gaussian noise channels [2].

To achieve arbitrarily small probability of decoding error over block fading DSC, the information has to be encoded into long codewords which span multiple blocks. Let $\underline{\mathcal{S}}$ denote the codebook with each of its codeword $\underline{\mathbf{s}}$ spanning K blocks. So, we have $\underline{\mathbf{s}} = [\mathbf{s}^{[0]\top}, \dots, \mathbf{s}^{[K-1]\top}]^\top$ where $\mathbf{s}^{[k]} \in \mathbb{C}^{N_s \times 1}$ corresponds to the k^{th} “segment” of the codeword $\underline{\mathbf{s}}$ transmitted during the k^{th} block. For coding over multiple blocks,

the data detector is modified as

$$\hat{\underline{\mathbf{s}}} = \operatorname{argmin}_{\underline{\mathbf{s}} \in \underline{\mathcal{S}}} \sum_{k=0}^{K-1} \|\mathbf{Q}\{\mathbf{y}^{[k]} - \sqrt{\rho} \hat{\mathbf{H}}^{[k]}(\mathbf{p} + \mathbf{B}\mathbf{s}^{[k]})\}\|^2, \quad (2.31)$$

where $\mathbf{y}^{[k]}$ and $\hat{\mathbf{H}}^{[k]}$ denotes the observations and the estimated channel matrix of the k^{th} block. We consider the codebook generated according to Gaussian distribution such that each codeword and its segments are generated independently with the covariance of the each segment being \mathbf{R}_s . Without loss of generality, we assume \mathbf{R}_s is full rank.³ Note that the Gaussian codes are capacity optimal for coherent Gaussian noise channels [33]. The information rate in bits per channel-use of the PAT scheme is given by $\mathcal{R} = \frac{1}{NK} \log |\underline{\mathcal{S}}|$, where $|\underline{\mathcal{S}}|$ denotes the size of the codebook. The rate \mathcal{R} is said to be *achievable* if the probability of detection error, i.e., $P_e = \Pr(\hat{\underline{\mathbf{s}}} \neq \underline{\mathbf{s}})$ is arbitrarily small.

2.4 PAT Design

The structure of the PAT scheme is governed mainly by the pilot vector \mathbf{p} and the data modulation matrix \mathbf{B} . We are interested in the joint design of (\mathbf{p}, \mathbf{B}) pair which meets some requirements. The criteria we will use are minimal channel estimation error variance and maximal pre-log factor of the asymptotic achievable rates.

Minimizing mean squared error is a popular PAT design criterion and many authors have studied it for different types of channels. For frequency selective channels, MMSE PAT schemes were obtained in [34–36] for OFDM systems. For time selective channels, MMSE-PAT schemes were obtained in [37] for SCM systems. For

³Noting that Gaussian codes are completely characterized by their covariance, it can be easily seen that, for a given \mathbf{B} and \mathbf{R}_s , we can always choose a full rank $\check{\mathbf{B}}$ and Gaussian vector $\check{\mathbf{s}}$ with positive definite covariance $\check{\mathbf{R}}_s$ so that the modulated data vectors $\mathbf{d} = \mathbf{B}\mathbf{s}$ and $\check{\mathbf{d}} = \check{\mathbf{B}}\check{\mathbf{s}}$ have the same covariance.

slowly time-varying frequency-selective channels, MMSE-PAT design for OFDM was discussed in [38,39] and [40] for single-antenna and multiple-antenna systems, respectively.

Since the bandwidth and power consumed by pilot symbols compromise the bandwidth and power available for information-bearing data symbols, PAT schemes may suffer a loss of information rate relative to noncoherent communication schemes that do not explicitly transmit pilots. To address this issue, several authors have studied the PAT design problem using information-theoretic tools (e.g., [13,20,41–44]). For example, [13,20,41–43] optimized the location, power, and number of pilot symbols by maximizing a lower bound on ergodic channel capacity, and, in doing so, uncovered connections between MSE minimization and achievable-rate maximization. In fact, for some channel models, MMSE-PAT and capacity-bound-maximizing PAT schemes coincide [13,20,41–43]. Furthermore, in the high signal-to-noise ratio regime, some of these MMSE-PAT schemes are first-order capacity-optimal, i.e., their achievable rates exhibit the same growth-rate (versus SNR) as the noncoherent channel’s capacity [7,9,20,41].

We consider the PAT design problem for DSC. For singly (time *or* frequency) selective channels (SSC), the channel matrix $\hat{\mathbf{H}}$ can be diagonalized with deterministic eigen vectors. For time selective channels, the standard basis vectors are the eigen vectors and for frequency selective channels, the complex exponential functions (columns of discrete Fourier transform matrix) are the eigen vectors. But, for DSC, there are no deterministic eigen vectors for the channel matrix $\hat{\mathbf{H}}$. This poses interesting challenges in the PAT design problem and shows the fundamental differences between SSC and DSC.

Motivated by the information-theoretic optimality of MMSE-PAT schemes for certain flat and frequency-selective channels [7, 20, 41], we start with the design and analysis of MMSE-PAT schemes for the CE-BEM DSC. While most studies on MMSE-PAT [34, 36, 37, 40] and achievable-rate-maximizing PAT [13, 20, 41–43, 45] assume a specific modulation scheme (e.g., OFDM or SCM), our study applies to the general class of cyclic-prefixed affine precoding [29] schemes. Affine precoding subsumes all schemes which transmit the sum of an energy-constrained pilot vector and a linearly precoded data vector, where the choice of the precoding matrix is arbitrary. For this setting, we derive necessary and sufficient conditions on the combination of pilot vector and precoding matrix that yield MMSE channel estimates. These conditions reveal the fundamental structure of all PAT schemes (i.e., irrespective of modulation format) that are MMSE for this DSC, paving the way for the design of novel MMSE-PAT schemes. Using these conditions, we then analyze the achievable rate of generic MMSE-PAT transmission and propose pilot/data power-allocation guidelines for it. The ergodic achievable rate is further analyzed in the high-SNR regime, and the spectral efficiency of systems which transmit a stream of MMSE-PAT blocks are characterized. This latter analysis provides insight into the roles of delay- and Doppler-spread on MMSE-PAT performance. In [13, 43], the authors investigated DSC PAT design for SCM with the goal of maximizing achievable rate. Due to connections between achievable-rate maximization and MSE-minimization, the PAT schemes proposed in [13, 43] turn out to be MMSE. However, as a consequence of our more general affine-precoding framework, we are able to show that the MMSE-PAT scheme obtained in [13] is but one of many MMSE-PAT schemes possible for the DSC. Furthermore, our achievable-rate analysis provides a means of comparing

among these different MMSE-PAT schemes and establishes that a particular multi-carrier MMSE-PAT scheme achieves higher rates than the single-carrier MMSE-PAT scheme from [13] when the DSC's discrete delay spread dominates its discrete Doppler spread. Numerical examples are also presented to illustrate the theoretical results.

Next, we consider the PAT design based on maximal pre-log factor of asymptotic achievable rates. We refer such PAT schemes as spectrally efficient PAT. While previous studies on the MMSE-PAT schemes established that they are spectrally efficient for the flat [7, 41], frequency-selective [20] and time-selective [9, 46–48] block fading channels, we establish that *all* the MMSE-PAT schemes of block-fading (strictly) doubly selective channels are spectrally *inefficient* [49]. We also provide guidelines for the design of spectrally efficient PAT schemes and obtain a novel spectrally efficient (non-MMSE) PAT scheme. We also present numerical and theoretical comparison of SE-PAT and MMSE-PAT schemes in the moderate SNR regime.

Our analysis of achievable rates of PAT schemes is limited to the high-SNR regime. As mentioned before, capacity optimal signaling techniques in the high-SNR regime are fundamentally different from those of low-SNR regime [3]. The work [50] provides insights on the PAT design for block-fading channels in the low-SNR regime.

CHAPTER 3

ASYMPTOTIC CHANNEL CAPACITY

3.1 Pre-log Factor

Recall that the input-output relation for a single block of doubly-selective block fading channel is

$$\mathbf{y} = \sqrt{\rho}\mathbf{H}\mathbf{x} + \mathbf{v}, \quad (3.1)$$

and the channel coefficients obey BEM (2.19). We study the ergodic capacity per-channel-use of the doubly selective block fading channel, which is expressed as [51]

$$\mathcal{C} = \sup_{\mathbf{x}: \mathbb{E}\{\|\mathbf{x}\|^2\} \leq N + N_{\text{delay}} - 1} \frac{1}{N} \mathbb{I}(\mathbf{y}; \mathbf{x}), \quad (3.2)$$

where $\mathbb{I}(\mathbf{y}; \mathbf{x})$ is the mutual information between random vectors \mathbf{y} and \mathbf{x} and the supremum is taken over all the input random vectors satisfying the power constraint. The mutual information in (3.2) is obtained by averaging over all the channel realizations and all the rates below the ergodic capacity can be achieved by coding over large number of block fading intervals [33].

We denote the pre-log factor in the high-SNR expression for the channel capacity by η . Precisely, we have

$$\eta = \lim_{\rho \rightarrow \infty} \frac{\mathcal{C}(\rho)}{\log \rho}. \quad (3.3)$$

For Rayleigh-fading SISO channels (flat, frequency-selective, or time-selective), in the coherent case, i.e., perfect receiver CSI, the pre-log factor η is unity. But in the noncoherent case, the pre-log factor is generally less than unity. The loss in pre-log factor has been shown to be proportional to the channel's discrete delay spread N_{delay} and discrete Doppler spread N_{Dopp} for frequency-selective and time-selective channels, respectively [9,20]. For the doubly selective block fading model, the ergodic coherent capacity, i.e., when the channel fading matrix \mathbf{H} is available at the receiver, is given by [33]

$$\mathcal{C}_{\text{coh}}(\rho) = \frac{1}{N} \sup_{\mathbf{R}_x \geq 0, \text{tr}\{\mathbf{R}_x\} \leq N + N_{\text{delay}} - 1} \mathbb{E}\{\log \det[\mathbf{I}_N + \rho \mathbf{H} \mathbf{R}_x \mathbf{H}^H]\} \quad (3.4)$$

where $\mathbf{R}_x = E\{\mathbf{x}\mathbf{x}^H\}$ and the expectation in (3.4) is taken over the random matrix \mathbf{H} . It easily follows that

$$\frac{1}{N} \mathbb{E}\{\log \det[\mathbf{I}_N + \rho \mathbf{H} \mathbf{H}^H]\} \leq \mathcal{C}_{\text{coh}} \leq \frac{1}{N} \mathbb{E}\{\log \det[\mathbf{I}_N + \rho(N + N_{\text{delay}} - 1) \mathbf{H} \mathbf{H}^H]\} \quad (3.5)$$

and since $\mathbf{H} \mathbf{H}^H$ is full rank (almost surely) with continuously distributed eigen values, we have

$$\lim_{\rho \rightarrow \infty} \frac{\mathcal{C}_{\text{coh}}(\rho)}{\log \rho} = 1.$$

So, the pre-log factor η of coherent DSC is unity. Now, we establish that, for the constrained capacity of noncoherent DSC with *continuous* random vector inputs, the pre-log factor is equal to $1 - \frac{N_{\text{delay}} N_{\text{Dopp}}}{N}$. So, the “loss” in pre-log factor is equal to the channel's spreading index $\gamma = N_{\text{Dopp}} N_{\text{delay}} / N$. To prove the result, first we obtain an upper bound on the pre-log factor of the mutual information between the input (with continuous distribution) and output of the block fading DSC in Theorem 1 and establish the achievability of the bound in Lemma 1.

Theorem 1 (Pre-log Factor). *For the block fading CE-BEM DSC, any sequence of continuous random input vectors $\{\mathbf{x}^\rho\}$ indexed by SNR ρ , satisfying the power constraint $\mathbb{E}\{\|\mathbf{x}\|^2\} \leq N + N_{\text{delay}} - 1$, and converging in distribution to a continuous random vector \mathbf{x}^∞ , yields*

$$\limsup_{\rho \rightarrow \infty} \frac{\frac{1}{N} \mathbb{I}(\mathbf{y}; \mathbf{x}^\rho)}{\log \rho} \leq \frac{N - N_{\text{Dopp}} N_{\text{delay}}}{N}. \quad (3.6)$$

Proof of Theorem 1 appears in Appendix B.1. The following lemma specifies a fixed input distribution which achieves equality in (3.6).

Lemma 1 (Achievability). *For the block fading CE-BEM DSC, i.i.d. inputs chosen from the zero-mean circular Gaussian distribution, i.e., $\mathbf{x} \sim \text{CN}(\mathbf{0}, \mathbf{I})$, yield*

$$\lim_{\rho \rightarrow \infty} \frac{\frac{1}{N} \mathbb{I}(\mathbf{y}; \mathbf{x})}{\log \rho} = \frac{N - N_{\text{Dopp}} N_{\text{delay}}}{N}. \quad (3.7)$$

See Appendix B.2 for proof of the above Lemma.

Since $\gamma \approx 2\mathcal{B}_{\text{Dopp}} \mathcal{T}_{\text{delay}}$, larger γ implies more time-frequency channel dispersion. Our result, which shows that channel dispersion limits the pre-log factor, is intuitively satisfying. For relatively small γ , the pre-log factor will be close to unity, i.e., that of the coherent case. A channel with small γ could be interpreted as one with few unknown parameters, and thus one which does not demand much training overhead. It has been established in [52] that, for overspread channels (i.e., $\gamma \geq 1$), the capacity grows only double logarithmically with SNR so that the pre-log factor is 0. The effect of the spreading index on the pre-log factor of the constrained capacity is illustrated in Fig. 3.1. The pre-log factor of the constrained capacity of block-fading DSC with continuous inputs coincides with the pre-log factor of the capacity its special cases: block fading - flat channels (i.e., $N_{\text{delay}} = 1, N_{\text{Dopp}} = 1$) [7, 19], time selective channels (i.e., $N_{\text{delay}} = 1$) [9], frequency selective channels (i.e., $N_{\text{Dopp}} = 1$) [20].

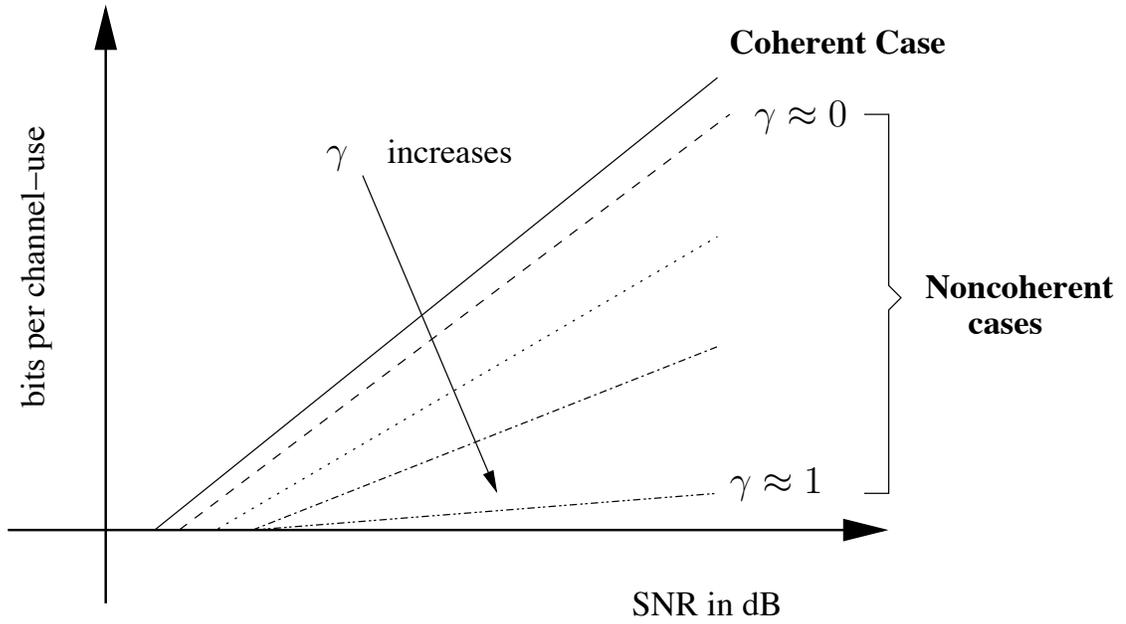


Figure 3.1: Effect of spreading index on the constrained channel capacity.

The loss in pre-log factor of noncoherent channel capacity can be intuitively seen as follows. The mutual information is expressed as $I(\mathbf{y}; \mathbf{x}) = h(\mathbf{y}) - h(\mathbf{y}|\mathbf{x})$, where $h(\cdot)$ denotes differential entropy. In both the coherent and noncoherent cases, the variance of elements of \mathbf{y} bounds the entropy of \mathbf{y} . In the coherent case, the uncertainty in \mathbf{y} given \mathbf{x} is only noise. So, $h(\mathbf{y}|\mathbf{x})$ is the entropy of noise. But, in the noncoherent case, the uncertainty in \mathbf{y} given \mathbf{x} is due to both channel and noise. Since the channel gets scaled by $\sqrt{\rho}\mathbf{X}$ (from (2.18)), $h(\mathbf{y}|\mathbf{x})$ scales with ρ and hence results in a loss. The evaluation of the mutual information for Gaussian codes in the following section gives the mathematical details behind the intuition.

3.2 Performance of Gaussian Codes

To illustrate the effect of spreading index, using i.i.d. Gaussian codes, we study the mutual information between the channel input and output of DSC. For zero-mean i.i.d. Gaussian codes, i.e., $\mathbf{x} \sim \text{CN}(\mathbf{0}, \mathbf{I})$, we first obtain an upper bound on the mutual information $I(\mathbf{y}; \mathbf{x})$. Since the channel is energy preserving, it easily follows that $E\{|y[i]|^2\} \leq \rho + 1$. So, applying independent Gaussian bound for differential entropy [51], we have

$$h(\mathbf{y}) \leq N \log(\rho + 1). \quad (3.8)$$

Given \mathbf{x} , the uncertainty in \mathbf{y} is due to both channel and noise. Recall that (2.21), the channel output can be written as

$$\mathbf{y} = \sqrt{\rho} \mathbf{X} \mathbf{U} \boldsymbol{\lambda} + \mathbf{v}. \quad (3.9)$$

Given \mathbf{x} , \mathbf{y} is Gaussian with covariance $\rho \mathbf{X} \mathbf{U} \mathbf{R}_\lambda (\mathbf{X} \mathbf{U})^H + \mathbf{I}$ and hence

$$h(\mathbf{y}|\mathbf{x}) = E\{\log \det[\mathbf{I} + \rho \mathbf{X} \mathbf{U} \mathbf{R}_\lambda (\mathbf{X} \mathbf{U})^H]\}. \quad (3.10)$$

Using (3.8) and (3.10), we numerically evaluate the upper bound on $I(\mathbf{y}; \mathbf{x})$ and plot the results in Fig. 3.2. The numerical results coincide with Theorem 1, illustrating the “degrading” effect of spreading factor. We see that the “slope” of the rate versus SNR plot decreases with the increase of spreading index.

For comparison, we plot the lower bound on the mutual information with i.i.d. Gaussian codes in Fig. 3.3. The lower bound is obtained as follows,

$$I(\mathbf{y}; \mathbf{x}) = I(\mathbf{y}; \mathbf{x}, \mathbf{H}) - I(\mathbf{y}; \mathbf{H}|\mathbf{x}) \quad (3.11)$$

$$\geq I(\mathbf{y}; \mathbf{x}|\mathbf{H}) - I(\mathbf{y}; \mathbf{H}|\mathbf{x}). \quad (3.12)$$

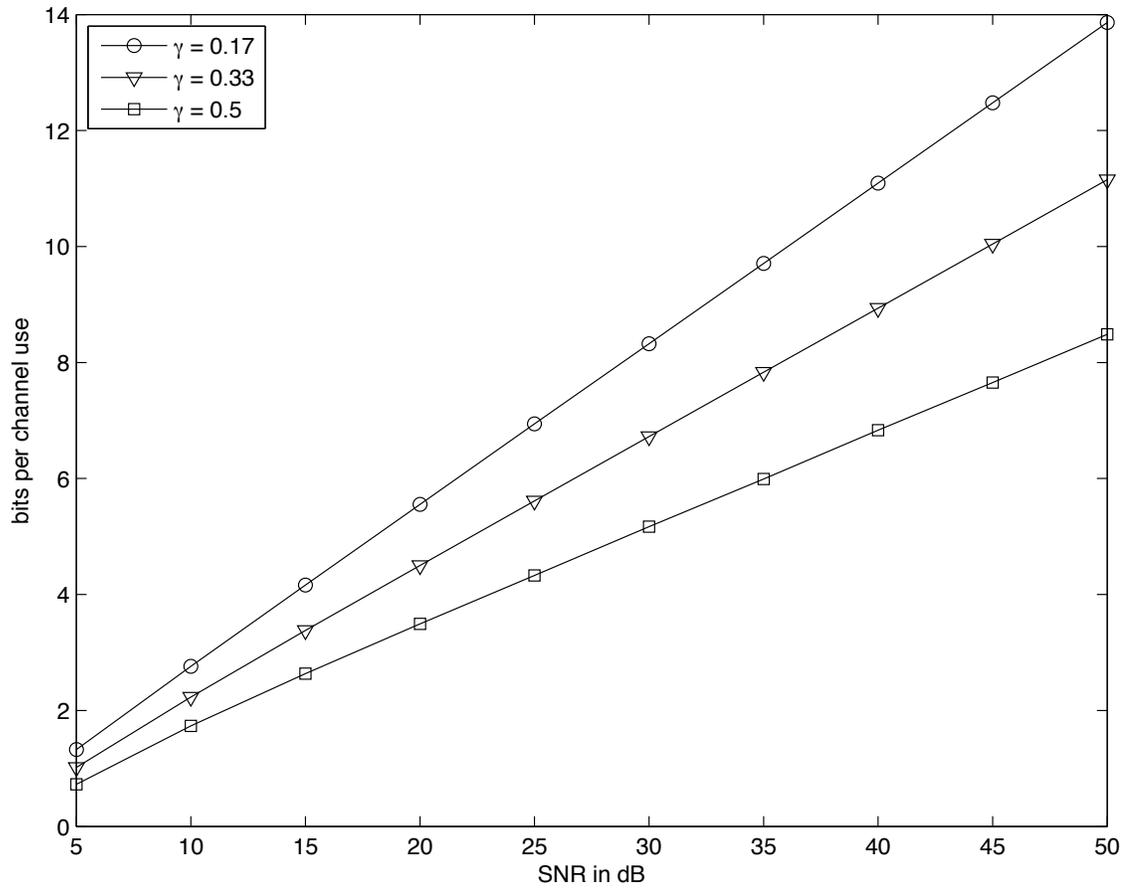


Figure 3.2: Upper bound on the mutual information for Gaussian codes.

Now, $I(\mathbf{y}; \mathbf{x}|\mathbf{H})$ corresponds the coherent case of known channel and with i.i.d. Gaussian codes,

$$I(\mathbf{y}; \mathbf{x}|\mathbf{H}) = E\{\log \det[\mathbf{I} + \rho\mathbf{H}\mathbf{H}^H]\}. \quad (3.13)$$

Similarly, we have

$$I(\mathbf{y}; \mathbf{H}|\mathbf{x}) = I(\mathbf{y}; \boldsymbol{\lambda}|\mathbf{X}) \quad (3.14)$$

$$= E\{\log \det[\mathbf{I} + \rho\mathbf{X}\mathbf{U}\mathbf{R}_\lambda(\mathbf{X}\mathbf{U})^H]\}. \quad (3.15)$$

Using (3.13) and (3.15) in (3.12), we evaluate the lower bound on $I(\mathbf{y}; \mathbf{x})$ and plot the results in Fig. 3.3, which shows similar effects of the spreading factor.

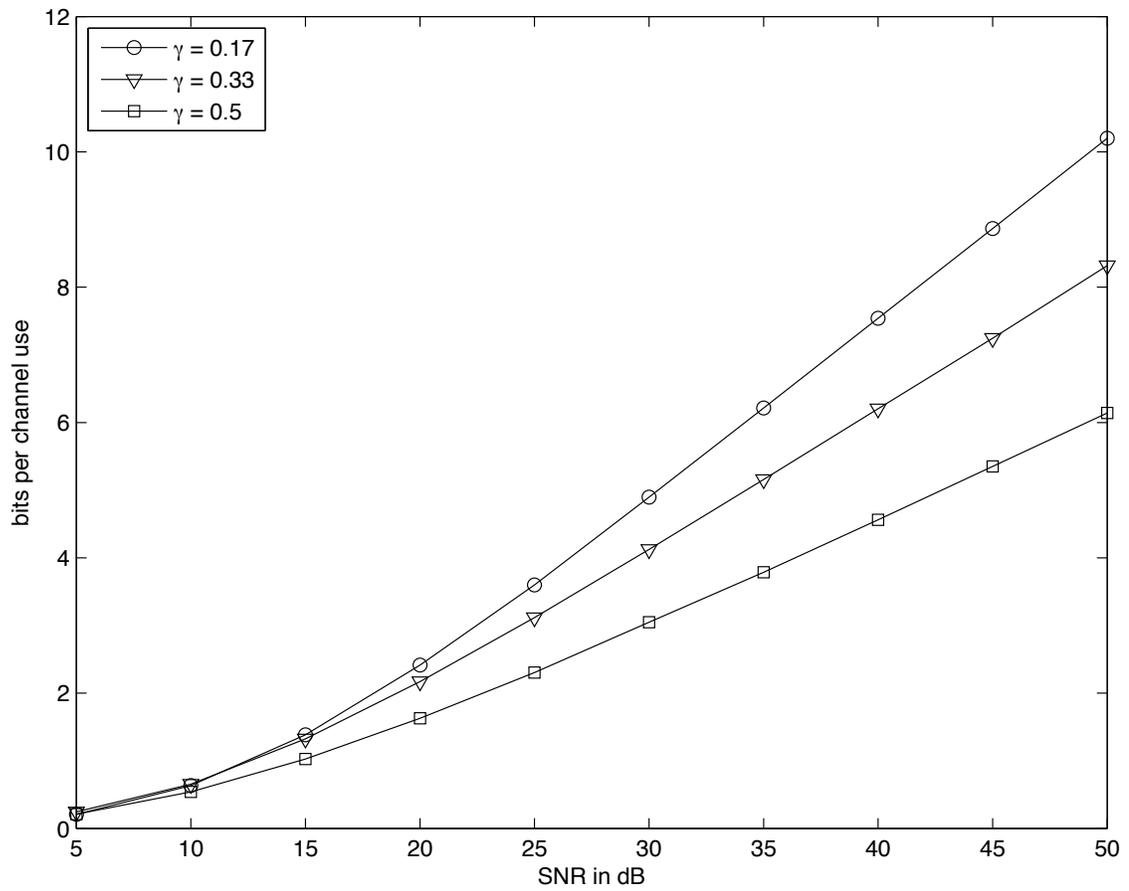


Figure 3.3: Lower bound on the mutual information for Gaussian codes.

CHAPTER 4

MINIMUM MEAN SQUARED ERROR - PAT

In this chapter, we consider cyclic prefixed PAT schemes over the single-antenna block fading DSC (Section 2.3), where a CE-BEM is used to characterize the channel variation over the block duration (Section 2.1.1). We outline a procedure for designing minimum mean squared error PAT schemes for this DSC and provide several novel MMSE-PAT examples. We derive the achievable rates of MMSE-PAT schemes and compare their rates in the high-SNR regime. Though we restricted our analysis to Rayleigh fading channels (Section 2.1.1), the MMSE-PAT design results in Sections 4.1-4.3 are valid for non-Gaussian fading scenarios as well. Throughout this chapter, we assume modulo- N indexing, i.e., $z[i] = z(\langle i \rangle_N)$.

4.1 MSE Lower Bound

In the MMSE-PAT design, we restrict ourselves to CP block transmissions so that $x[-i] = x[N - i]$, $i \in \{-1, \dots, -N_{\text{delay}} + 1\}$. Recall the DSC block transmission model (2.21),

$$\mathbf{y} = \sqrt{\rho} \mathbf{X} \mathbf{h} + \mathbf{v}, \quad (4.1)$$

where $\mathbf{X} \in \mathbb{C}^{N \times NN_{\text{delay}}}$ and $\mathbf{h} \in \mathbb{C}^{NN_{\text{delay}} \times 1}$. Recall that (Section 2.3.1), for PAT schemes, the transmitted signal is constructed as $x[i] = p[i] + d[i]$, where $\{p[i]\}$ is the

pilot sequence and $\{d[i]\}$ is the zero-mean data sequence. We also recall that (2.29) the output of CE-BEM DSC for a PAT scheme is written as

$$\mathbf{y} = \sqrt{\rho}\mathbf{P}\mathbf{U}\boldsymbol{\lambda} + \sqrt{\rho}\mathbf{D}\mathbf{U}\boldsymbol{\lambda} + \mathbf{v}. \quad (4.2)$$

As we will see, the channel estimate MSE is a direct function of the pilot energy

$$\|\mathbf{p}\|^2 = E_p. \quad (4.3)$$

The linear-MMSE estimate of \mathbf{h} given the knowledge of $\{\mathbf{y}, \mathbf{P}\}$ and the knowledge of the second-order statistics of $\{\mathbf{h}, \mathbf{D}, \mathbf{v}\}$ is [53]

$$\hat{\mathbf{h}} = \mathbf{R}_{y,h}^H \mathbf{R}_y^{-1} \mathbf{y}, \quad (4.4)$$

where $\mathbf{R}_{y,h} = E\{\mathbf{y}\mathbf{h}^H\}$ and $\mathbf{R}_y = E\{\mathbf{y}\mathbf{y}^H\}$. With our transmission and channel models, we have

$$\begin{aligned} \mathbf{R}_{y,h} &= \sqrt{\rho}\mathbf{P}\mathbf{U}\mathbf{R}_\lambda\mathbf{U}^H \\ \mathbf{R}_y &= \rho\mathbf{P}\mathbf{U}\mathbf{R}_\lambda\mathbf{U}^H\mathbf{P}^H + \rho E\{\mathbf{D}\mathbf{U}\mathbf{R}_\lambda\mathbf{U}^H\mathbf{D}^H\} + \mathbf{I}_N. \end{aligned}$$

Our channel estimator is pilot-aided, non-iterative and non-decision-aided. The channel estimation error $\tilde{\mathbf{h}} = \mathbf{h} - \hat{\mathbf{h}}$ has variance $\sigma_e^2 = E\{\|\tilde{\mathbf{h}}\|^2\}$ given by [53]

$$\sigma_e^2 = \text{tr}\{\mathbf{U}\mathbf{R}_\lambda\mathbf{U}^H - \mathbf{R}_{y,h}^H \mathbf{R}_y^{-1} \mathbf{R}_{y,h}\}. \quad (4.5)$$

We are interested in finding PAT schemes which minimize the MSE (4.5) of the LMMSE channel estimate (4.4) subject to a pilot power constraint E_p . Specifically, we are interested in finding the combination of E_p -constrained pilot vector \mathbf{p} and precoding matrix \mathbf{B} which minimize the estimation MSE σ_e^2 . We refer to such combinations of (\mathbf{p}, \mathbf{B}) as MMSE-PAT schemes. We shall see later that there are many

(\mathbf{p}, \mathbf{B}) combinations which lead to minimal MSE. We now proceed to the design of MMSE-PAT schemes.

Theorem 2 (MSE Lower Bound). *For N -block CP affine PAT over the CE-BEM DSC with the pilot-aided channel estimator (4.4), the channel estimate MSE (4.5) obeys*

$$\sigma_e^2 \geq \text{tr} \left\{ \left(\mathbf{R}_\lambda^{-1} + \frac{\rho E_p}{N} \mathbf{I}_{N_{\text{Dopp}} N_{\text{delay}}} \right)^{-1} \right\}, \quad (4.6)$$

where equality in (4.6) occurs if and only if the following two conditions hold:

1. *Pilot-Data Orthogonality:*

$$(\mathbf{P}\mathbf{U})^H \mathbf{D}\mathbf{U} = \mathbf{0}, \forall \mathbf{D}. \quad (4.7)$$

2. *Optimal Excitation:*

$$(\mathbf{P}\mathbf{U})^H \mathbf{P}\mathbf{U} = \frac{E_p}{N} \mathbf{I}_{N_{\text{Dopp}} N_{\text{delay}}}. \quad (4.8)$$

When (4.7)-(4.8) are met, $\mathbf{R}_{\tilde{\mathbf{h}}} = E\{\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H\}$ and $\mathbf{R}_{\hat{\mathbf{h}}} = E\{\hat{\mathbf{h}}\hat{\mathbf{h}}^H\}$ are given by

$$\mathbf{R}_{\tilde{\mathbf{h}}} = \mathbf{U} \left(\mathbf{R}_\lambda^{-1} + \frac{\rho E_p}{N} \mathbf{I}_{N_{\text{Dopp}} N_{\text{delay}}} \right)^{-1} \mathbf{U}^H, \quad (4.9)$$

$$\mathbf{R}_{\hat{\mathbf{h}}} = \mathbf{U} \left[\mathbf{R}_\lambda - \left(\mathbf{R}_\lambda^{-1} + \frac{\rho E_p}{N} \mathbf{I}_{N_{\text{Dopp}} N_{\text{delay}}} \right)^{-1} \right] \mathbf{U}^H. \quad (4.10)$$

Proof. See Appendix C.1. □

Condition (4.7) says that pilots and data should be multiplexed in a way that preserves orthogonality at the channel output, while condition (4.8) says that pilots should be constructed so that the channel modes are independently excited with equal

energy. Defining $b_q[i] = [\mathbf{B}]_{i,q}$, we rephrase the MSE optimality requirements (4.7)-(4.8) in terms of the pilot sequence $\{p[i]\}$ and modulation sequences $\{b_q[i]\}$, using the index sets

$$\begin{aligned}\mathcal{N}_{\text{delay}} &= \{-N_{\text{delay}} + 1, \dots, N_{\text{delay}} - 1\}, \\ \mathcal{N}_{\text{Dopp}} &= \{-N_{\text{Dopp}} + 1, \dots, N_{\text{Dopp}} - 1\}.\end{aligned}$$

Lemma 2 (Time Domain Conditions). *For N -block CP affine PAT over the CE-BEM DSC, the pair (4.11)-(4.12) form necessary and sufficient conditions for equality in (4.6).*

$$\sum_{i=0}^{N-1} p[i]p^*[i-k]e^{-j\frac{2\pi}{N}mi} = E_p\delta[k]\delta[m] \quad \forall k \in \mathcal{N}_{\text{delay}}, \forall m \in \mathcal{N}_{\text{Dopp}}, \quad (4.11)$$

$$\sum_{i=0}^{N-1} b_q[i]p^*[i-k]e^{-j\frac{2\pi}{N}mi} = 0 \quad \forall k \in \mathcal{N}_{\text{delay}}, \forall m \in \mathcal{N}_{\text{Dopp}}, \forall q \in \{0, \dots, N_s - 1\}. \quad (4.12)$$

Proof. See Appendix C.2. □

We are not aware of previous results which specify the relationship, between pilots and general forms of linearly modulated data, that is necessary and sufficient to minimize the MSE of LMMSE DSC estimates (4.4). Previous work on DSC PAT design [13, 43] was based on the maximization of a channel-capacity lower bound for the specific case of SCM. Requirements on pilot-data orthogonality to minimize the least squares (LS) estimation error variance were discussed, for frequency selective channels, in [35]. Our pilot-data orthogonality requirement (4.12) establishes that, for DSC MMSE-PAT, the data modulation basis must be orthogonal to certain time- and frequency-shifts of the pilot vector.

Geometric interpretation

The MMSE-PAT design requirements in Theorem 2 has intuitive geometric interpretation. The received observation \mathbf{y} is the sum of pilot component $\mathbf{PU}\boldsymbol{\lambda}$ and the data component $\mathbf{DU}\boldsymbol{\lambda}$ corrupted by additive white Gaussian noise. Notice that the pilot component lies in the column space of \mathbf{PU} while data component lies in the subspace $\bigoplus_{\mathbf{D}} \text{col}(\mathbf{DU})$, where \bigoplus denotes direct sum of vector spaces. Note that \mathbf{D} corresponds to random data which takes values from finite set of possibilities.

In general, the random undetected data (\mathbf{D}) causes interference in the estimator (4.4) and this interference will contribute to increase in the estimation error σ_e^2 given in (4.5). Our first MMSE-PAT design condition (4.7) requires that the pilot subspace and the data subspace have to be orthogonal as illustrated in Fig. 4.1. When the orthogonality condition is satisfied, the pilot component and the data component in \mathbf{y} can be “linearly separated” without changing the statistics of the noise. To see this, let the columns of $\mathbf{B}_p \in \mathbb{C}^{N \times K}$ and $\mathbf{B}_d \in \mathbb{C}^{N \times (N-K)}$ form an orthonormal basis for the column space of \mathbf{PU} and the left null space of \mathbf{PU} respectively. Now, the unitary rotation $[\mathbf{B}_p \mathbf{B}_d]^H \mathbf{y}$ separates the pilot and data components since $\mathbf{B}_d^H \mathbf{PU} = \mathbf{0}$ and $\mathbf{B}_p^H \mathbf{DU} = \mathbf{0}, \forall \mathbf{D}$. This corresponds to the projection of observation into the pilot and the data subspaces. Also, the unitary rotation does not result in any loss of pilot energy or the data energy and the noise statistics remain the same. Because of this linear separability property, the interference of the data component in the linear MMSE channel estimator is eliminated when the orthogonality requirement (4.7) is met.

Only the observation component in the pilot subspace $\text{col}(\mathbf{PU})$ is “useful” for the channel estimation since the data component $\mathbf{DU}\boldsymbol{\lambda}$ is uncorrelated with the channel

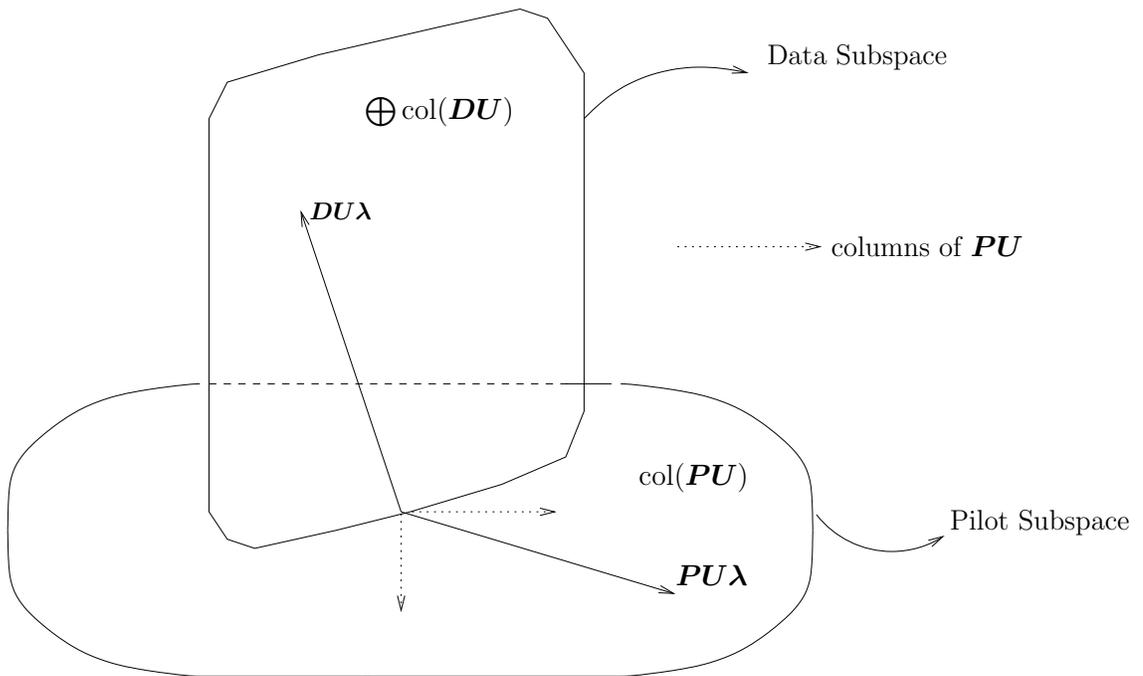


Figure 4.1: Geometric interpretation of MMSE-PAT design requirements.

$\boldsymbol{\lambda}$ since data \boldsymbol{D} is zero-mean. Now, the pilot subspace component is corrupted by the white noise component in that subspace $\boldsymbol{v}_p = \boldsymbol{B}_p^H \boldsymbol{v}$. The uncorrupted pilot observation $\boldsymbol{P}\boldsymbol{U}\boldsymbol{\lambda}$ corresponds to the sum of columns of $\boldsymbol{P}\boldsymbol{U}$ weighted by uncorrelated BEM coefficients. If say, two columns of $\boldsymbol{P}\boldsymbol{U}$ are closely aligned, it is difficult to differentiate the effect of weighting by the corresponding BEM coefficients. Intuitively, the best possible choice is that the columns of $\boldsymbol{P}\boldsymbol{U}$ are orthogonal (as illustrated in Fig. 4.1). Because of the columns of $\boldsymbol{P}\boldsymbol{U}$ have equal norm for any pilot vector \boldsymbol{p} due to the structural constraints imposed by the channel, we arrive at the optimal excitation requirement (4.8).

4.2 MMSE-PAT Design

Now, we outline a two-step DSC MMSE-PAT design procedure based on the necessary and sufficient conditions (4.11)-(4.12). In the first step, one obtains a \boldsymbol{p} which satisfies (4.11). A procedure for doing so is outlined in Appendix C.3. In the second step, an admissible \boldsymbol{p} is used to construct a $\boldsymbol{B} \in \mathbb{C}^{N \times N_s}$ which satisfies (4.12). This can be done as follows. Defining $\check{\boldsymbol{F}} \in \mathbb{C}^{2N_{\text{Dopp}}-1 \times N}$ element-wise as $[\check{\boldsymbol{F}}]_{n,m} = \frac{1}{\sqrt{N}} e^{-j\frac{2\pi}{N}(n-N_{\text{Dopp}}+1)m}$, and then defining $\boldsymbol{P}_i = \text{diag}(p[i], \dots, p[i+N-1])$ and

$$\boldsymbol{W}_i = \check{\boldsymbol{F}} \boldsymbol{P}_i^H \quad (4.13)$$

$$\boldsymbol{W} = [\boldsymbol{W}_{-N_{\text{delay}}+1}^\top \cdots \boldsymbol{W}_{N_{\text{delay}}-1}^\top]^\top, \quad (4.14)$$

we see that (4.12) can be written as $\boldsymbol{W}\boldsymbol{b}_q = \mathbf{0}$, implying that each \boldsymbol{b}_q must lie in the null space of \boldsymbol{W} . This latter condition is achieved by choosing \boldsymbol{B} such that its columns form a basis for $\text{null}(\boldsymbol{W})$. Clearly, N_s , the number of columns in \boldsymbol{B} , must obey $N_s \leq \dim(\text{null}(\boldsymbol{W}))$. For the case when $N_s = \dim(\text{null}(\boldsymbol{W}))$, we can bound the data dimension N_s as follows. Notice from (4.14) that the $N_{\text{Dopp}}N_{\text{delay}}$ rows of

$(\mathbf{P}\mathbf{U})^H$ are contained within the $(2N_{\text{Dopp}} - 1)(2N_{\text{delay}} - 1)$ rows of \mathbf{W} . To satisfy (4.8), the rows of $(\mathbf{P}\mathbf{U})^H$ must be orthogonal. Thus, $N_{\text{Dopp}}N_{\text{delay}} \leq \text{rank}(\mathbf{W}) \leq (2N_{\text{Dopp}} - 1)(2N_{\text{delay}} - 1)$, and so (4.11)-(4.12) imply

$$N - (2N_{\text{Dopp}} - 1)(2N_{\text{delay}} - 1) \leq N_s \leq N - N_{\text{Dopp}}N_{\text{delay}}. \quad (4.15)$$

From (4.15), we see that MMSE-PAT dedicates at least $N_{\text{delay}}N_{\text{Dopp}}$, but no more than $(2N_{\text{delay}} - 1)(2N_{\text{Dopp}} - 1)$, signaling dimensions to pilots. The implication is that the PAT schemes which achieve equality in (4.6) obey $N_s \leq N - N_{\text{Dopp}}N_{\text{delay}}$. This result is intuitive, because, to minimize MSE, the pilot symbols must estimate all $N_{\text{Dopp}}N_{\text{delay}}$ independent BEM coefficients, consuming at least $N_{\text{Dopp}}N_{\text{delay}}$ signaling dimensions. Our analysis also shows that PAT schemes can support a data dimension of at least $N - (2N_{\text{Dopp}} - 1)(2N_{\text{delay}} - 1)$ without sacrificing MSE.

The design procedure for MMSE-PAT (\mathbf{p}, \mathbf{B}) for the DSC can be summarized as follows.

1. Find a pilot sequence \mathbf{p} which satisfies (4.11). One such procedure is given in Appendix C.3.
2. Using \mathbf{p} , construct \mathbf{W} from (4.13)-(4.14).
3. Find an orthonormal basis for the null space of \mathbf{W} and create \mathbf{B} column-wise from this basis.

This constructive MMSE-PAT design procedure establishes the achievability of the lower bound in (4.6). We give four examples of MMSE-PAT in Section 4.3.

Time-frequency duality is a well-known and fundamental concept in communication theory (see, e.g., [17]). While it has recently been applied to the design of

space-time codes (e.g., [56]), we now demonstrate that it can be applied to the design of affine MMSE-PAT schemes for the DSC. We refer to $(\check{\mathbf{p}}, \check{\mathbf{B}})$, for $\check{\mathbf{p}} = \mathbf{F}_N \mathbf{p}$ and $\check{\mathbf{B}} = \mathbf{F}_N \mathbf{B}$ by taking DFT, as the ‘‘frequency-domain counterpart’’ of (\mathbf{p}, \mathbf{B}) . Using this notation, Lemma 3 gives the frequency-domain analogy of Lemma 2.

Lemma 3 (Frequency Domain Conditions). *For N -block CP affine PAT over the CE-BEM DSC, the pair (4.16)-(4.17) form necessary and sufficient conditions for equality in (4.6).*

$$\sum_{i=0}^{N-1} \check{p}[i] \check{p}^*[i-k] e^{-j \frac{2\pi}{N} m i} = E_p \delta[k] \delta[m] \quad \forall k \in \mathcal{N}_{\text{Dopp}}, \quad \forall m \in \mathcal{N}_{\text{delay}}, \quad (4.16)$$

$$\sum_{i=0}^{N-1} \check{b}_q[i] \check{p}^*[i-k] e^{-j \frac{2\pi}{N} m i} = 0 \quad \forall k \in \mathcal{N}_{\text{Dopp}}, \quad \forall m \in \mathcal{N}_{\text{delay}}, \quad \forall q \in \{0, \dots, N_s - 1\}. \quad (4.17)$$

Proof. Plugging $p[i] = \frac{1}{\sqrt{N}} \sum_{r=0}^{N-1} \check{p}[r] e^{j \frac{2\pi}{N} i r}$ into (4.11), we have

$$\begin{aligned} \sum_{i=0}^{N-1} p[i] p^*[i-k] e^{-j \frac{2\pi}{N} m i} &= \sum_{i=0}^{N-1} \left(\frac{1}{\sqrt{N}} \sum_{r_1=0}^{N-1} \check{p}[r_1] e^{j \frac{2\pi}{N} i r_1} \right) \left(\frac{1}{\sqrt{N}} \sum_{r_2=0}^{N-1} \check{p}^*[r_2] e^{-j \frac{2\pi}{N} (i-k) r_2} \right) e^{-j \frac{2\pi}{N} m i} \\ &= \frac{1}{N} \sum_{r_1=0}^{N-1} \sum_{r_2=0}^{N-1} \check{p}[r_1] \check{p}^*[r_2] e^{j \frac{2\pi}{N} k r_2} \underbrace{\sum_{i=0}^{N-1} e^{j \frac{2\pi}{N} (r_1 - r_2 - m) i}}_{N \delta[r_1 - r_2 - m]} \\ &= e^{-j \frac{2\pi}{N} m k} \sum_{r_1=0}^{N-1} \check{p}[r_1] \check{p}^*[r_1 - m] e^{j \frac{2\pi}{N} r_1 k}, \end{aligned}$$

and since $\{k \in \mathcal{N}_{\text{delay}}, m \in \mathcal{N}_{\text{Dopp}}\} \Leftrightarrow \{-k \in \mathcal{N}_{\text{delay}}, -m \in \mathcal{N}_{\text{Dopp}}\}$, the conditions (4.11) and (4.16) are equivalent. In a similar manner, the equivalence between the conditions (4.12) and (4.17) can be established. □

Note that the conditions in Lemma 3 mirror those in Lemma 2, except that the discrete delay spread N_{delay} and discrete Doppler spread N_{Dopp} have interchanged their roles. Our duality result can be stated concisely as follows.

Lemma 4 (Duality). *If (\mathbf{p}, \mathbf{B}) parameterizes N -block CP MMSE-PAT over the CE-BEM DSC with discrete delay spread N_1 and discrete Doppler spread N_2 , then $(\mathbf{F}_N \mathbf{p}, \mathbf{F}_N \mathbf{B})$ parameterizes N -block CP MMSE-PAT for the CE-BEM DSC with discrete delay spread N_2 and discrete Doppler spread N_1 and vice versa.*

To our knowledge, the application of time-frequency duality to the design of DSC MMSE-PAT schemes is novel. Similarities between the structure of MMSE-PAT schemes for SCM over time-selective channels [57] and OFDM over frequency selective channels [35] have been previously noted in [57]. However, our result on the duality of DSC MMSE-PAT is more general, and applies to any affine modulation scheme (including, but not limited to, SCM and OFDM), as illustrated by the MMSE-PAT examples given in Section 4.3.

4.3 Examples of MMSE-PAT

Here we give several examples of N -block CP affine MMSE-PAT schemes for the CE-BEM DSC with discrete delay spread N_{delay} and discrete Doppler spread N_{Dopp} , using the (\mathbf{p}, \mathbf{B}) parameterization. The proofs for the MMSE-optimality of the following examples are given in Appendix C.4.

Example 1 (TDKD). *Assuming $\frac{N}{N_{\text{Dopp}}} \in \mathbb{Z}$, define the pilot index set $\mathcal{P}_t^{[\ell]}$ and the guard index set $\mathcal{G}_t^{[\ell]}$:*

$$\begin{aligned} \mathcal{P}_t^{[\ell]} &= \left\{ \ell, \ell + \frac{N}{N_{\text{Dopp}}}, \dots, \ell + \frac{(N_{\text{Dopp}}-1)N}{N_{\text{Dopp}}} \right\} \\ \mathcal{G}_t^{[\ell]} &= \bigcup_{k \in \mathcal{P}_t^{[\ell]}} \{-N_{\text{delay}} + 1 + k, \dots, N_{\text{delay}} - 1 + k\}. \end{aligned}$$

An N -block CP MMSE-PAT scheme for the CE-BEM DSC is given by

$$p[k] = \begin{cases} \sqrt{\frac{E_p}{N_{\text{Dopp}}}} e^{j\theta[k]} & k \in \mathcal{P}_t^{[\ell]} \\ 0 & k \notin \mathcal{P}_t^{[\ell]} \end{cases} \quad (4.18)$$

and by \mathbf{B} constructed from the columns of \mathbf{I}_N with indices not in the set $\mathcal{G}_t^{[\ell]}$. Both $\ell \in \{0, \dots, \frac{N}{N_{\text{Dopp}}} - 1\}$ and $\theta[k] \in \mathbb{R}$ are arbitrary. The corresponding data dimension is $N_s = N - N_{\text{Dopp}}(2N_{\text{delay}} - 1)$.

Example 1 specifies a PAT scheme in which the data and pilot sequences are non-overlapping in the time domain, where the pilot pattern consists of periodic time-domain bursts, and where each burst has a truncated Kronecker-delta structure, with ℓ controlling the time-offset of the first burst. (See Fig. 4.2). Note that the burst period $\frac{N}{N_{\text{Dopp}}} \approx \frac{1}{2B_{\text{Dopp}}T_s}$ satisfies a Nyquist-sampling criterion. This “time-domain Kronecker delta” (TDKD) scheme bears similarity to the PAT scheme proposed in [58] (heuristically) and later in [13], with the difference that [13] focused on zero-padded (ZP) block transmission, which allows for efficient concatenation of blocks. We undertake a detailed comparison of CP and ZP schemes in Section 4.6.

Example 2 (FDKD). Assuming $\frac{N}{N_{\text{delay}}} \in \mathbb{Z}$, define the pilot index set $\mathcal{P}_f^{[\ell]}$ and the guard index set $\mathcal{G}_f^{[\ell]}$:

$$\begin{aligned}\mathcal{P}_f^{[\ell]} &= \left\{ \ell, \ell + \frac{N}{N_{\text{delay}}}, \dots, \ell + \frac{(N_{\text{delay}}-1)N}{N_{\text{delay}}} \right\} \\ \mathcal{G}_f^{[\ell]} &= \bigcup_{k \in \mathcal{P}_f^{[\ell]}} \{-N_{\text{Dopp}} + 1 + k, \dots, N_{\text{Dopp}} - 1 + k\}.\end{aligned}$$

An N -block CP MMSE-PAT scheme for the CE-BEM DSC is given by $\mathbf{p} = \mathbf{F}_N^H \check{\mathbf{p}}$ and $\mathbf{B} = \mathbf{F}_N^H \check{\mathbf{B}}$ with

$$\check{\mathbf{p}}[k] = \begin{cases} \sqrt{\frac{E_p}{N_{\text{delay}}}} e^{j\theta[k]} & k \in \mathcal{P}_f^{[\ell]} \\ 0 & k \notin \mathcal{P}_f^{[\ell]} \end{cases}, \quad (4.19)$$

and by $\check{\mathbf{B}}$ constructed from the columns of the \mathbf{I}_N with indices not in the set $\mathcal{G}_f^{[\ell]}$. Both $\ell \in \{0, \dots, \frac{N}{N_{\text{delay}}} - 1\}$ and $\theta[k] \in \mathbb{R}$, are arbitrary. The corresponding data dimension is $N_s = N - N_{\text{delay}}(2N_{\text{Dopp}} - 1)$.

Example 2 specifies a PAT scheme in which the data and pilot sequences are non-overlapping in the frequency domain, where the pilot pattern consists of periodic sub-carrier clusters, and where each cluster has a truncated Kronecker-delta structure, with ℓ controlling the offset of the first cluster. (See Fig. 4.2). Note that the cluster spacing $\frac{N}{N_{\text{delay}}}$ satisfies a frequency-domain Nyquist-sampling criterion. This “frequency-domain Kronecker delta” (FDKD) scheme is the time-frequency dual of TDKD and bears similarity to the heuristic PAT schemes proposed in [58] and [59]. For the special case of frequency-selective channels (i.e., $N_{\text{Dopp}} = 1$), FDKD coincides with the MSE-optimal OFDM system identified in [34, 35], where the pilot clusters reduce to pilot tones.

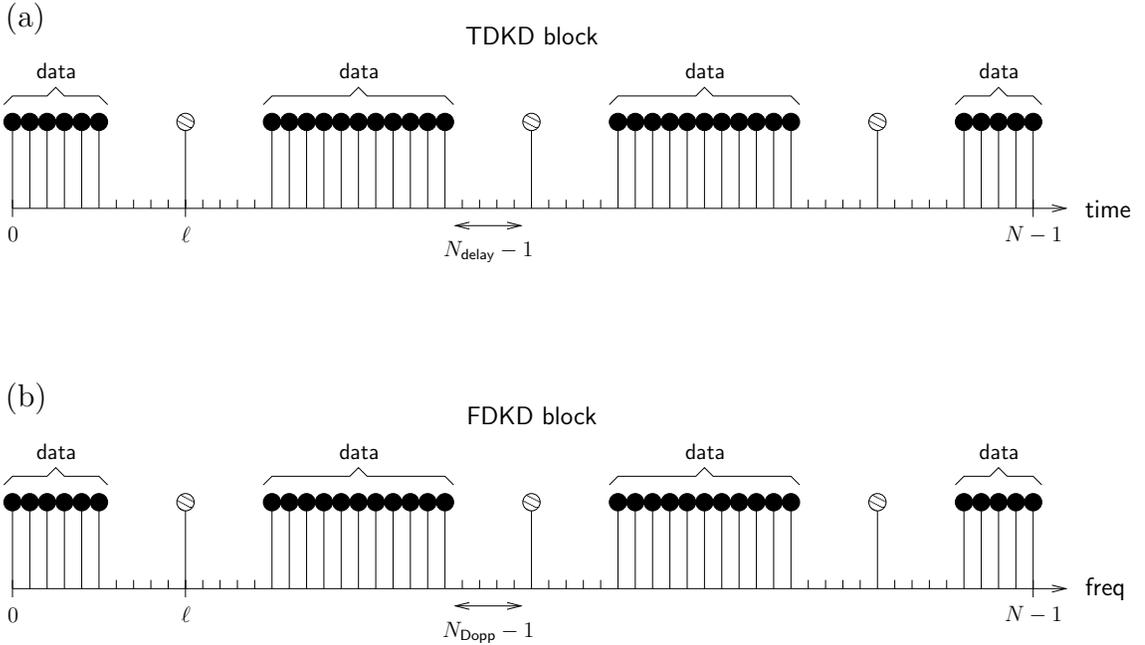


Figure 4.2: (a) Structure of TDKD with $N_{\text{delay}} = 5$ and $N_{\text{Dopp}} = 3$. (b) Structure of FDKD with $N_{\text{delay}} = 3$ and $N_{\text{Dopp}} = 5$.

Example 3 (Time domain Chirps). *Assuming even N , an N -block CP MMSE-PAT scheme for the CE-BEM DSC is given by*

$$p[k] = \sqrt{\frac{E_p}{N}} e^{j \frac{2\pi}{N} \frac{N_{\text{Dopp}}}{2} k^2} \quad (4.20)$$

$$[\mathbf{B}]_{k,q} = \frac{1}{\sqrt{N}} e^{j \frac{2\pi}{N} (q + N_{\text{Dopp}} N_{\text{delay}}) k} e^{j \frac{2\pi}{N} \frac{N_{\text{Dopp}}}{2} k^2}, \quad (4.21)$$

for $k \in \{0, \dots, N-1\}$ and $q \in \{0, \dots, N_s-1\}$, where the data dimension $N_s = N - 2N_{\text{Dopp}}N_{\text{delay}} + 1$.

Example 4 (Frequency domain Chirps). *Assuming even N , an N -block CP MMSE-PAT scheme for the CE-BEM DSC is given by $\mathbf{p} = \mathbf{F}_N^H \check{\mathbf{p}}$ and $\mathbf{B} = \mathbf{F}_N^H \check{\mathbf{B}}$, with*

$$\check{p}[k] = \sqrt{\frac{E_p}{N}} e^{j \frac{2\pi}{N} \frac{N_{\text{delay}}}{2} k^2} \quad (4.22)$$

$$[\check{\mathbf{B}}]_{k,q} = \frac{1}{\sqrt{N}} e^{j \frac{2\pi}{N} (q + N_{\text{Dopp}} N_{\text{delay}}) k} e^{j \frac{2\pi}{N} \frac{N_{\text{delay}}}{2} k^2}, \quad (4.23)$$

for $k \in \{0, \dots, N-1\}$ and $q \in \{0, \dots, N_s-1\}$, where the data dimension $N_s = N - 2N_{\text{Dopp}}N_{\text{delay}} + 1$.

To our knowledge, the “chirp” schemes specified in Examples 3 and 4 are novel DSC MMSE PAT schemes. We refer to examples 3 and 4 as “chirp” schemes because the pilot and data waveforms are linear chirps, as evidenced by the k^2 term in the complex exponentials. Though a chirp-based PAT scheme was suggested in [60], it was not MMSE. With our chirp schemes, the pilots and data are superimposed in both the time and frequency domains, though still *linearly separable* at the channel output, as required by (4.7). Note that the PATs in Example 3 and Example 4 are time-frequency duals. Other novel MMSE-PAT schemes can be obtained from the design procedure outlined in Section 4.2.

4.4 Achievable-Rate Analysis of MMSE-PAT

We now calculate bounds on the ergodic achievable rate of affine precoded MMSE-PAT schemes for the CE-BEM DSC. Lemma 2, which distills the fundamental properties of these MMSE-PAT schemes, allows this achievable-rate analysis to be conducted in an unified manner. We deepen our analysis in the high-SNR regime and compare the MMSE-PAT examples from Section 4.3.

As before, we consider the transmitter and channel models discussed in Section 4.1. Suppose that the MMSE-PAT scheme (\mathbf{p}, \mathbf{B}) has pilot energy E_p [recall (4.3)] and yields data dimension N_s . It will be convenient to write the input-output relation (4.1) as

$$\mathbf{y} = \sqrt{\rho} \hat{\mathbf{H}} \mathbf{p} + \sqrt{\rho} \hat{\mathbf{H}} \mathbf{B} \mathbf{s} + \mathbf{v}. \quad (4.24)$$

where $\hat{\mathbf{H}} \in \mathbb{C}^{N \times N}$ is defined element-wise as

$$[\hat{\mathbf{H}}]_{n,m} = h[n; \langle n - m \rangle_N], \quad (4.25)$$

with $h[n; l] = 0$ for $l \neq \{0, \dots, N_{\text{delay}} - 1\}$. We also define $\hat{\mathbf{H}}$ and $\tilde{\mathbf{H}}$ in the same way as $\hat{\mathbf{H}}$, but from $\hat{\mathbf{h}}$ and $\tilde{\mathbf{h}}$, respectively. It can be verified that $E\{\hat{\mathbf{H}}\hat{\mathbf{H}}^H\} = \mathbf{I}$, $E\{\hat{\mathbf{H}}\hat{\mathbf{H}}^H\} = \mathbf{I} \text{tr}\{\mathbf{R}_{\hat{\mathbf{h}}}\}/N$ and $E\{\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H\} = \mathbf{I} \text{tr}\{\mathbf{R}_{\tilde{\mathbf{h}}}\}/N$, where $\mathbf{R}_{\hat{\mathbf{h}}}$ and $\mathbf{R}_{\tilde{\mathbf{h}}}$ were given in (4.9) and (4.10), respectively.

For Section 4.4, we make an additional assumption on our model. We assume that the columns of precoding matrix \mathbf{B} are orthonormal (whereas earlier \mathbf{B} was specified as full rank). Note that, for the purpose of ergodic achievable-rate analysis, this assumption can be made w.l.o.g., since the mutual information between \mathbf{s} and \mathbf{y} remains unaffected by invertible transformations of \mathbf{s} . We then define the data

energy E_s ,

$$E_s = E\{\|\mathbf{d}\|^2\} = E\{\|\mathbf{s}\|^2\}, \quad (4.26)$$

the total energy $E_{\text{tot}} = E_p + E_s$. To meet the average power constraint, We require that $E_{\text{tot}} \leq N$. We have the average SNR of the system equals ρ , due to our energy-preserving channel model. Finally, we define the normalized data power $\sigma_s^2 = E_s/N_s$. Note that these energy and power definitions do not take the CP into account; the effects of the CP will be discussed in Section 4.6.

Now, we obtain achievable-rate bounds for the MMSE-PAT schemes using Gaussian codes and minimum distance decoding (Section 2.3.2). To present the bounds, we use the normalized channel estimate $\bar{\mathbf{H}} = \hat{\mathbf{H}}\sqrt{N/\text{tr}\{\mathbf{R}_{\hat{h}}\}}$, where $E\{\bar{\mathbf{H}}\bar{\mathbf{H}}^H\} = \mathbf{I}$.

Theorem 3 (Achievable-Rate Bounds). *For the N -block CP affine MMSE-PAT scheme (\mathbf{p}, \mathbf{B}) with zero-mean i.i.d. Gaussian $\mathbf{s} \in \mathbb{C}^{N_s}$ over the CE-BEM DSC, the per-block ergodic achievable rate $\mathcal{R}_{\text{mmse-blk}}$ obeys $\mathcal{R}_{\text{mmse-blk-lb}} \leq \mathcal{R}_{\text{mmse-blk}} \leq \mathcal{R}_{\text{mmse-blk-ub}}$, where*

$$\mathcal{R}_{\text{mmse-blk-lb}} = E\{\log \det[\mathbf{I}_{N_s} + \rho_{\text{lb}}\mathbf{B}^H\bar{\mathbf{H}}^H\bar{\mathbf{H}}\mathbf{B}]\} \text{ bits/block} \quad (4.27)$$

$$\mathcal{R}_{\text{mmse-blk-ub}} = E\{\log \det[\mathbf{I}_{N_s} + \rho_{\text{ub}}\mathbf{B}^H\hat{\mathbf{H}}^H\hat{\mathbf{H}}\mathbf{B}]\} \text{ bits/block} \quad (4.28)$$

$$\rho_{\text{lb}} = \frac{\rho\sigma_s^2 \text{tr}\{\mathbf{R}_{\hat{h}}\}}{\rho\sigma_s^2 \text{tr}\{\mathbf{R}_{\hat{h}}\} + N} \quad (4.29)$$

$$\rho_{\text{ub}} = \rho\sigma_s^2. \quad (4.30)$$

Proof. See Appendix C.5. □

The lower bound (4.27) describes the “worst case” scenario of channel estimation error acting as additive Gaussian noise. This concept was previously used in, e.g., [41] and [61]. The upper bound (4.28) describes the “best case” scenario of perfect channel

estimates. In Section 4.6, we consider the effects of the CP and write the achievable rate in units of bits/sec/Hz.

It is insightful to compare $\mathcal{R}_{\text{mmse-blk}}$, the achievable rate of MMSE-PAT, to $\mathcal{R}_{\text{csi-blk}}$, the achievable rate of a system with perfect receiver-CSI, under equal transmission power. (With perfect CSI, there is, of course, no need for pilots.) With i.i.d. Gaussian data, the achievable rates are known to be [33]

$$\mathcal{R}_{\text{csi-blk}} = E\{\log \det[\mathbf{I}_N + \rho_{\text{csi}} \hat{\mathbf{H}}^H \hat{\mathbf{H}}]\} \text{ bits/block} \quad (4.31)$$

$$\rho_{\text{csi}} = \rho. \quad (4.32)$$

Two principle factors separate $\mathcal{R}_{\text{csi-blk}}$ from $\mathcal{R}_{\text{mmse-blk}}$. First, MMSE-PAT suffers from channel estimation error, which degrades $\mathcal{R}_{\text{mmse-blk}}$ by affecting the “effective SNR” (4.29). Second, MMSE-PAT uses only N_s out of N total signaling dimensions for data transmission. Note, from (4.24), (4.27) and (4.28), that MMSE-PAT communicates the data \mathbf{s} through the “effective channel” $\hat{\mathbf{H}}\mathbf{B} \in \mathbb{C}^{N \times N_s}$, which offers only N_s degrees of freedom. The perfect-receiver-CSI system, on the other hand, communicates data through the effective channel $\hat{\mathbf{H}} \in \mathbb{C}^{N \times N}$, which offers N degrees of freedom. Our asymptotic achievable rates in Theorem 4 provides further insight on these issues.

Theorem 4 (High-SNR Achievable Rate). *For the N -block CP MMSE-PAT scheme (\mathbf{p}, \mathbf{B}) with data dimension N_s over the CE-BEM DSC, the per-block ergodic rate is given by*

$$\mathcal{R}_{\text{mmse-blk}} = N_s \log \rho + O(1) \text{ bits/block} \quad (4.33)$$

as $\rho \rightarrow \infty$.

Proof. See Appendix C.6. □

In the case of N -block transmission under perfect receiver-CSI, it was shown in Chapter 3 that the asymptotic capacity obeys

$$\mathcal{C}_{\text{coh-blk}} = N \log \rho + O(1) \text{ bits/block} \quad (4.34)$$

as $\rho \rightarrow \infty$. Notice that, in the high-SNR regime, $\mathcal{R}_{\text{mmse-blk}}$ and $\mathcal{C}_{\text{coh-blk}}$ increase linearly in $\log \rho$ with slopes N_s and N , respectively. Recalling the range of N_s from (4.15), it is evident that MMSE-PAT suffers a pre-log factor penalty relative to the perfect-receiver-CSI system. In fact, at high SNR, the loss in pre-log factor becomes the dominant cost of imperfect receiver-CSI, since there the MMSE-PAT channel estimates will be accurate and losses due to channel estimation error will be insignificant. For channels with low spreading index (i.e., $\gamma \ll 1$), N_s will be close to N (recall $\gamma = \frac{N_{\text{Dopp}} N_{\text{delay}}}{N}$), and hence the pre-log factors of $\mathcal{R}_{\text{mmse-blk}}$ and $\mathcal{C}_{\text{coh-blk}}$ will be similar. Thus, for low spreading indices and moderately high SNR, the rates achieved by MMSE-PAT should not be far from those of a perfect-receiver-CSI system. This is intuitively satisfying, since smaller γ implies relatively few unknown channel parameters and thus relatively small pilot overhead. On the other hand, for channels with high spreading index (γ close to 1), N_s will be significantly less than N [recall (4.15)]. In this case, $\mathcal{R}_{\text{mmse-blk}}$ can deviate significantly from $\mathcal{C}_{\text{coh-blk}}$, even at moderately high SNR. These trends are confirmed by the numerical results in Section 4.7.

4.5 Pilot/Data Power Allocation

Until now, the MMSE-PAT schemes were designed using fixed pilot energy E_p . Now we consider the problem of judiciously allocating a fixed energy E_{tot} between pilots and data. Notice the inherent tradeoff: increasing the pilot power decreases the channel estimation error but also decreases the data power, which in turn increases

the sensitivity to noise and channel estimation error. Intuitively, power should be allocated to maximize an “effective SNR” which takes into account both the noise and channel estimation errors. The approach we take is to maximize ρ_{lb} .

Let $\alpha \in (0, 1)$ denote the fraction of energy allocated to the data symbols, i.e., $E_s = \alpha E_{\text{tot}}$ and $E_p = (1 - \alpha)E_{\text{tot}}$. We are interested in finding $\alpha_\star = \arg \max_\alpha \rho_{\text{lb}}(\alpha)$. Recall that α_\star must satisfy $\left. \frac{\partial \rho_{\text{lb}}(\alpha)}{\partial \alpha} \right|_{\alpha=\alpha_\star} = 0$, which is equivalent to satisfying

$$(\varphi(\alpha_\star) + 1) \left(\frac{\rho E_{\text{tot}}}{N_s} - \varphi'(\alpha_\star) \right) = \left(\frac{\alpha_\star \rho E_{\text{tot}}}{N_s} - \varphi(\alpha_\star) \right) \varphi'(\alpha_\star), \quad (4.35)$$

$$\varphi(\alpha) = \sum_{i=0}^{N_{\text{Dopp}} N_{\text{delay}} - 1} \frac{\alpha \rho E_{\text{tot}}}{N N_s [\mathbf{R}_\lambda]_{i,i}^{-1} + (1 - \alpha) \rho E_{\text{tot}} N_s}, \quad (4.36)$$

where $\varphi'(\alpha) = \partial \varphi / \partial \alpha$. Numerical techniques can be used to find the roots, within the interval $(0, 1)$, of the polynomial (4.35). Among these roots, α_\star is the one which maximizes ρ_{lb} .

In the case of identically distributed BEM coefficients, i.e., $\mathbf{R}_\lambda = \frac{N}{N_{\text{Dopp}} N_{\text{delay}}} \mathbf{I}_{N_{\text{Dopp}} N_{\text{delay}}}$, it can be shown that the maximizer of $\rho_{\text{lb}}(\alpha)$ is

$$\alpha_{\star, \text{iid}} = \begin{cases} \beta - \sqrt{\beta^2 - \beta} & \text{if } N_s > N_{\text{delay}} N_{\text{Dopp}} \\ \beta + \sqrt{\beta^2 - \beta} & \text{if } N_s < N_{\text{delay}} N_{\text{Dopp}} \\ \frac{1}{2} & \text{if } N_s = N_{\text{delay}} N_{\text{Dopp}} \end{cases} \quad (4.37)$$

$$\beta = \frac{1 + \frac{N_{\text{Dopp}} N_{\text{delay}}}{\rho E_{\text{tot}}}}{1 - \frac{N_{\text{Dopp}} N_{\text{delay}}}{N_s}}. \quad (4.38)$$

Furthermore, it can be shown that $\alpha_{\star, \text{iid}}$ maximizes the achievable-rate lower bound given in (4.27). To see this, note that the channel \mathbf{h} and the normalized channel estimate $\bar{\mathbf{h}} = \hat{\mathbf{h}} \sqrt{N / \text{tr}\{\mathbf{R}_{\hat{\mathbf{h}}}\}}$ have the same covariance, and hence $\hat{\mathbf{H}}$ and $\bar{\mathbf{H}}$ have the same distribution, so that the power allocation fraction affects the $\mathcal{R}_{\text{mmse-blk-lb}}$ only through ρ_{lb} . Since $\mathcal{R}_{\text{mmse-blk-lb}}$ is an increasing function of ρ_{lb} , maximizing ρ_{lb} is equivalent to maximizing $\mathcal{R}_{\text{mmse-blk-lb}}$.

For the case of general \mathbf{R}_λ , closed-form solutions for α_{\max} are possible in the high-SNR and low-SNR asymptotic cases. It can be shown that $\arg \max_\alpha \lim_{\rho \rightarrow 0} \rho_{\text{lb}} = \frac{1}{2}$ and $\arg \max_\alpha \lim_{\rho \rightarrow \infty} \rho_{\text{lb}} = \lim_{\rho \rightarrow \infty} \alpha_{\star, \text{iid}}$, where $\alpha_{\star, \text{iid}}$ is calculated from (4.37) using $\beta = (1 - N_{\text{Dopp}} N_{\text{delay}} / N_s)^{-1}$. Note that all affine precoding MMSE-PAT schemes with data dimension N_s have ρ_{lb} maximized by the same pilot/data power allocation.

4.6 Streaming MMSE-PAT

In this section, we analyze the spectral efficiency of systems which transmit a stream of blocks, where each block is constructed according to the MMSE-PAT principles discussed earlier and separated from its neighbors by time-domain guard intervals. We consider guards based on CP (as assumed in Section 4.1) as well as ZP (as assumed, e.g., in [13]). In particular, we examine the time-bandwidth resources consumed by these systems and analyze their achievable rates in units of bits/sec/Hz. To quantify time-bandwidth consumption, we consider the use of an arbitrary continuous-time baseband-equivalent pulse. The analysis in this section facilitates a direct comparison between the MMSE-PAT schemes in Section 4.3 and the zero-padded SCM scheme from [13].

Let $\{x^{[m]}[i]\}_{i=0}^{N-1}$ denote the discrete-time transmitted sequence within the m^{th} block. We assume that $\{x^{[m]}[i]\}_{i=0}^{N-1}$ is constructed in the manner of $\{x[i]\}_{i=0}^{N-1}$ from (2.25), but with pilot and data that satisfy the MMSE-PAT conditions in Lemma 2. Recall from Section 2.1 that the baseband modulation is accomplished by a time-limited and band-limited pulse $\psi(t)$. The streaming cyclic-prefixed (SCP) modulator generates the continuous-time transmitted waveform as,

$$x(t) = \sum_{m \in \mathbb{Z}} \sum_{i=-N_{\text{delay}}+1}^{N-1} x^{[m]}[i] \psi(t - i\mathcal{T}_s - m(N + N_{\text{delay}} - 1)\mathcal{T}_s), \quad (4.39)$$

where modulo- N indexing is assumed for $x^{[m]}[i]$ in (4.39). We refer to SCP with $\{x^{[m]}[i]\}_{i=0}^{N-1}$ constructed according to the FDKD example from Example 2, with $\ell = N/N_{\text{delay}} - N_{\text{Dopp}}$ and arbitrary $\theta[k]$, as “SCP-FDKD.” Similarly, we use “SCP-Chirp” to refer to the corresponding schemes constructed from either the time- or frequency-domain Chirp examples from Section 4.3. (See Fig. 4.3.) Because of the CP, the SCP block period equals $(N + N_{\text{delay}} - 1)\mathcal{T}_s$ seconds. For the streaming zero-padded (SZP) PAT scheme from [13], the modulator generates

$$x(t) = \sum_{m \in \mathbb{Z}} \sum_{i=0}^{N-1} x^{[m]}[i] \psi(t - i\mathcal{T}_s - mN\mathcal{T}_s). \quad (4.40)$$

We refer to SZP with $\{x^{[m]}[i]\}_{i=0}^{N-1}$ constructed according to the TDKD example from Example 1, with $\ell = N/N_{\text{Dopp}} - N_{\text{delay}}$ and arbitrary $\theta[k]$, as “SZP-TDKD.” (See Fig. 4.3.) Noting that the TDKD samples $\{x^{[m]}[i]\}_{i=N-N_{\text{delay}}+1}^{N-1}$ are zero-valued for this choice of ℓ , it can be seen that SZP-TDKD takes advantage of a “built in” ZP, which permits efficient concatenation of blocks (Fig. 4.3) at a block period of $N\mathcal{T}_s$ seconds. The continuous-time channel output is

$$y(t) = \int h(t; \tau) x(t - \tau) + v(t), \quad (4.41)$$

where $h(t; \tau)$ denotes the (time-varying) continuous-time DSC impulse response and $v(t)$ denotes the noise waveform.

We now compute the spectral efficiencies of SZP-TDKD, SCP-FDKD and SCP-Chirp schemes in bits/sec/Hz by quantifying the per-block time-bandwidth resources consumed by these systems. To do this, we examine the continuous time channel *output* signal $y(t)$. The decision to examine the channel output rather than the channel input is somewhat arbitrary, but results in cleaner expressions. In addition, channel output properties are more relevant than channel input properties when analyzing

gress onto adjacent frequency bands. In the sequel, we make the approximation $2\mathcal{B}_{\text{Dopp}} \approx \frac{N_{\text{Dopp}}-1}{N\mathcal{T}_s}$, which is accurate for large N .⁴

As illustrated in Fig. 4.3, each block of SCP-FDKD and SCP-Chirp schemes consumes $(N + N_{\text{delay}} - 1)\mathcal{T}_s$ seconds while that of SZP-TDKD consumes $N\mathcal{T}_s$ seconds. To find the bandwidth consumption of these systems, we assume uncorrelated information symbols $\{s^{[m]}[k]\}$ (e.g., from an i.i.d. Gaussian codebook). SZP-TDKD is an SCM system and its transmit signal occupies a bandwidth of $\frac{1}{\mathcal{T}_s}$ Hz. Because of the frequency dispersion of DSC, the channel output has bandwidth of $\frac{1}{\mathcal{T}_s} + 2\mathcal{B}_{\text{Dopp}} = \frac{N+N_{\text{Dopp}}-1}{N\mathcal{T}_s}$ Hz, so that SZP-TDKD consumes $N + N_{\text{Dopp}} - 1$ Hz sec per block. SCP-FDKD is a multi-carrier system transmitting pilots and data in the frequency domain, where each sub-carrier consumes a bandwidth of $\frac{1}{N\mathcal{T}_s}$ Hz. Because of the zero-valued pilot sub-carriers at the edge of the band (with the choice of $\ell = \frac{N}{N_{\text{delay}}} - N_{\text{Dopp}}$), the SCP-FDKD output signal consumes a bandwidth of $\frac{N-N_{\text{Dopp}}+1}{N\mathcal{T}_s} + 2\mathcal{B}_{\text{Dopp}} = \frac{1}{\mathcal{T}_s}$ Hz. Thus, SCP-FDKD consumes $N - N_{\text{delay}} + 1$ Hz sec per block. Notice that SCP-FDKD consumes more time resources, but less frequency resources, (per block) than SZP-TDKD. This follows from the fact that SCP-FDKD contains built-in frequency-domain guard intervals while SZP-TDKD contains built-in time-domain guard intervals. Both the SCP-Chirp schemes consume a bandwidth of $\frac{1}{\mathcal{T}_s} + 2\mathcal{B}_{\text{Dopp}} = \frac{N+N_{\text{Dopp}}-1}{N\mathcal{T}_s}$ Hz, so that their time-bandwidth consumption is $(N + N_{\text{delay}} - 1)(N + N_{\text{Dopp}} - 1)/N$ Hz sec per block.

For CP-based systems, we have assumed throughout that the receiver discards the samples corresponding to the CP. (Recall Section 4.1.) In contrast, the SZP-TDKD

⁴By writing $N_{\text{Dopp}} = 2\lceil \mathcal{B}_{\text{Dopp}}\mathcal{T}_s N \rceil + 1 = 2\mathcal{B}_{\text{Dopp}}\mathcal{T}_s N - \epsilon + 1$ where $\epsilon \in [0, 1)$, we see that $2\mathcal{B}_{\text{Dopp}} = \frac{N_{\text{Dopp}}-1}{N\mathcal{T}_s} + \frac{\epsilon}{N\mathcal{T}_s}$. We desire that the approximation error satisfies $\frac{\epsilon}{N\mathcal{T}_s} \ll \frac{1}{\mathcal{T}_s}$, or equivalently $N \gg \epsilon$, and this is guaranteed when $N \gg 1$.

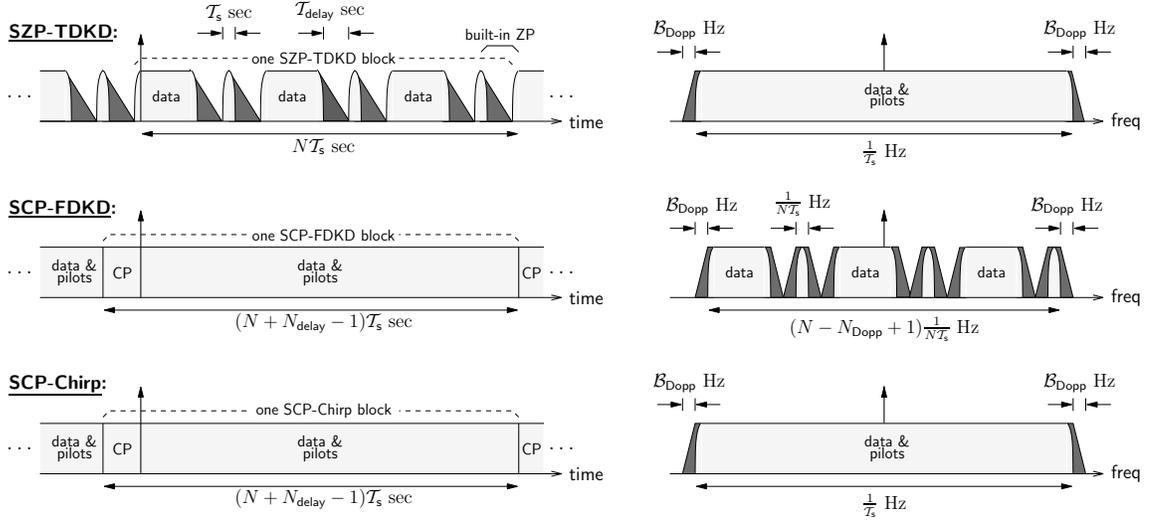


Figure 4.3: Time and bandwidth occupation for several streaming MMSE-PAT systems designed for $N_{\text{delay}} = 3$ and $N_{\text{Dopp}} = 3$. Lightly shaded shows channel input and darkly shaded shows channel output. Note the use of a time-frequency concentrated modulation pulse $\psi(t)$.

receiver does not discard samples. Thus, when comparing these systems, we must be careful when defining SNR.

Consider SNR defined as *the ratio of signal power to noise power observed at the output of the receiver's pulse-shaping filter*. Because we assume an energy preserving discrete-time channel $\{h[i; l]\}$, this SNR can be equivalently described as the ratio of transmitted signal power to received noise power. Since SZP-TDKD has a built-in ZP, the transmitted signal power is $\sigma_{\text{zp}}^2 = \rho \frac{1}{N} \sum_{i=0}^{N-1} \text{E}\{|x^{[m]}[i]|^2\} = \frac{\rho E_{\text{tot}}}{N}$. For SCP schemes, the transmitted signal power is $\sigma_{\text{cp}}^2 = \rho \frac{1}{N + N_{\text{delay}} - 1} \sum_{i=-N_{\text{delay}}+1}^{N-1} \text{E}\{|x^{[m]}[i]|^2\}$ taking the cyclic prefix into account. However, since it is easily verified that SCP-FDKD and SCP-Chirp both guarantee that

$$\frac{1}{N + N_{\text{delay}} - 1} \sum_{i=-N_{\text{delay}}+1}^{N-1} \text{E}\{|x^{[m]}[i]|^2\} = \frac{1}{N} \sum_{i=0}^{N-1} \text{E}\{|x^{[m]}[i]|^2\},$$

we find that $\sigma_{\text{zp}}^2 = \sigma_{\text{cp}}^2$ for these schemes. In other words, SZP-TDKD, SCP-FDKD, and SCP-Chirp all have the same average transmitted power, and thus have the same SNR, which is ρ .

Now, we combine the per-block achievable rates (from Section 4.4) with the per-block time-bandwidth consumption of streaming systems (from Section 4.6) in order to write the achievable rates in units of bits/sec/Hz. As shown in Section 4.6, SZP-TDKD, SCP-FDKD, and SCP-Chirp schemes have same SNR of ρ and can be fairly compared. In addition, the channel matrix definition (4.25) holds for SZP-TDKD as well as SCP schemes because, as evident from Fig. 4.3, the ZP portion of each SZP-TDKD block acts as if it were a CP for the next SZP-TDKD block. Thus, with i.i.d. Gaussian input, upper and lower bounds on the achievable rate of the streaming schemes under consideration are

$$\mathcal{R}_{\text{mmse-ub}} = \zeta \mathcal{R}_{\text{mmse-blk-ub}} \text{ bits/sec/Hz} \quad (4.42)$$

$$\mathcal{R}_{\text{mmse-lb}} = \zeta \mathcal{R}_{\text{mmse-blk-lb}} \text{ bits/sec/Hz} \quad (4.43)$$

respectively, where

$$\zeta = \begin{cases} \frac{1}{N+N_{\text{Dopp}}-1} & \text{SZP-TDKD} \\ \frac{1}{N+N_{\text{delay}}-1} & \text{SCP-FDKD} , \\ \frac{N}{(N+N_{\text{delay}}-1)(N+N_{\text{Dopp}}-1)} & \text{SCP-Chirp} \end{cases} \quad (4.44)$$

and where $\mathcal{R}_{\text{mmse-blk-ub}}$ and $\mathcal{R}_{\text{mmse-blk-lb}}$ were defined in Theorem 3. Using similar arguments, the achievable rate of an SCP system with i.i.d. Gaussian inputs and perfect receiver-CSI (and no pilots) is

$$\mathcal{R}_{\text{csi}} = \frac{N}{(N+N_{\text{delay}}-1)(N+N_{\text{Dopp}}-1)} \mathcal{R}_{\text{csi-blk}} \text{ bits/sec/Hz}. \quad (4.45)$$

The asymptotic achievable rates can then be written as

$$\mathcal{R}_{\text{SZP-TDKD}} = \underbrace{\left(\frac{N - (2N_{\text{delay}} - 1)N_{\text{Dopp}}}{N + N_{\text{Dopp}} - 1} \right)}_{\eta_{\text{SZP-TDKD}}} \log(\rho) + O(1) \text{ bits/sec/Hz} \quad (4.46)$$

$$\mathcal{R}_{\text{SCP-FDKD}} = \underbrace{\left(\frac{N - (2N_{\text{Dopp}} - 1)N_{\text{delay}}}{N + N_{\text{delay}} - 1} \right)}_{\eta_{\text{SCP-FDKD}}} \log(\rho) + O(1) \text{ bits/sec/Hz} \quad (4.47)$$

$$\mathcal{R}_{\text{SCP-Chirp}} = \underbrace{\left(\frac{(N - 2N_{\text{Dopp}}N_{\text{delay}} + 1)N}{(N + N_{\text{Dopp}} - 1)(N + N_{\text{delay}} - 1)} \right)}_{\eta_{\text{SCP-Chirp}}} \log(\rho) + O(1) \text{ bits/sec/Hz}, \quad (4.48)$$

as $\rho \rightarrow \infty$. For integer N_{delay} and N_{Dopp} , it easily follows that $N_{\text{delay}} \geq N_{\text{Dopp}} \Leftrightarrow \eta_{\text{SCP-FDKD}} \geq \eta_{\text{SZP-TDKD}}$ and vice versa. Thus, in the high-SNR regime, SCP-FDKD dominates SZP-TDKD (from [13]) when $N_{\text{delay}} > N_{\text{Dopp}}$, while SZP-TDKD dominates SCP-FDKD when $N_{\text{delay}} < N_{\text{Dopp}}$.

4.7 Numerical Results

In this section, we present numerical examples of the spectral efficiencies of several streaming MMSE-PAT schemes. For this purpose, we evaluate the achievable rate bounds of streaming MMSE-PAT (in units of bits/sec/Hz) for the SZP-TDKD, SCP-FDKD, and SCP-Chirp schemes, using the power allocation procedure described in Section 4.5. In all cases, we consider block size $N = 120$ and plot the bounds over the SNR range of practical interest. As discussed in Section 4.6, ρ is the SNR of all the PAT schemes. We used a Monte-Carlo approach to evaluate the expectation in the bounding expressions (4.27)-(4.28).

In Fig. 4.4, we show results for a channel with i.i.d. BEM coefficients (i.e., $\mathbf{R}_\lambda = \frac{N}{N_{\text{Dopp}}N_{\text{delay}}}\mathbf{I}$) such that $N_{\text{delay}} = N_{\text{Dopp}} = 3$. These discrete spreading parameters correspond to a spreading index of $\gamma \approx 0.07$, as results from, e.g., carrier frequency $f_c = 80$ GHz, sampling interval $\mathcal{T}_s = 1$ μsec , maximum mobile velocity $v_{\text{max}} = 150$ km/hr, and a channel delay spread $\mathcal{T}_{\text{delay}}$ of 3 μsec . These physical channel parameters are related to $\mathcal{B}_{\text{Dopp}}$ via $\mathcal{B}_{\text{Dopp}} = f_c v_{\text{max}}/c$, where c denotes the speed of light. Both the SCP-Chirp schemes yield identical bounds since they have the same spectral efficiency. Also, as can be seen from Fig. 4.4, SZP-TDKD and SCP-FDKD yield identical bounds on achievable rate, which is expected since $N_{\text{delay}} = N_{\text{Dopp}}$. The bounds for SCP-Chirp schemes are uniformly lower than those for SZP-TDKD and SCP-FDKD, which is also expected since $N_{\text{Dopp}} \neq 1 \neq N_{\text{delay}}$. For reference, Fig. 4.4 also plots the performance of an SCP system with perfect receiver CSI (and no pilots) via (4.45).

In Fig. 4.5, we show results for a channel with i.i.d. BEM coefficients such that $N_{\text{delay}} = 15$ and $N_{\text{Dopp}} = 3$. This channel is primarily time-spreading with spreading index $\gamma \approx 0.37$ and results from, e.g., the same physical channel parameters as before, but with a channel delay spread $\mathcal{T}_{\text{delay}}$ of 15 μsec . Note that, since $N_{\text{delay}} > N_{\text{Dopp}}$, the SCP-FDKD bounds dominate SZP-TDKD bounds, which in turn dominate the SCP-Chirp bounds. Compared to Fig. 4.4, there is a much larger gap between the MMSE-PAT bounds and the perfect-receiver-CSI bounds, as a consequence of the higher spreading index.

In Fig. 4.6, we compare SZP-TDKD to SCP-FDKD on two channels with $N_{\text{delay}} = 15$ and $N_{\text{Dopp}} = 3$. One has i.i.d. BEM coefficients (i.e., $\text{E}|\lambda[k; l]|^2 = \frac{N}{N_{\text{Dopp}}N_{\text{delay}}}$ for $k \in \{-\frac{N_{\text{Dopp}}-1}{2}, \dots, \frac{N_{\text{Dopp}}-1}{2}\}$ and $l \in \{0, \dots, N_{\text{delay}}-1\}$), and the other has independent

but non-identically distributed BEM coefficients, as would result from a non-uniform delay profile and a non-uniform Doppler spectrum. In particular, we consider a channel with a “Jakes” Doppler spectrum and an exponential delay profile, for which $\mathbb{E}\{|\lambda[k; l]|^2\} = \chi e^{-0.1l} (\mathcal{B}_{\text{Dopp}}^2 - k^2 (N\mathcal{T}_s)^{-2})^{-0.5}$ for $k \in \{-\frac{N_{\text{Dopp}}-1}{2}, \dots, \frac{N_{\text{Dopp}}-1}{2}\}$ and $l \in \{0, \dots, N_{\text{delay}} - 1\}$, and where χ is chosen such that $\text{tr}\{\mathbf{R}_\lambda\} = N$. For the channel with i.i.d. BEM coefficients, we allocate pilot/data power according to the procedure in Section 4.5, while, for the non-i.i.d. channel, we allocate equal power between pilots and data. In both cases, we see that the achievable rate bounds grow (asymptotically) linearly in $\log \rho$ with slopes proportional to the pre-log factors in the asymptotic rate expressions (4.46)-(4.47). Since, for this channel, $N_{\text{delay}} > N_{\text{Dopp}}$, SCP-FDKD’s higher pre-log factor translates into significant rate gains over SZP-TDKD at high SNR.

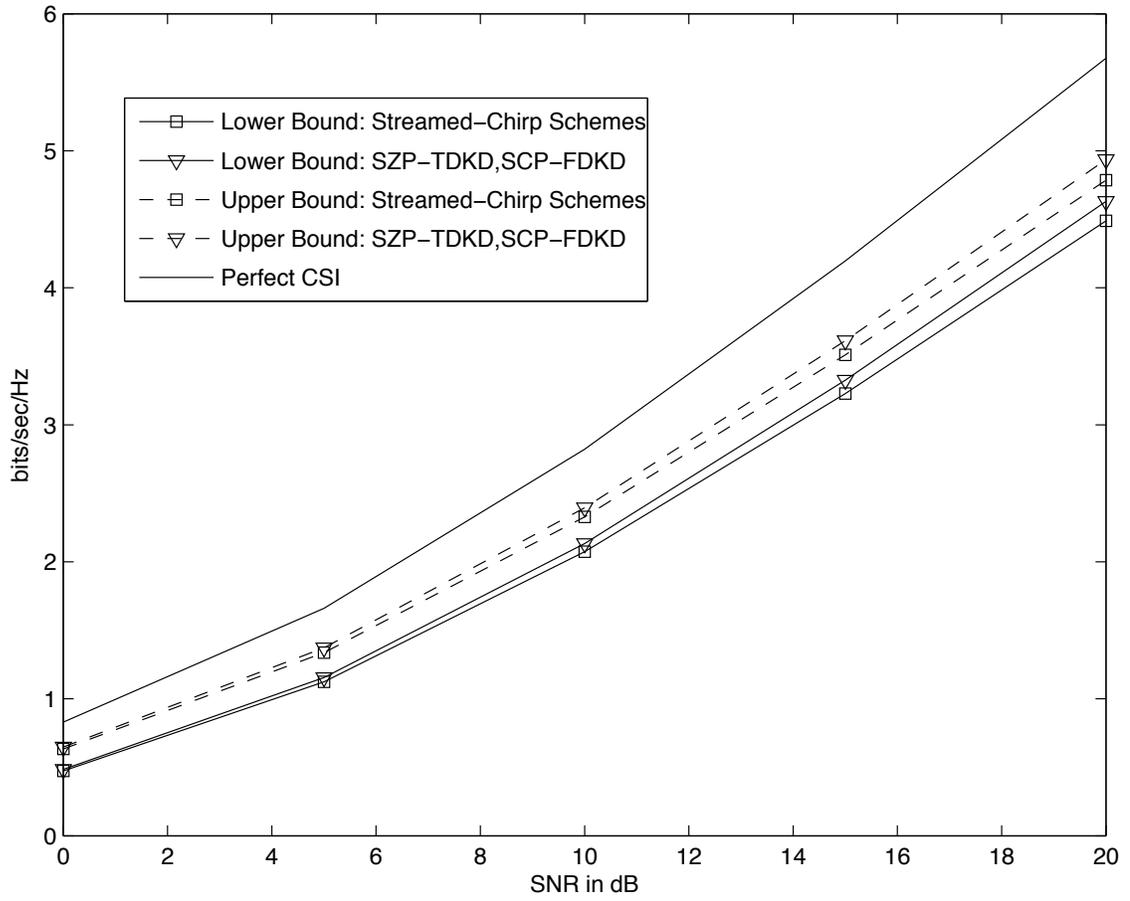


Figure 4.4: Bounds on ergodic achievable rates for $N_{\text{delay}} = N_{\text{Dopp}} = 3$.

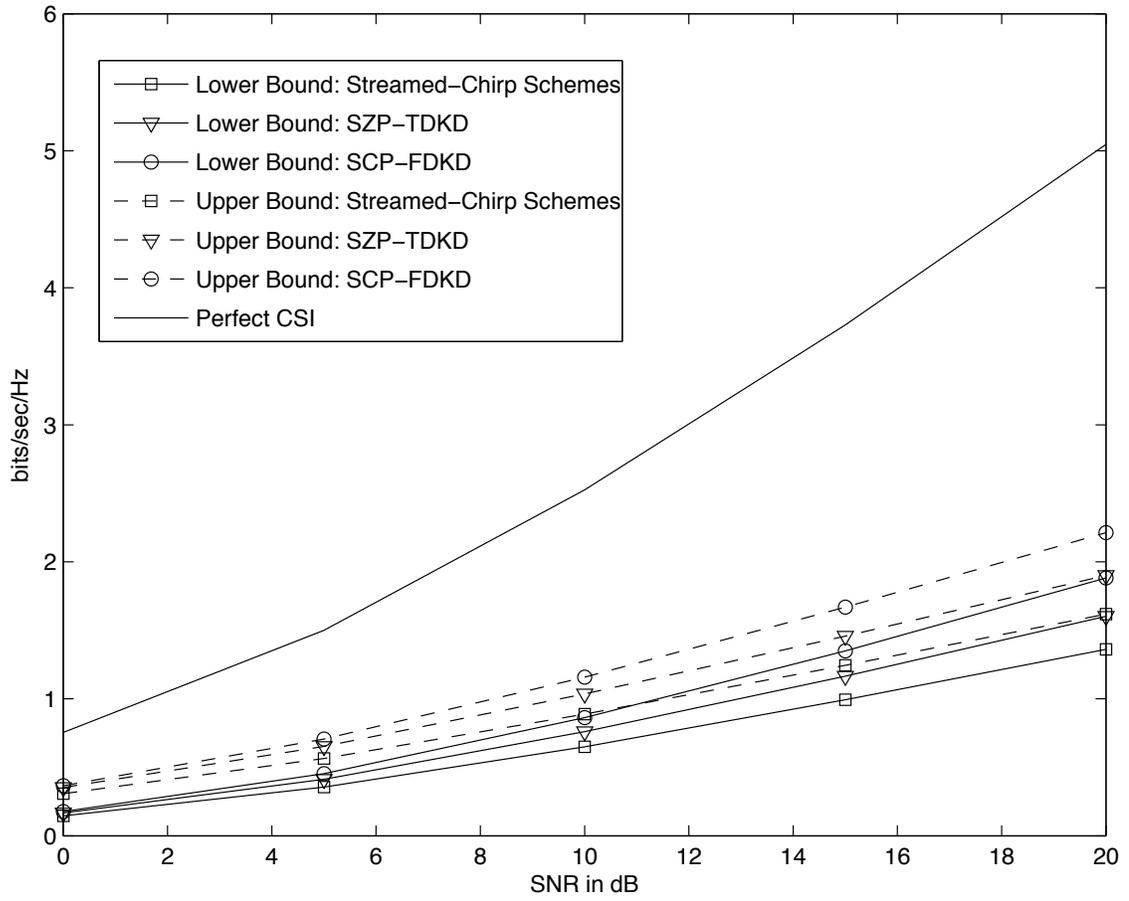


Figure 4.5: Bounds on ergodic achievable rates for $N_{\text{delay}} = 15$ and $N_{\text{Dopp}} = 3$.

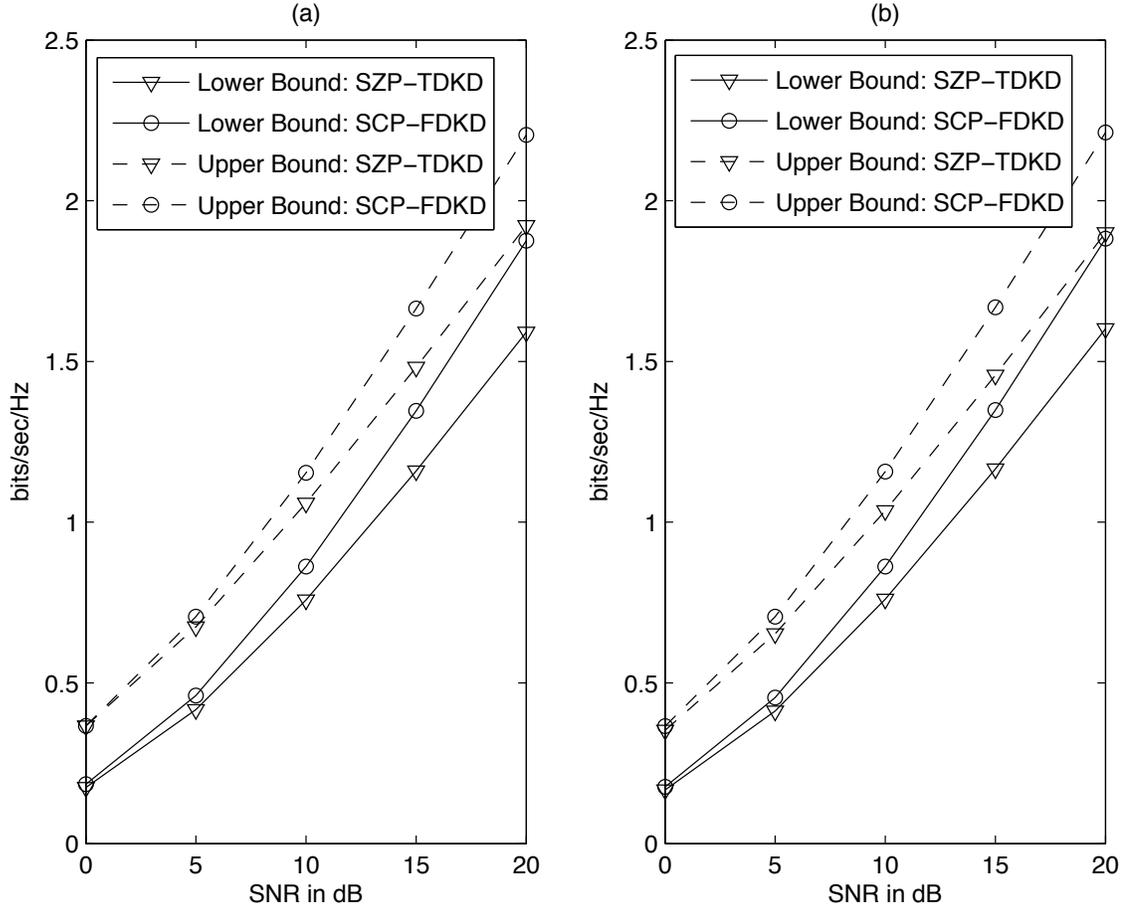


Figure 4.6: Bounds on ergodic achievable rates for $N_{\text{delay}} = 15$ and $N_{\text{Dopp}} = 3$ for (a) non-uniform and (b) uniform decay and Doppler profiles.

CHAPTER 5

SPECTRALLY EFFICIENT PAT

Recall that a rate \mathcal{R} of a PAT scheme is said to be achievable if the probability of decoding error for that rate can be made arbitrarily small. Since our PAT schemes use Gaussian codes, based on Theorem 1, which gives the bound on the pre-log factor of the constrained capacity of noncoherent DSC with continuous inputs, we make the following definition about the PAT schemes.

Definition 1. *A PAT scheme is called spectrally efficient if its per-channel-use achievable rate $\mathcal{R}(\rho)$ over the CE-BEM DSC satisfies $\lim_{\rho \rightarrow \infty} \frac{\mathcal{R}(\rho)}{\log \rho} = \frac{N - N_{\text{Dopp}} N_{\text{delay}}}{N}$.*

For the case of flat or frequency-selective channels, PAT schemes designed to minimize the channel estimation error variance have been shown to be spectrally efficient [7, 20, 41]. We study the MMSE-PAT schemes for DSC and establish that they are spectrally inefficient. We also derive design conditions for spectrally efficient PAT and present a novel spectrally efficient PAT.

5.1 Is any MMSE-PAT Spectrally Efficient?

From the asymptotic rates of MMSE-PAT schemes (Theorem 4), the prelog factor of the per-channel-use achievable rates of TDKD, FDKD and the chirp schemes are

$\frac{N-(2N_{\text{delay}}-1)N_{\text{Dopp}}}{N}$, $\frac{N-(2N_{\text{Dopp}}-1)N_{\text{delay}}}{N}$ and $\frac{N-2N_{\text{delay}}N_{\text{Dopp}}}{N}$ respectively. Note that, for singly selective channels, some MMSE-PAT schemes are spectrally efficient. For example, for time selective channels ($N_{\text{delay}} = 1$), the TDKD scheme is spectrally efficient. Similarly, for frequency selective channels ($N_{\text{Dopp}} = 1$), the FDKD scheme is spectrally efficient. For strictly doubly selective channels (i.e., $N_{\text{delay}} > 1$ and $N_{\text{Dopp}} > 1$), the four MMSE-PAT Examples from Section 4.3 yield $N_s < N - N_{\text{Dopp}}N_{\text{delay}}$, and are clearly not spectrally efficient. The ZP-MMSE-PAT scheme from [13] which resembles TDKD is also spectrally inefficient. But does there exist some other MMSE-PAT scheme which is spectrally efficient over the CE-BEM DSC? The answer is given in the following theorem.

Theorem 5. *In CE-BEM DSC with $N_{\text{delay}} > 1$ and $N_{\text{Dopp}} > 1$, the data dimension N_s of any CP-MMSE-PAT scheme (\mathbf{p}, \mathbf{B}) satisfying the necessary requirements given in Lemma 2 is strictly bounded as $N_s < N - N_{\text{Dopp}}N_{\text{delay}}$.*

Proof. See Appendix D.1. □

Combining the achievable rates of MMSE-PAT from Theorem 4 and the result from Theorem 5, we have the following result.

Corollary 1 (Spectral Inefficiency). *No CP-MMSE-PAT scheme is spectrally efficient over the strict CE-BEM DSC with $N_{\text{delay}} > 1$ and $N_{\text{Dopp}} > 1$.*

Our results show that there are some fundamental differences between singly selective channels and doubly selective channels. The reason for spectral inefficiency of MMSE-PAT schemes over DSC can be intuitively explained as follows. In the case of the singly selective (time selective or frequency selective) channels, the “effective” channel matrix $\hat{\mathbf{H}}$ for the PAT schemes (recall (2.27)) has deterministic eigen vectors

and the unknown channel coefficients correspond to the eigen values. This property allows the “optimal” estimation of the unknown channel parameters (in the MSE sense) by sacrificing the signaling dimensions whose number equals the number of unknown channel parameters. But channel matrix corresponding to DSC does not have deterministic eigen vectors and the optimal estimation of the channel coefficients takes away more signaling dimensions than the number of unknown channel coefficients.

Note on CP overhead: In calculating the spectral efficiency of MMSE-PAT schemes in units bits per channel-use, we ignored the overhead due to CP. If the delay spread N_{delay} is small compared to the block size N , the fraction of CP overhead $\frac{N_{\text{delay}}-1}{N}$ becomes negligible. Even if $N_{\text{delay}} \ll N$, the MMSE-PAT schemes from Section 4.3 can be highly inefficient in terms of achievable rates, if $N_{\text{Dopp}}N_{\text{delay}}$ is comparable to N .

5.2 Design of Spectrally Efficient PAT

We proceed towards the design of spectrally efficient PAT. First, we note that for our decoder in Section 2.3.2, if $E_p = 0$, then the LMMSE estimate of the channel from pilots is $\hat{\mathbf{H}} = 0$. In that case, the mismatched minimum distance decoder in (2.31) can not differentiate between any two distinct codewords and the achievable rate is 0. Since this case is not interesting, we restrict ourselves to the case of non-zero pilot energy E_p , specifically $\liminf_{\rho \rightarrow \infty} E_p > 0$. Now, recalling (2.29), the output vector is represented as,

$$\mathbf{y} = \sqrt{\rho} \mathbf{P} \mathbf{U} \boldsymbol{\lambda} + \sqrt{\rho} \mathbf{D} \mathbf{U} \boldsymbol{\lambda} + \mathbf{v}. \quad (5.1)$$

Since we are considering non-data-aided estimators, the pilot-data orthogonality at the channel output is desirable; otherwise the channel estimation will suffer interference from data and the channel estimates will not be perfect even in the absence of noise (i.e., asymptotically as $\rho \rightarrow \infty$). Using pilot data orthogonality criterion, we introduce the notion of “linearly separable” PAT.

Definition 2. A PAT scheme is said to be linearly separable if the pilots and data are orthogonal so that $(\mathbf{P}\mathbf{U})^H \mathbf{D}\mathbf{U} = \mathbf{0}$, $\forall \mathbf{D}$.

Recall from Theorem 2 that all the MMSE-PAT schemes are linearly separable since they satisfy the pilot-data orthogonality. Now, we present the achievable rates of linearly separable PAT schemes with Gaussian codes of covariance \mathbf{R}_s and the mismatched decoder specified in Section 2.3.2. To do so, we specify the weighting matrix \mathbf{Q} and use a result from [30]. To obtain the weighting matrix, first we “separate” the pilots and data at the observation. Recalling (2.27), we have

$$\mathbf{y} = \sqrt{\rho} \dot{\mathbf{H}}(\mathbf{p} + \mathbf{B}\mathbf{s}) + \mathbf{v}. \quad (5.2)$$

Let the columns of \mathbf{B}_d form an orthonormal basis for the left null space of $\mathbf{P}\mathbf{U}$. Due to pilot-data orthogonality, the projection

$$\mathbf{y}_d = \mathbf{B}_d^H \mathbf{y} \quad (5.3)$$

$$= \underbrace{\mathbf{B}_d^H \dot{\mathbf{H}} \mathbf{B}}_{\mathbf{H}_d} \mathbf{s} + \underbrace{\mathbf{B}_d^H \mathbf{v}}_{\mathbf{v}_d}, \quad (5.4)$$

does not result in any loss of data energy. Splitting \mathbf{H}_d into estimate $\hat{\mathbf{H}}_d$ and error $\tilde{\mathbf{H}}_d$ components, we have

$$\mathbf{y}_d = \hat{\mathbf{H}}_d \mathbf{s} + \underbrace{\tilde{\mathbf{H}}_d \mathbf{s}}_{\mathbf{n}} + \mathbf{v}_d. \quad (5.5)$$

From [30], denoting $\mathbf{R}_n = \text{E}\{\mathbf{n}\mathbf{n}^H\}$, the optimal weighting factor to be applied to \mathbf{y}_d in order to maximize the achievable rates is $\mathbf{R}_n^{-1/2}$, which can be thought of as “whitening” filter. So, we apply the weighting factor

$$\mathbf{Q} = \mathbf{R}_n^{-1/2} \mathbf{B}_d^H, \quad (5.6)$$

in the decoder (2.31). Now, let the columns of \mathbf{B}_p form an orthonormal basis for the column space of $\mathbf{P}\mathbf{U}$ and consider the projection

$$\mathbf{y}_p = \mathbf{B}_p^H \mathbf{y} \quad (5.7)$$

$$= \sqrt{\rho} \mathbf{B}_p^H \mathbf{P}\mathbf{U} \boldsymbol{\lambda} + \underbrace{\mathbf{B}_p^H \mathbf{v}}_{\mathbf{v}_p}, \quad (5.8)$$

where (5.8) follows from the linear separability. Since the projection \mathbf{y}_p does not result in any loss of pilot energy, LMMSE estimate of the channel from \mathbf{y} and \mathbf{y}_p are same. Now, since \mathbf{y}_p and $\boldsymbol{\lambda}$ are jointly Gaussian LMMSE estimate coincide with the MMSE estimate. Because of this, our PAT decoder in Section 2.3.2 falls within the hypothesis of [30] and we can use the result from [30] to find the achievable rates of our PAT schemes. The result is given in the following lemma.

Lemma 5. *For a linearly separable PAT scheme, with the weighting factor in (5.6) applied in the decoder (2.31), the per-block achievable rates are given by*

$$\mathcal{R}_{\text{ls-blk}} = \text{E}\{\log \det[\mathbf{I} + \rho \mathbf{R}_n^{-1} \hat{\mathbf{H}}_d \mathbf{R}_s \hat{\mathbf{H}}_d^H]\} \text{ bits/block}. \quad (5.9)$$

The above rate expression resembles that of coherent case [33] with \mathbf{n} acting as “effective” Gaussian noise. Now, note that there are $N_{\text{Dopp}} N_{\text{delay}}$ independent unknown BEM coefficients in each N -length block. All the CP-MMSE-PAT schemes are shown to sacrifice more than $N_{\text{Dopp}} N_{\text{delay}}$ dimensions for pilots, leaving $N_s < N - N_{\text{Dopp}} N_{\text{delay}}$

dimensions for data symbols, and hence are spectrally inefficient. To design a spectrally efficient PAT, we need to relax the MMSE requirements in Lemma 2. As noted before, since we are considering non-data-aided estimators, the pilot-data orthogonality at the channel output (4.7) is desirable. Thus, a PAT scheme which preserves pilot-data orthogonality at the channel output, and which can yield perfect channel estimates in the absence of noise using only $N_{\text{Dopp}}N_{\text{delay}}$ pilot dimensions, is a candidate for spectrally efficient PAT. In the following, we establish the sufficient conditions for a PAT scheme to be spectrally efficient.

Theorem 6. *Suppose a PAT scheme parameterized by pilot vector \mathbf{p} and data modulation matrix \mathbf{B} satisfies following conditions;*

1. \mathbf{PU} is full rank.
2. $\text{rank}(\mathbf{B}) = N - N_{\text{Dopp}}N_{\text{delay}}$.
3. It is linearly separable.

Then the PAT scheme is spectrally efficient.

Proof. See Appendix D.2. □

First condition gives the requirement for the pilot vector. Second condition stipulates the requirement of the precoder matrix. Third condition gives the joint requirement between the pilot vector and precoder matrix. A PAT scheme satisfying the above requirements is given below.

Example 5 (SE-PAT). *With the pilot index set $\mathcal{P}_s = \{0, N_{\text{delay}}, \dots, (N_{\text{Dopp}} - 1)N_{\text{delay}}\}$ and the guard index set $\mathcal{G}_s = \{0, \dots, N_{\text{Dopp}}N_{\text{delay}} - 1\}$, an N -block ZP PAT scheme for*

the CE-BEM DSC is given by

$$p[k] = \begin{cases} \sqrt{\frac{E_p}{N_{\text{Dopp}}}} e^{j\theta[k]} & k \in \mathcal{P}_s \\ 0 & k \notin \mathcal{P}_s \end{cases} \quad (5.10)$$

and by \mathbf{B} constructed from the columns of \mathbf{I}_N with indices not in the set \mathcal{G}_s . $\theta(k) \in \mathbb{R}$ is arbitrary.

In the above PAT, the first $N_{\text{Dopp}}N_{\text{delay}}$ time slots are used by the pilots and the remaining time slots are used for data transmission ensuring linear separability. The structure of the PAT is illustrated in Fig. 5.1. Also the rank of the precoding matrix is $N - N_{\text{Dopp}}N_{\text{delay}}$. It can also be easily verified that \mathbf{PU} is full rank.

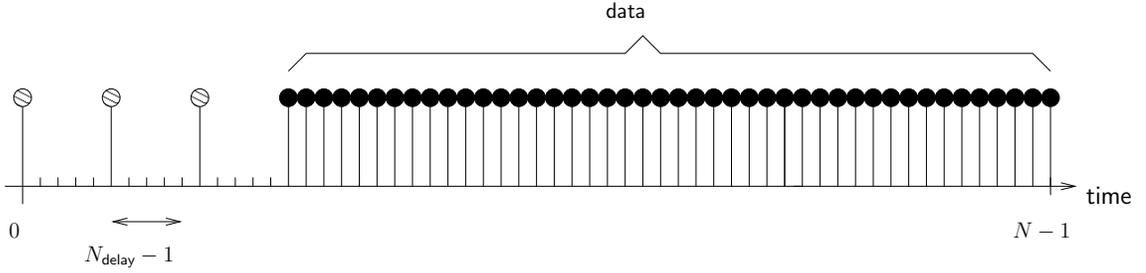


Figure 5.1: Structure of SE-PAT with $N_{\text{delay}} = 5$ and $N_{\text{Dopp}} = 3$.

5.3 Comparison of PAT schemes

In this section, we compare the achievable rates of MMSE-PAT schemes from Section 4.3 and the SE-PAT scheme from the previous section. We numerically evaluate the achievable rates and also perform some approximate theoretical computations for comparison. First, we numerically evaluate the achievable rates of SE-PAT and MMSE-PAT schemes with i.i.d. Gaussian codes and equal energy allocation between

pilots and data. With $\sigma_s^2 = \frac{E_s}{N_s}$, the per-channel-use achievable rates of SE-PAT and MMSE-PAT schemes with i.i.d. Gaussian codes are given by (Lemma 5)

$$\mathcal{R} = \frac{1}{N} \mathbb{E}\{\log \det[\mathbf{I} + \rho\sigma_s^2 \mathbf{R}_n^{-1} \hat{\mathbf{H}}_d \hat{\mathbf{H}}_d^H]\}, \quad (5.11)$$

where $\hat{\mathbf{H}}_d = \mathbf{B}_d^H \hat{\mathbf{H}} \mathbf{B}$ and \mathbf{R}_n is the covariance of $\mathbf{B}_d^H (\sqrt{\rho} \tilde{\mathbf{H}} \mathbf{B} \mathbf{s} + \mathbf{v})$. Since $\mathbf{B} \mathbf{B}^H \leq \mathbf{I}$, we have,

$$\mathbf{R}_n = \rho\sigma_s^2 \mathbf{B}_d^H \mathbb{E} \left\{ \tilde{\mathbf{H}} \mathbf{B} \mathbf{B}^H \tilde{\mathbf{H}}^H \right\} \mathbf{B}_d + \mathbf{I} \quad (5.12)$$

$$\leq \rho\sigma_s^2 \mathbf{B}_d^H \mathbb{E} \left\{ \tilde{\mathbf{H}} \tilde{\mathbf{H}}^H \right\} \mathbf{B}_d + \mathbf{I}. \quad (5.13)$$

For ease of computation, we evaluate the lower bound on the achievable rates (5.11) by replacing \mathbf{R}_n with its upper bound (5.13). In fact, for MMSE-PAT schemes, this lower bound on achievable rates coincides with the lower bound in Theorem 3. For different values of N , N_{delay} and N_{Dopp} , with identically distributed BEM coefficients, the plots of lower bound on the achievable rates (in bits per channel-use) versus SNR are shown in Figures 5.2-5.4. For comparison, we also plot the rates of coherent case of perfect receiver CSI apriori, with i.i.d. Gaussian codes. Since the asymptotic achievable rates of SE-PAT scheme has higher pre-log factor than that of MMSE-PAT schemes, we see that the rates of SE-PAT scheme grow faster with SNR than the MMSE-PAT schemes, in the high SNR regime.

5.3.1 Effective SNR

In the decoder, since $\hat{\mathbf{H}}_d \mathbf{s}$ corresponds to the signal component and since $\mathbf{n} = \tilde{\mathbf{H}}_d \mathbf{s} + \mathbf{v}_d$ acts as the effective noise (from Lemma 5), we define the effective SNR of the PAT scheme as

$$\rho_{\text{eff}} = \frac{\rho\sigma_s^2 \text{tr}\{\mathbf{R}_{\hat{h}}/N\}}{\rho\sigma_s^2 \text{tr}\{\mathbf{R}_{\tilde{h}}/N\} + 1}.$$

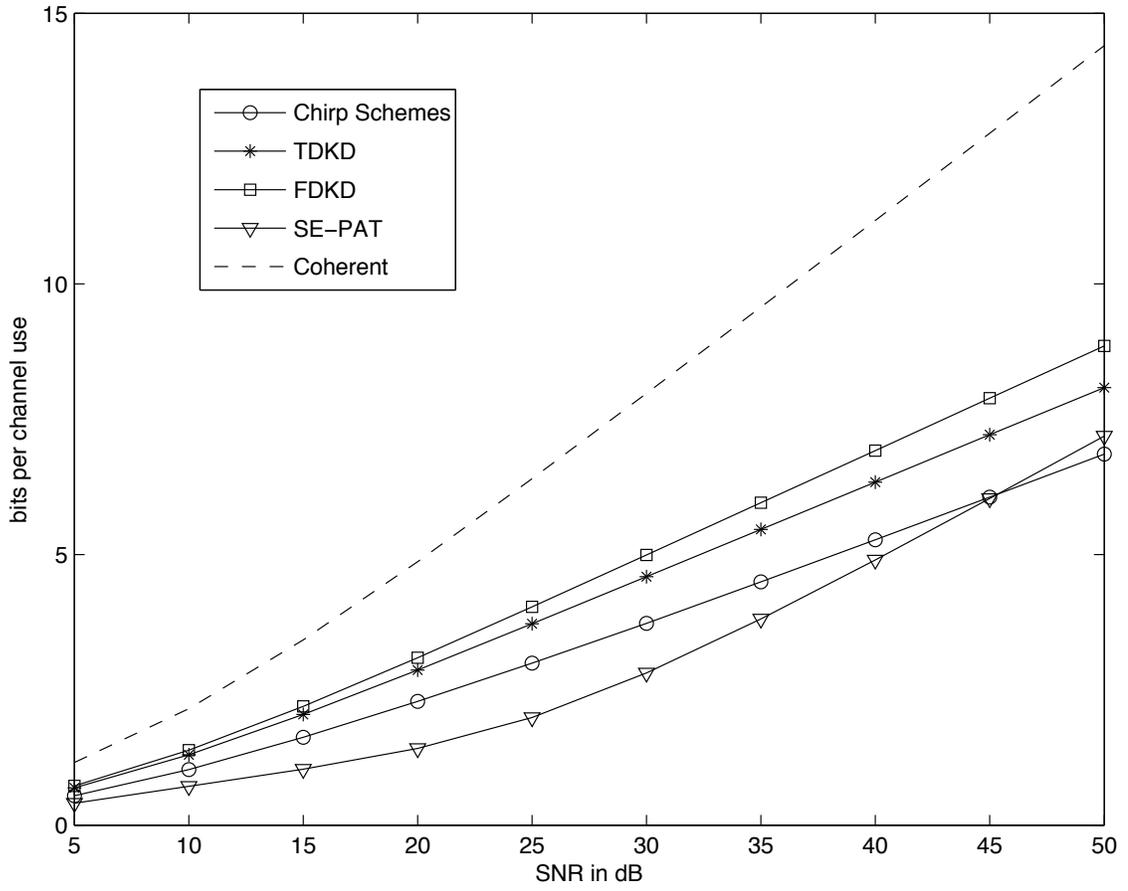


Figure 5.2: Lower bound on achievable rates for $N = 108$, $N_{\text{delay}} = 9$, $N_{\text{Dopp}} = 3$.

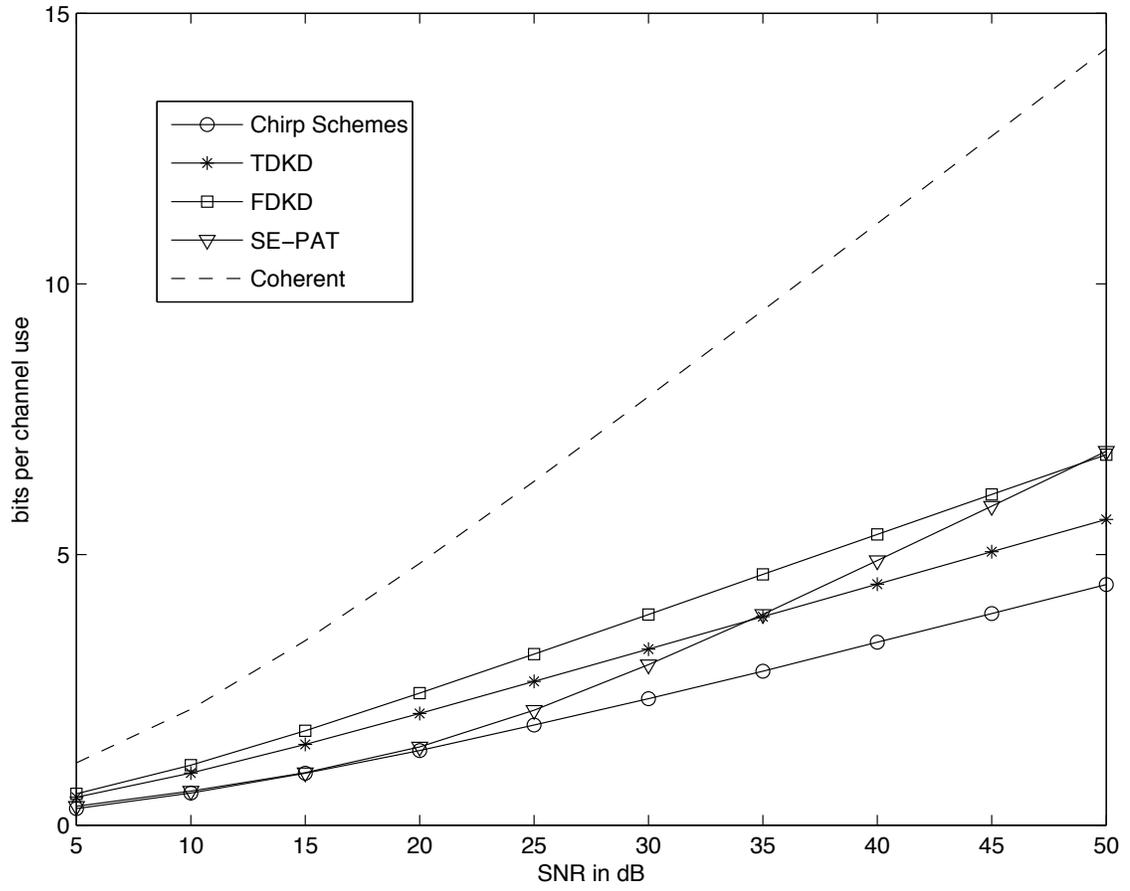


Figure 5.3: Lower bound on achievable rates for $N = 108$, $N_{\text{delay}} = 12$, $N_{\text{Dopp}} = 3$.

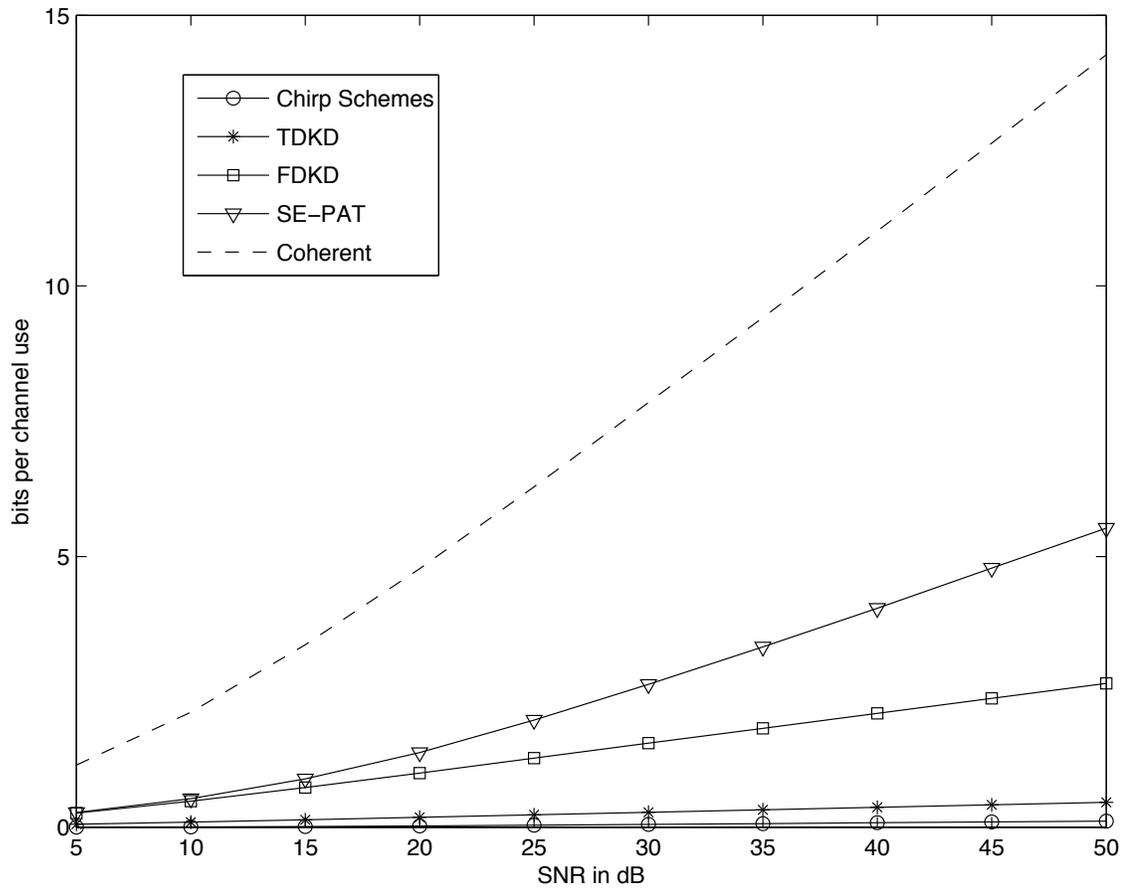


Figure 5.4: Lower bound on achievable rates for $N = 108$, $N_{\text{delay}} = 18$, $N_{\text{Dopp}} = 3$.

We will see that the effective SNR is an important parameter of the PAT schemes which gives insights on their achievable rates. For MMSE-PAT schemes, the covariance of the channel estimate and the estimation error are given by (4.10) and (4.9) respectively. For the spectrally efficient (SE) PAT, the pilot observations can be written as

$$\mathbf{y}_p = \sqrt{\frac{\rho E_p}{N}} \mathbf{G} \boldsymbol{\lambda} + \mathbf{v}_p, \quad (5.14)$$

for a suitably constructed \mathbf{G} with unit-norm rows. From the structure of the SE-PAT scheme, it follows that \mathbf{G} is full rank. Now, the error covariance can be written as

$$\mathbf{R}_h^{\text{se}} = \mathbf{U} \left(\mathbf{R}_\lambda^{-1} + \frac{\rho E_p}{N} \mathbf{G}^H \mathbf{G} \right)^{-1} \mathbf{U}^H, \quad (5.15)$$

$$\mathbf{R}_h^{\text{se}} = \mathbf{U} \left[\mathbf{R}_\lambda - \left(\mathbf{R}_\lambda^{-1} + \frac{\rho E_p}{N} \mathbf{G}^H \mathbf{G} \right)^{-1} \right] \mathbf{U}^H. \quad (5.16)$$

Denoting the minimum eigen value of $\mathbf{G}^H \mathbf{G}$ by μ_{\min} , we obtain the following bounds on the covariance matrices,

$$\mathbf{R}_h^{\text{se}} \leq \mathbf{U} \left(\mathbf{R}_\lambda^{-1} + \frac{\rho E_p \mu_{\min}}{N} \mathbf{I}_{N_{\text{Dopp}} N_{\text{delay}}} \right)^{-1} \mathbf{U}^H = \mathbf{R}_h^{\text{se-ub}}, \quad (5.17)$$

$$\mathbf{R}_h^{\text{se}} \geq \mathbf{U} \left[\mathbf{R}_\lambda - \left(\mathbf{R}_\lambda^{-1} + \frac{\rho E_p \mu_{\min}}{N} \mathbf{I}_{N_{\text{Dopp}} N_{\text{delay}}} \right)^{-1} \right] \mathbf{U}^H = \mathbf{R}_h^{\text{se-lb}}. \quad (5.18)$$

Since \mathbf{G} is full rank and $\text{tr}\{\mathbf{G}^H \mathbf{G}\} = N_{\text{Dopp}} N_{\text{delay}}$, the minimum eigen value is bounded as

$$0 < \mu_{\min} \leq 1. \quad (5.19)$$

For the SE-PAT schemes, using (5.17), we have the following bound on the effective SNR,

$$\rho_{\text{eff}}^{\text{se}} \geq \rho_{\text{eff}}^{\text{se-lb}} = \frac{\rho \sigma_s^2 \text{tr}\{\mathbf{R}_h/N\}}{\rho \sigma_s^2 \text{tr}\{\mathbf{R}_h^{\text{se-ub}}/N\} + 1},$$

which will be used in our theoretical computations. If the effective SNR of a PAT scheme is high, i.e., $\rho_{\text{eff}} \gg 1$, then it operates in degrees of freedom (DOF) limited regime, i.e., the achievable rates are limited by the available DOF [3]. If the effective SNR is small, i.e., $\rho_{\text{eff}} \ll 1$, then it operates in the noise limited regime. Now, we make some approximate calculations to find the cutoff SNR value for the PAT schemes at which they switch from the noise limited regime to the DOF limited regime. In the DOF limited regime, SNR is high, i.e., $\rho \gg 1$, since the effective SNR (which is limited by both the estimation error and the additive noise) is high. For the case of MMSE-PAT schemes, with $\rho \gg 1$, using Taylor approximation, we have

$$\text{tr}\{\mathbf{R}_{\tilde{h}}\} = \text{tr}\left(\mathbf{R}_{\lambda}^{-1} + \frac{\rho E_p}{N} \mathbf{I}\right)^{-1} \quad (5.20)$$

$$= \sum_{i=0}^{N_{\text{Dopp}} N_{\text{delay}} - 1} \frac{1}{[\mathbf{R}_{\lambda}]_{i,i}^{-1} + \frac{\rho E_p}{N}} \quad (5.21)$$

$$\approx \frac{N}{\rho E_p} \sum_{i=0}^{N_{\text{Dopp}} N_{\text{delay}} - 1} 1 - \frac{N}{\rho E_p [\mathbf{R}_{\lambda}]_{i,i}} \quad (5.22)$$

$$\approx \frac{N N_{\text{Dopp}} N_{\text{delay}}}{\rho E_p}, \quad (5.23)$$

and hence $\text{tr}\{\mathbf{R}_{\tilde{h}}\} = N(1 - \frac{N_{\text{Dopp}} N_{\text{delay}}}{\rho E_p}) \approx N$. Using these approximations, defining $\sigma_p^2 = \frac{E_p}{N_{\text{Dopp}} N_{\text{delay}}}$, we have

$$\rho_{\text{eff}}^{\text{mmse}} \approx \frac{\rho \sigma_s^2 \sigma_p^2}{\sigma_s^2 + \sigma_p^2}. \quad (5.24)$$

Similar calculations for SE-PAT schemes yield,

$$\rho_{\text{eff}}^{\text{se-lb}} \approx \frac{\rho \mu_{\min} \sigma_e^2 \sigma_p^2}{\sigma_s^2 + \mu_{\min} \sigma_p^2}. \quad (5.25)$$

We define the SNR-cutoff for the MMSE-PAT schemes as

$$\rho_{\text{cut-off}}^{\text{mmse}} = \frac{\sigma_s^2 + \sigma_p^2}{\sigma_s^2 \sigma_p^2},$$

while for SE-PAT schemes we define

$$\rho_{\text{cut-off}}^{\text{se}} = \frac{\sigma_s^2 + \mu_{\min}\sigma_p^2}{\mu_{\min}\sigma_s^2\sigma_p^2}. \quad (5.26)$$

Beyond this cutoff, the achievable rates of PAT schemes are not limited by effective noise but are limited by the available DOF. For the SE-PAT scheme, in Fig. 5.5, we plot the rates versus SNR curves for the two cases: (a) $N = 135$, $N_{\text{delay}} = 9$ and $N_{\text{Dopp}} = 3$, (b) $N = 135$, $N_{\text{delay}} = 15$ and $N_{\text{Dopp}} = 3$. For these cases, theoretically calculated $\rho_{\text{cut-off}}^{\text{se}}$ using (5.25) is 18 dB and 25 dB respectively. From Fig. 5.5, we see that, after the cutoff SNR, the performance is limited by the available DOF, and the rates grow linearly with SNR with a slope proportional to N_s .

5.3.2 Intersection Point

Since SE-PAT scheme has higher estimation error, its effective SNR might be smaller than the MMSE-PAT schemes. So, in the moderate SNR regime, MMSE-PAT schemes may support higher rates than the SE-PAT scheme. (See Figures 5.2, 5.3.) But as ρ increases beyond $\rho_{\text{cut-off}}^{\text{se}}$, rates of SE-PAT increases with higher slope and eventually beats the rates of MMSE-PAT schemes. (See Figures 5.2, 5.3.) We now make some rough calculations on the SNR value for which the rates of the SE-PAT intersects with that of the MMSE-PAT schemes. These SNR values provide the thresholds for choosing between the SE-PAT and MMSE-PAT schemes. For convenience, we define the normalized channel estimates $\bar{\mathbf{H}}_d = \hat{\mathbf{H}}_d\sqrt{N/\text{tr}\{\mathbf{R}_h\}}$, and the normalized input covariance $\bar{\mathbf{R}}_s = \mathbf{R}_s/\sigma_s^2$. Now, the per-block achievable rates of MMSE-PAT schemes and SE-PAT schemes are bounded as,

$$\mathcal{R}_{\text{mmse-blk}} \geq \mathcal{R}_{\text{mmse-blk-lb}} = E\{\log \det[\mathbf{I} + \rho_{\text{eff}}^{\text{mmse}} \bar{\mathbf{H}}_d^{\text{H}} \bar{\mathbf{R}}_s \bar{\mathbf{H}}_d]\}, \quad (5.27)$$

$$\mathcal{R}_{\text{se-blk}} \geq \mathcal{R}_{\text{se-blk-lb}} = E\{\log \det[\mathbf{I} + \rho_{\text{eff}}^{\text{se-lb}} \bar{\mathbf{H}}_d^{\text{H}} \bar{\mathbf{R}}_s \bar{\mathbf{H}}_d]\}. \quad (5.28)$$

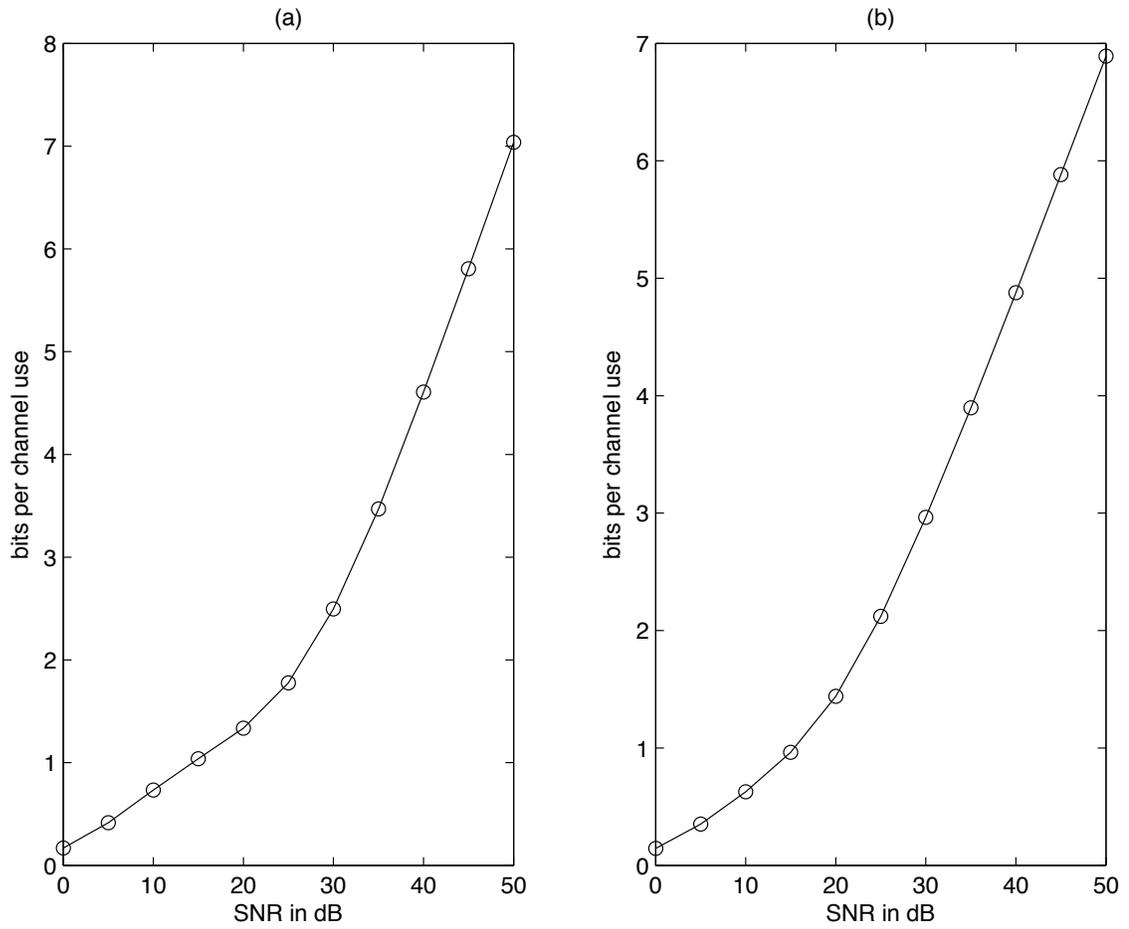


Figure 5.5: Achievable rates of SE-PAT: (a) $N = 135$, $N_{\text{delay}} = 9$ and $N_{\text{Dopp}} = 3$, (b) $N = 135$, $N_{\text{delay}} = 15$ and $N_{\text{Dopp}} = 3$.

We wish to compare the rates of an SE-PAT scheme with data dimension N_s^{se} and SNR cut-off $\rho_{\text{cut-off}}^{\text{se}}$ with the rates of MMSE-PAT scheme with data dimension N_s^{mmse} and cut-off SNR $\rho_{\text{cut-off}}^{\text{mmse}}$. We approximately compute SNR value at which the rates of those two schemes intersect. Specifically, we compute the SNR at which the $\mathcal{R}_{\text{se-blk-lb}} - \mathcal{R}_{\text{mmse-blk-lb}} \approx 0$. Let $\{\alpha_i\}_0^{N_s^{\text{se}}-1}$ and $\{\beta_i\}_0^{N_s^{\text{mmse}}-1}$ denote the non-zero eigen values of the normalized effective channel $\bar{\mathbf{H}}_d \bar{\mathbf{R}}_s \bar{\mathbf{H}}_d^H$ for MMSE-PAT and SE-PAT schemes respectively. Now,

$$\mathcal{R}_{\text{mmse-blk-lb}} = \mathbb{E}\left\{ \sum_{i=0}^{N_s^{\text{mmse}}} \log(1 + \rho_{\text{eff}}^{\text{mmse}} \alpha_i) \right\}, \quad (5.29)$$

$$\mathcal{R}_{\text{se-blk-lb}} = \mathbb{E}\left\{ \sum_{i=0}^{N_s^{\text{se}}} \log(1 + \rho_{\text{eff}}^{\text{se-lb}} \beta_i) \right\}, \quad (5.30)$$

and using the high SNR approximations of $\rho_{\text{eff}}^{\text{mmse}}$ and $\rho_{\text{eff}}^{\text{se-lb}}$, we have

$$\mathcal{R}_{\text{mmse-blk-lb}} \approx \mathbb{E}\left\{ \sum_{i=0}^{N_s^{\text{mmse}}} \log\left(1 + \frac{\rho}{\rho_{\text{cut-off}}^{\text{mmse}}} \alpha_i\right) \right\}, \quad (5.31)$$

$$\mathcal{R}_{\text{se-blk-lb}} \approx \mathbb{E}\left\{ \sum_{i=0}^{N_s^{\text{se}}} \log\left(1 + \frac{\rho}{\rho_{\text{cut-off}}^{\text{se}}} \beta_i\right) \right\}. \quad (5.32)$$

Now,

$$\begin{aligned} \mathcal{R}_{\text{se-blk-lb}} - \mathcal{R}_{\text{mmse-blk-lb}} &\approx N_s^{\text{se}} \log\left(\frac{\rho}{\rho_{\text{cut-off}}^{\text{se}}}\right) - N_s^{\text{mmse}} \log\left(\frac{\rho}{\rho_{\text{cut-off}}^{\text{mmse}}}\right) \\ &\quad + \mathbb{E}\left\{ \sum_i \log\left(\frac{\rho_{\text{cut-off}}^{\text{mmse}}}{\rho} + \beta_i\right) - \log\left(\frac{\rho_{\text{cut-off}}^{\text{se}}}{\rho} + \alpha_i\right) \right\} \\ &\approx N_s^{\text{se}} \log\left(\frac{\rho}{\rho_{\text{cut-off}}^{\text{se}}}\right) - N_s^{\text{mmse}} \log\left(\frac{\rho}{\rho_{\text{cut-off}}^{\text{mmse}}}\right) \end{aligned}$$

and hence $\mathcal{R}_{\text{se-blk-lb}} - \mathcal{R}_{\text{mmse-blk-lb}} \approx 0$ implies $N_s^{\text{se}} \log\left(\frac{\rho}{\rho_{\text{cut-off}}^{\text{se}}}\right) \approx N_s^{\text{mmse}} \log\left(\frac{\rho}{\rho_{\text{cut-off}}^{\text{mmse}}}\right)$. So,

we compute the point of the intersection of the rates as

$$\rho_{\text{int}} = \frac{N_s^{\text{se}} \log \rho_{\text{cut-off}}^{\text{se}} - N_s^{\text{mmse}} \log \rho_{\text{cut-off}}^{\text{mmse}}}{N_s^{\text{se}} - N_s^{\text{mmse}}}. \quad (5.33)$$

Note that, in calculating the intersection points, we made approximations that $\rho_{\text{eff}} \approx \frac{\rho}{\rho_{\text{cut-off}}}$ which is accurate when $\rho \gg 1$. So, our intersection point SNR will be accurate when the intersection of the rates occur in the high SNR regime. In Fig. 5.6, we plot the rates of SE-PAT and FDKD scheme for the parameters $N = 126$, $N_{\text{delay}} = 14$ and $N_{\text{Dopp}} = 3$. For this case, ρ_{int} calculated from (5.33) is approximately 51 dB. From Fig. 5.6, we see that our theoretical intersection point is close to the point obtained through simulations. For the parameters $N = 96$, $N_{\text{delay}} = 12$ and $N_{\text{Dopp}} = 3$, the achievable rates of SE-PAT and FDKD are plotted in Fig. 5.7 and the theoretically calculated ρ_{int} , which is approximately 39 dB, is close to the value from simulations.

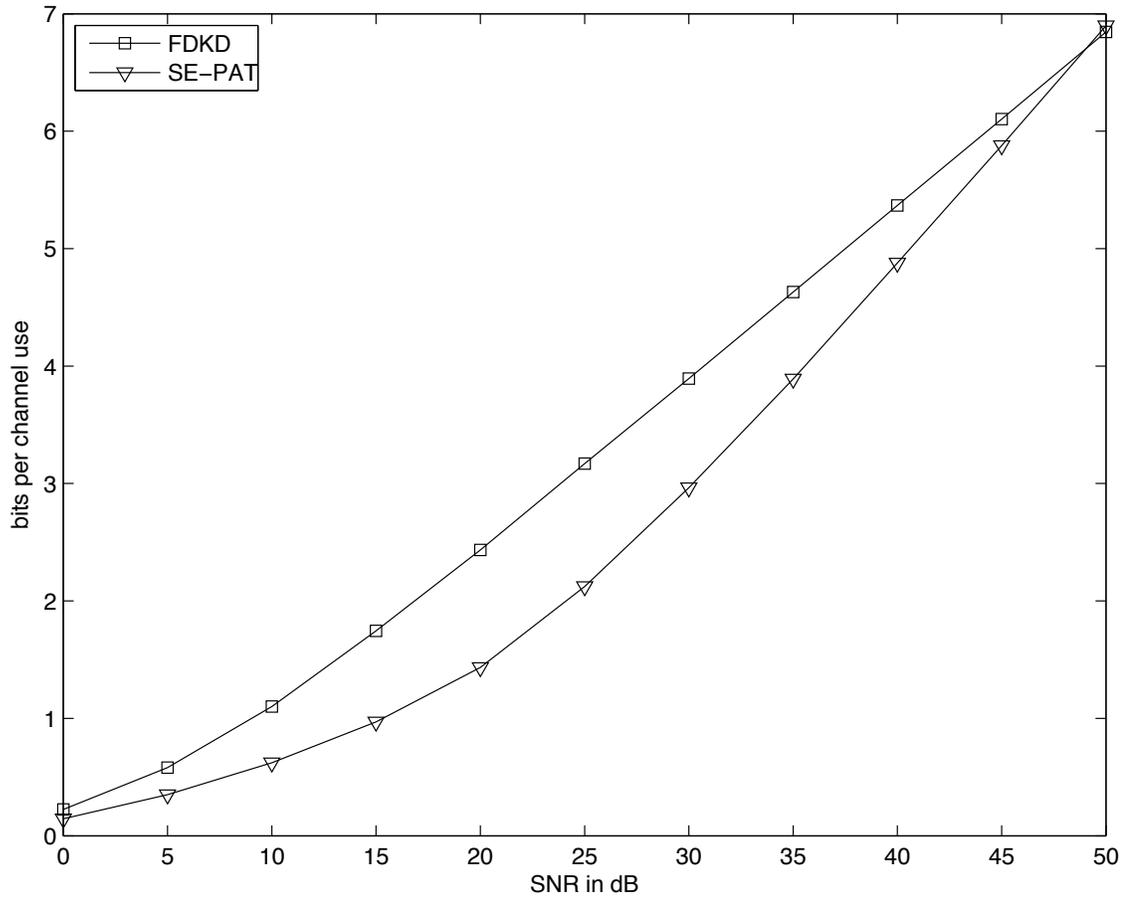


Figure 5.6: Intersection of achievable rates for $N = 126$, $N_{\text{delay}} = 14$, $N_{\text{Dopp}} = 3$.

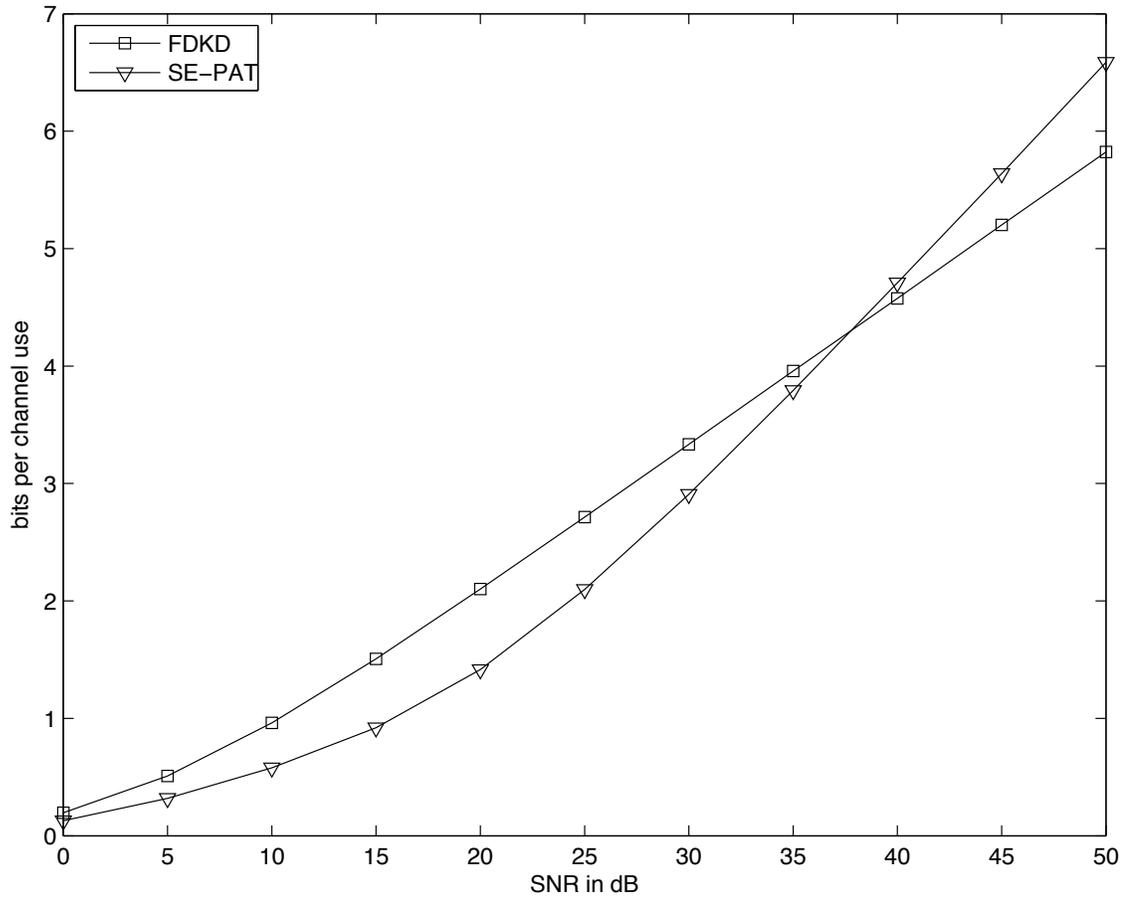


Figure 5.7: Intersection of achievable rates for $N = 96$, $N_{\text{delay}} = 12$, $N_{\text{Dopp}} = 3$.

CHAPTER 6

MMSE-PAT FOR MIMO SYSTEMS

Multi-antenna systems are widely used in the wireless communications due to their ability to provide high spectral efficiency by creating many “equivalent” parallel channels between the transmitter and the receiver. Several researchers have analyzed the performance of MIMO systems under different fading assumptions, for e.g. [7, 19, 28, 33, 42]. In this chapter, we extend the MMSE-PAT design for MIMO doubly selective channels [62] and new intuitions are uncovered in the MIMO case. For example, we show that, for independent fading between different antenna pairs, the number of receive antennas does not affect the design of MIMO-MMSE-PAT for DSC. We establish that the spectral efficiency of MIMO-MMSE-PAT does not necessarily increase even if the number of transmit and receive antennas are increased simultaneously. We also present the optimal number of active antennas which maximizes the spectral efficiency. Numerical examples are presented to illustrate the theoretical results. The works [13, 43], while similar in direction, were restricted to non-superimposed time-domain pilots. In fact, we show that a frequency-domain MMSE-PAT scheme achieves higher rates than the time-domain MMSE-PAT scheme from [43] when the MIMO doubly selective channel’s time spread dominates its frequency spread.

6.1 MIMO System Model

We consider a MIMO system with T transmit and R receive antennas, cyclic-prefix block transmission, and a DSC that satisfies a complex exponential basis expansion model. Details are given below.

The sampled complex-baseband output signal $\{y^{[r]}[n]\}$ at the r^{th} receive antenna is related to the transmitted signal $\{x^{[t]}[n]\}$ from the t^{th} transmit antenna via

$$y^{[r]}[n] = \sqrt{\rho} \sum_{t=0}^{T-1} \sum_{\ell=0}^{N_{\text{delay}}-1} h^{[r,t]}[n;\ell] x^{[t]}[n-\ell] + v^{[r]}[n], \quad (6.1)$$

where $\{v^{[r]}[n]\}$ is zero-mean unit-variance circular (spatially and temporally) white Gaussian noise and $h^{[r,t]}[n;\ell]$ is the time- n channel response at the r^{th} receive antenna to an impulse applied at time $n - \ell$ on the t^{th} transmit antenna. Here, N_{delay} denotes the channel's maximum time spread normalized to the sampling interval \mathcal{T}_s , which is assumed equal for all (r, t) antenna pairs. The length- N transmission block⁵ $\{x^{[t]}[n]\}_{n=0}^{N-1}$ is preceded by a cyclic prefix of length $N_{\text{delay}} - 1$, whose contribution is discarded in forming $\mathbf{y}^{[r]} = [y^{[r]}[0], \dots, y^{[r]}[N - 1]]^\top$. Throughout this chapter, we

⁵When we refer to a “transmission block of length N ,” we do not include the contribution from the CP.

assume modulo-N indexing, i.e., $z[i] = z[\langle i \rangle_N]$. With the definitions

$$\begin{aligned}
\mathbf{X} &= \text{diag}(\mathbf{X}^{[0]}, \dots, \mathbf{X}^{[T-1]}) \\
\mathbf{X}^{[t]} &= [\mathbf{X}_0^{[t]} \cdots \mathbf{X}_{-N_{\text{delay}}+1}^{[t]}] \\
\mathbf{X}_k^{[t]} &= \text{diag}(x^{[t]}[k], \dots, x^{[t]}[k+N-1]) \\
\mathbf{h}^{[r]} &= [\mathbf{h}^{[r,0]\top} \cdots \mathbf{h}^{[r,T-1]\top}]^\top \\
\mathbf{h}^{[r,t]} &= [\mathbf{h}_0^{[r,t]\top} \cdots \mathbf{h}_{N_{\text{delay}}-1}^{[r,t]\top}]^\top \\
\mathbf{h}_k^{[r,t]} &= [h^{[r,t]}[0;k], \dots, h^{[r,t]}[N-1;k]]^\top \\
\mathbf{v}^{[r]} &= [v^{[r]}[0], \dots, v^{[r]}[N-1]]^\top,
\end{aligned}$$

the DSC model (6.1) can be rewritten as

$$\mathbf{y}^{[r]} = \sqrt{\rho} \sum_{j=0}^{T-1} \mathbf{X}^{[j]} \mathbf{h}^{[r,j]} + \mathbf{v}^{[r]}, \quad (6.2)$$

$$= \sqrt{\rho} \underbrace{(\mathbf{X}^{[0]} \cdots \mathbf{X}^{[T-1]})}_{\mathbf{X}} \underbrace{\begin{pmatrix} \mathbf{h}^{[r,0]} \\ \vdots \\ \mathbf{h}^{[r,T-1]} \end{pmatrix}}_{\mathbf{h}^{[r]}} + \mathbf{v}^{[r]}. \quad (6.3)$$

Collecting the observations from different receive antennas in $\bar{\mathbf{y}} = [\mathbf{y}^{[0]\top}, \dots, \mathbf{y}^{[R-1]\top}]^\top$, we have

$$\bar{\mathbf{y}} = \sqrt{\rho} \bar{\mathbf{X}} \bar{\mathbf{h}} + \bar{\mathbf{v}}, \quad (6.4)$$

with $\bar{\mathbf{X}} = \mathbf{I}_R \otimes \mathbf{X}$, $\bar{\mathbf{h}} = [\mathbf{h}^{[0]\top}, \dots, \mathbf{h}^{[R-1]\top}]^\top$, and $\bar{\mathbf{v}} = [\mathbf{v}^{[0]\top}, \dots, \mathbf{v}^{[R-1]\top}]^\top$.

The transmit signal is constructed as $x^{[t]}[n] = p^{[t]}[n] + d^{[t]}[n]$, where $\{p^{[t]}[n]\}$ is the deterministic pilot sequence and $\{d^{[t]}[n]\}$ is the zero-mean data sequence. Note the superposition of pilots and data. Using $\{p^{[t]}[n]\}$ and $\{d^{[t]}[n]\}$ to construct \mathbf{P} and \mathbf{D} , respectively, in the manner of \mathbf{X} , we see that

$$\mathbf{X} = \mathbf{P} + \mathbf{D}, \quad (6.5)$$

which again shows the superposition of pilots and data. Similarly, we have $\bar{\mathbf{X}} = \bar{\mathbf{P}} + \bar{\mathbf{D}}$. Defining the pilot vector $\mathbf{p}^{[t]} = [p^{[t]}[0], \dots, p^{[t]}[N-1]]^\top$, the pilot energy is constrained as

$$\sum_{t=0}^{T-1} \|\mathbf{p}^{[t]}\|^2 = E_p. \quad (6.6)$$

The data vector $\mathbf{d}^{[t]} = [d^{[t]}[0], \dots, d^{[t]}[N-1]]^\top$ is obtained by linear modulation of N_s information bearing symbols $\mathbf{s}^{[t]} = [s^{[t]}[0], \dots, s^{[t]}[N_s-1]]^\top$ according to

$$\mathbf{d}^{[t]} = \mathbf{B}^{[t]} \mathbf{s}^{[t]}, \quad (6.7)$$

where $\mathbf{B}^{[t]}$ is the t^{th} transmit antenna's precoding matrix. We require that the columns of $\mathbf{B}^{[t]}$ are linearly independent.

We assume that the channel coefficients between different antenna pairs are independent with identical second-order statistics. The following CE-BEM describes the channel response between the r^{th} receive and t^{th} transmit antenna over the N -length block duration. For $n \in \{0, \dots, N-1\}$ and $l \in \{0, \dots, N_{\text{delay}}-1\}$,

$$h^{[r,t]}[n; l] = \frac{1}{\sqrt{N}} \sum_{k=-(N_{\text{Dopp}}-1)/2}^{(N_{\text{Dopp}}-1)/2} \lambda^{[r,t]}[k; l] e^{j \frac{2\pi}{N} kn}, \quad (6.8)$$

where CE-BEM coefficients $\{\lambda^{[r,t]}[k; l]\}$ are assumed to be zero-mean uncorrelated Gaussian with positive variance. We allow CE-BEM coefficients with possibly unequal variances in order to model arbitrary delay profiles and Doppler spectra.

We define the $N \times N_{\text{Dopp}}$ matrix $\bar{\mathbf{F}}$ element-wise as $[\bar{\mathbf{F}}]_{n,m} = \frac{1}{\sqrt{N}} e^{j \frac{2\pi}{N} n(m - \frac{N_{\text{Dopp}}-1}{2})}$ and notice that $\bar{\mathbf{F}}^H \bar{\mathbf{F}} = \mathbf{I}_{N_{\text{Dopp}}}$. With the definitions

$$\begin{aligned} \mathbf{U}_0 &= \mathbf{I}_{N_{\text{delay}}} \otimes \bar{\mathbf{F}} \\ \boldsymbol{\lambda}_l^{[r,t]} &= [\lambda^{[r,t]}[-\frac{N_{\text{Dopp}}-1}{2}; l], \dots, \lambda^{[r,t]}[\frac{N_{\text{Dopp}}-1}{2}; l]^\top \\ \boldsymbol{\lambda}^{[r,t]} &= [\boldsymbol{\lambda}_0^{[r,t]\top} \dots \boldsymbol{\lambda}_{N_{\text{delay}}-1}^{[r,t]\top}]^\top, \end{aligned}$$

(6.8) becomes

$$\mathbf{h}^{[r,t]} = \mathbf{U}_0 \boldsymbol{\lambda}^{[r,t]}, \quad (6.9)$$

which is the Karhunen-Loeve expansion of $\mathbf{h}^{[r,t]}$, since $\mathbf{U}_0^H \mathbf{U}_0 = \mathbf{I}_{N_{\text{Dopp}} N_{\text{delay}}}$ and since $\mathbf{R}_\lambda^{[r,t]} = \text{E}\{\boldsymbol{\lambda}^{[r,t]} \boldsymbol{\lambda}^{[r,t]H}\} > 0$ is diagonal. Since we assume identical second-order statistics for every transmit-receive antenna pair, we abbreviate $\mathbf{R}_\lambda^{[r,t]}$ by \mathbf{R}_λ . We assume an energy preserving channel between each transmit-receive antenna pair such that $\frac{1}{N} \text{E}\{\|\mathbf{h}^{[r,t]}\|^2\} = \frac{1}{N} \text{tr}\{\mathbf{R}_\lambda\} = 1$. Now, with

$$\begin{aligned} \mathbf{U} &= \mathbf{I}_T \otimes \mathbf{U}_0, \\ \bar{\mathbf{U}} &= \mathbf{I}_R \otimes \mathbf{U}, \\ \boldsymbol{\lambda}^{[r]} &= [\boldsymbol{\lambda}^{[r,0]T}, \dots, \boldsymbol{\lambda}^{[r,T-1]T}]^T, \\ \bar{\boldsymbol{\lambda}} &= [\boldsymbol{\lambda}^{[0]T}, \dots, \boldsymbol{\lambda}^{[R-1]T}]^T, \end{aligned}$$

we have $\mathbf{h}^{[r]} = \mathbf{U} \boldsymbol{\lambda}^{[r]}$, $\bar{\mathbf{h}} = \bar{\mathbf{U}} \bar{\boldsymbol{\lambda}}$ and

$$\bar{\mathbf{y}} = \sqrt{\rho}(\bar{\mathbf{P}} + \bar{\mathbf{D}})\bar{\mathbf{U}}\bar{\boldsymbol{\lambda}} + \bar{\mathbf{v}}. \quad (6.10)$$

From the channel independence assumptions between different antenna pairs, we have

$$\mathbf{R}_{\bar{\boldsymbol{\lambda}}} = \text{E}\{\bar{\boldsymbol{\lambda}} \bar{\boldsymbol{\lambda}}^H\} = \mathbf{I}_{RT} \otimes \mathbf{R}_\lambda.$$

As in the SISO case, we refer to N_{delay} as the discrete time spread and to N_{Dopp} as the discrete frequency spread. In addition, we refer to $\mathcal{B}_{\text{Dopp}} \mathcal{T}_s$ as the normalized Doppler spread and to $\gamma = N_{\text{delay}} N_{\text{Dopp}} / N$ as the channel's spreading index. We restrict our focus to underspread channels for which $\gamma < 1$.

6.2 MIMO-MMSE-PAT Design

In this section, we present the MMSE-PAT design requirements for the MIMO DSC and provide several novel MIMO-MMSE-PAT schemes. Though we assume Gaussian fading statistics, the results in this Section are valid for non-Gaussian fading as well.

The LMMSE estimate of $\bar{\mathbf{h}}$ given knowledge of $\{\bar{\mathbf{y}}, \bar{\mathbf{P}}\}$ and knowledge of the second-order statistics of $\{\bar{\mathbf{h}}, \bar{\mathbf{D}}, \bar{\mathbf{v}}\}$ is [53]

$$\hat{\mathbf{h}} = \mathbf{R}_{\bar{\mathbf{y}}, \bar{\mathbf{h}}}^{\text{H}} \mathbf{R}_{\bar{\mathbf{y}}}^{-1} \bar{\mathbf{y}}, \quad (6.11)$$

where $\mathbf{R}_{\bar{\mathbf{y}}, \bar{\mathbf{h}}} = \text{E}\{\bar{\mathbf{y}}\bar{\mathbf{h}}^{\text{H}}\}$ and $\mathbf{R}_{\bar{\mathbf{y}}} = \text{E}\{\bar{\mathbf{y}}\bar{\mathbf{y}}^{\text{H}}\}$. Given our assumptions,

$$\begin{aligned} \mathbf{R}_{\bar{\mathbf{y}}, \bar{\mathbf{h}}} &= \sqrt{\rho} \bar{\mathbf{P}} \bar{\mathbf{U}} \mathbf{R}_{\bar{\lambda}} \bar{\mathbf{U}}^{\text{H}} \\ \mathbf{R}_{\bar{\mathbf{y}}} &= \rho \bar{\mathbf{P}} \bar{\mathbf{U}} \mathbf{R}_{\bar{\lambda}} \bar{\mathbf{U}}^{\text{H}} \bar{\mathbf{P}}^{\text{H}} + \rho \text{E}\{\bar{\mathbf{D}} \bar{\mathbf{U}} \mathbf{R}_{\bar{\lambda}} \bar{\mathbf{U}}^{\text{H}} \bar{\mathbf{D}}^{\text{H}}\} + \mathbf{I}_{NR}. \end{aligned}$$

The channel estimation error $\tilde{\mathbf{h}} = \bar{\mathbf{h}} - \hat{\mathbf{h}}$ has total MSE $\sigma_e^2 = \text{E}\{\|\tilde{\mathbf{h}}\|^2\}$, which is given by [53]

$$\sigma_e^2 = \text{tr}\{\bar{\mathbf{U}} \mathbf{R}_{\bar{\lambda}} \bar{\mathbf{U}}^{\text{H}} - \mathbf{R}_{\bar{\mathbf{y}}, \bar{\mathbf{h}}}^{\text{H}} \mathbf{R}_{\bar{\mathbf{y}}}^{-1} \mathbf{R}_{\bar{\mathbf{y}}, \bar{\mathbf{h}}}\}. \quad (6.12)$$

We are interested in finding the combination of energy constrained pilot vectors $\{\mathbf{p}^{[t]}\}$ and data precoding matrices $\{\mathbf{B}^{[t]}\}$ such that the resulting MSE σ_e^2 is minimal. We refer to such combinations $\{\mathbf{p}^{[t]}, \mathbf{B}^{[t]}\}, t \in \{0, \dots, T-1\}$ as MIMO-MMSE-PAT schemes.

Theorem 7 (MSE Lower Bound). *For T -transmit R -receive antenna N -block CP PAT over the CE-BEM DSC, with the non-data-aided channel estimator (6.11), the channel estimate MSE (6.12) obeys*

$$\sigma_e^2 \geq \text{tr} \left\{ \left(\mathbf{R}_{\bar{\lambda}}^{-1} + \frac{\rho E_p}{NT} \mathbf{I}_{N_{\text{Dopp}} N_{\text{delay}} RT} \right)^{-1} \right\}, \quad (6.13)$$

where equality in (6.13) occurs if and only if the following conditions hold:

1. Pilot-Data Orthogonality:

$$(\mathbf{P}\mathbf{U})^H \mathbf{D}\mathbf{U} = \mathbf{0}, \forall \mathbf{D} \quad (6.14)$$

2. Optimal Excitation:

$$(\mathbf{P}\mathbf{U})^H \mathbf{P}\mathbf{U} = \frac{E_p}{NT} \mathbf{I}. \quad (6.15)$$

When (6.14)-(6.15) are met, $\mathbf{R}_{\tilde{\mathbf{h}}} = \mathbb{E}\{\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H\}$ and $\mathbf{R}_{\hat{\mathbf{h}}} = \mathbb{E}\{\hat{\mathbf{h}}\hat{\mathbf{h}}^H\}$ are given by

$$\mathbf{R}_{\tilde{\mathbf{h}}} = \bar{\mathbf{U}} \left(\mathbf{R}_{\tilde{\lambda}}^{-1} + \frac{\rho E_p}{NT} \mathbf{I}_{N_{\text{Dopp}} N_{\text{delay}} RT} \right)^{-1} \bar{\mathbf{U}}^H \quad (6.16)$$

$$\mathbf{R}_{\hat{\mathbf{h}}} = \bar{\mathbf{U}} \left\{ \mathbf{R}_{\hat{\lambda}} - \left(\mathbf{R}_{\tilde{\lambda}}^{-1} + \frac{\rho E_p}{NT} \mathbf{I}_{N_{\text{Dopp}} N_{\text{delay}} RT} \right)^{-1} \right\} \bar{\mathbf{U}}^H. \quad (6.17)$$

Proof. See Appendix E.1. □

$\mathbf{P}\mathbf{U}$ and $\mathbf{D}\mathbf{U}$ are composed of blocks of the form $\mathbf{P}^{[t]}\mathbf{U}_0$ and $\mathbf{D}^{[t]}\mathbf{U}_0$, respectively, corresponding to each transmit-receive antenna pair. An analysis of these blocks using techniques similar to the SISO case (Chapter 4) can be used to re-state the MSE optimality requirements (6.14)-(6.15) in terms of the pilot sequence $\{p^{[t]}[i]\}$ and data basis sequence $\{b_q^{[t]}[i]\}$, where $p^{[t]}[i] = [\mathbf{p}^{[t]}]_i$ and $b_q^{[t]}[i] = [\mathbf{B}^{[t]}]_{i,q}$. We use the following index sets in the sequel: $\mathcal{N}_{\text{delay}} = \{-N_{\text{delay}} + 1, \dots, N_{\text{delay}} - 1\}$, $\mathcal{N}_{\text{Dopp}} = \{-N_{\text{Dopp}} + 1, \dots, N_{\text{Dopp}} - 1\}$, $\mathcal{T} = \{0, \dots, T - 1\}$ and $\mathcal{Q} = \{0, \dots, N_s - 1\}$.

Lemma 6 (Time Domain). *For T -transmit R -receive antenna N -block CP PAT over the CE-BEM DSC, the following are necessary and sufficient conditions for equality*

in (6.13). $\forall k \in \mathcal{N}_{\text{delay}}, \forall m \in \mathcal{N}_{\text{Dopp}}, \forall t_i \in \mathcal{T}, \forall q \in \mathcal{Q}$

$$\sum_{i=0}^{N-1} b_q^{[t_1]}[i] p^{[t_2]*}[i-k] e^{-j \frac{2\pi}{N} m i} = 0 \quad (6.18)$$

$$\sum_{i=0}^{N-1} p^{[t_1]}[i] p^{[t_2]*}[i-k] e^{-j \frac{2\pi}{N} m i} = \frac{E_p}{T} \delta[k] \delta[m] \delta[t_1 - t_2]. \quad (6.19)$$

Similar to the SISO case, the MMSE optimality (6.18) requires that the pilot and the data components of the received signal lie in orthogonal subspaces so that the linear estimate (6.11) is impervious to interference from unknown data. With MIMO systems, the MMSE optimality condition (6.19) requires that the received pilot components corresponding to different transmit antennas lie in orthogonal subspaces. This is intuitively satisfying since our channel independence assumptions imply that the received pilot contribution from one transmit antenna is not useful for the estimation of channel coefficients corresponding to other transmit antennas. We note also that the number of receive antennas R does not affect the MIMO-MMSE-PAT design requirements in Lemma 6, and hence the multiple receive-antenna MMSE-PAT design problem reduces to a single receive-antenna MMSE-PAT design problem. In fact, the symmetry among different antenna pairs implies that the MMSE-PAT design for a particular receive antenna suffices as the MMSE-PAT design for the other receive antennas. Here again, cooperation among the estimators of channels for different receive-antennas is ineffective due to the independence of these channels.

Next we establish the duality of time- and frequency-domain MIMO-MMSE-PAT. Using the N -point unitary DFT matrix \mathbf{F}_N , we define the frequency-domain pilot vectors and data precoding matrices, $\check{\mathbf{p}}^{[t]} = \mathbf{F}_N \mathbf{p}^{[t]}$ and $\check{\mathbf{B}}^{[t]} = \mathbf{F}_N \mathbf{B}^{[t]}$, respectively. The MIMO-MMSE-PAT requirements (6.18)-(6.19) are rewritten in terms of $\{\check{\mathbf{p}}^{[t]}, \check{\mathbf{B}}^{[t]}\}_{t \in \mathcal{T}}$ in the following Lemma, using $\check{p}^{[t]}[i] = [\check{\mathbf{p}}^{[t]}]_i$ and $\check{b}_q^{[t]}[i] = [\check{\mathbf{B}}^{[t]}]_{i,q}$.

Lemma 7 (Frequency Domain). *For T -transmit R -receive antenna N -block CP PAT over the CE-BEM DSC, the following are necessary and sufficient conditions for equality in (6.13). $\forall k \in \mathcal{N}_{\text{Dopp}}, \forall m \in \mathcal{N}_{\text{delay}}, \forall t_i \in \mathcal{T}, \forall q \in \mathcal{Q}$*

$$\sum_{i=0}^{N-1} \check{b}_q^{[t_1]}[i] \check{p}^{[t_2]*}[i-k] e^{-j\frac{2\pi}{N}mi} = 0. \quad (6.20)$$

$$\sum_{i=0}^{N-1} \check{p}^{[t_1]}[i] \check{p}^{[t_2]*}[i-k] e^{-j\frac{2\pi}{N}mi} = \frac{E_p}{T} \delta[k] \delta[m] \delta[t_1 - t_2]. \quad (6.21)$$

Since the conditions in Lemma 7 are same as those of Lemma 6, with the exception of the time spread N_{delay} and frequency spread N_{Dopp} interchanging their roles, we see that every time-domain MIMO-MMSE-PAT has a frequency-domain dual. This result is stated precisely as follows.

Lemma 8 (Duality). *If $\{\mathbf{p}^{[t]}, \mathbf{B}^{[t]}\}_{t \in \mathcal{T}}$ specifies a T -transmit antenna MIMO-MMSE N -block CP-PAT scheme for a CE-BEM DSC with time spread N_1 and frequency spread N_2 , then $\{\mathbf{F}_N \mathbf{p}^{[t]}, \mathbf{F}_N \mathbf{B}^{[t]}\}_{t \in \mathcal{T}}$ specifies a T -transmit antenna MIMO-MMSE N -block CP-PAT scheme for a CE-BEM DSC with time spread N_2 and frequency spread N_1 .*

Time-frequency duality is a well-known and fundamental concept (see, e.g., [17]) and its applicability to SISO-MMSE-PAT schemes for the DSC was demonstrated in Chapter 4. Here we have established that the time-frequency duality also extends to the more general case of MIMO-MMSE-PAT for the DSC.

For MIMO-MMSE-PAT, we now present bounds on the number of data symbols that can be transmitted from each antenna, i.e., the rank of the data precoding matrices $\{\mathbf{B}^{[t]}\}_{t \in \mathcal{T}}$. Given pilot vectors $\{\mathbf{p}^{[t]}\}_{t \in \mathcal{T}}$ satisfying (6.19), precoding matrix $\mathbf{B}^{[t]} \in \mathbb{C}^{N \times N_s}$ which satisfies (6.18) can be constructed as follows. Defining the

$(2N_{\text{Dopp}} - 1) \times N$ matrix $\check{\mathbf{F}}$ element-wise as $[\check{\mathbf{F}}]_{n,m} = \frac{1}{\sqrt{N}} e^{-j\frac{2\pi}{N}(n-N_{\text{Dopp}}+1)m}$, and then defining

$$\begin{aligned}
\mathbf{P}_k^{[t]} &= \text{diag}(p^{[t]}[k], \dots, p^{[t]}[k + N - 1]), \\
\mathbf{W}_k^{[t]} &= \check{\mathbf{F}} \mathbf{P}_k^{[t]H}, \\
\mathbf{W}^{[t]} &= [\mathbf{W}_{-N_{\text{delay}}+1}^{[t]\top} \cdots \mathbf{W}_{N_{\text{delay}}-1}^{[t]\top}]^\top, \\
\mathbf{W} &= [\mathbf{W}^{[0]\top} \cdots \mathbf{W}^{[T-1]\top}]^\top,
\end{aligned} \tag{6.22}$$

condition (6.18) becomes $\mathbf{W} \mathbf{b}_q^{[t]} = \mathbf{0}$, implying that, for each q , the vector $\mathbf{b}_q^{[t]}$ must lie in the null space of \mathbf{W} . This can be achieved by choosing the columns of $\mathbf{B}^{[t]}$ as a basis for null space of \mathbf{W} , and hence data dimension N_s must obey $N_s \leq \dim(\text{null}(\mathbf{W}))$. For the case when $N_s = \dim(\text{null}(\mathbf{W}))$, N_s , the number of information symbols per MIMO-MMSE N -block CP-PAT per transmit antenna, can be bounded as follows. Note from (6.22) that the $TN_{\text{Dopp}}N_{\text{delay}}$ rows of $(\mathbf{P}\mathbf{U})^H$ are contained within the $T(2N_{\text{Dopp}} - 1)(2N_{\text{delay}} - 1)$ rows of \mathbf{W} . In order to satisfy (6.15), those rows must be orthogonal. Thus, $TN_{\text{Dopp}}N_{\text{delay}} \leq \text{rank}(\mathbf{W}) \leq T(2N_{\text{Dopp}} - 1)(2N_{\text{delay}} - 1)$, which means that (6.18)-(6.19) imply

$$N - T(2N_{\text{Dopp}} - 1)(2N_{\text{delay}} - 1) \leq N_s \leq N - TN_{\text{Dopp}}N_{\text{delay}}. \tag{6.23}$$

From (6.23), we see that MIMO-MMSE-PAT sacrifices at least $TN_{\text{delay}}N_{\text{Dopp}}$, but no more than $T(2N_{\text{delay}} - 1)(2N_{\text{Dopp}} - 1)$, signaling dimensions per transmit antenna. Recall that $N_{\text{Dopp}}N_{\text{delay}}$ describes the number of degrees of freedom in the DSC per each transmit-receive antenna pair and $TN_{\text{Dopp}}N_{\text{delay}}$ denotes the number of (independent) channel coefficients to be estimated at each receive antenna. Increasing T , increases the number of parameters to be estimated and thus reduces the number of data symbols from each antenna.

6.2.1 MIMO-MMSE-PAT Examples

Now, we present several examples of N -block CP MIMO-MMSE-PAT schemes for the CE-BEM DSC using the $\{\mathbf{p}^{[t]}, \mathbf{B}^{[t]}\}_{t \in \mathcal{T}}$ parameterization.

Example 6 (MIMO-TDKD). Assuming $\frac{N}{N_{\text{Dopp}}} \in \mathbb{Z}$, consider the pilot index sets $\mathcal{P}_t^{[t]}$ and the guard index set \mathcal{G}_t :

$$\begin{aligned}\mathcal{P}_t^{[t]} &= \left\{ \ell + tN_{\text{delay}}, \ell + tN_{\text{delay}} + \frac{N}{N_{\text{Dopp}}}, \dots, \ell + tN_{\text{delay}} + \frac{(N_{\text{Dopp}}-1)N}{N_{\text{Dopp}}} \right\} \\ \mathcal{G}_t &= \bigcup_{t \in \mathcal{T}} \bigcup_{k \in \mathcal{P}_t^{[t]}} \{-N_{\text{delay}} + 1 + k, \dots, N_{\text{delay}} - 1 + k\}.\end{aligned}$$

An N -block CP MIMO MMSE-PAT scheme for the CE-BEM DSC is given by

$$p^{[t]}[q] = \begin{cases} \sqrt{\frac{E_p}{TN_{\text{Dopp}}}} e^{j\theta[q]} & q \in \mathcal{P}_t^{[t]} \\ 0 & q \notin \mathcal{P}_t^{[t]} \end{cases} \quad (6.24)$$

and by $\mathbf{B}^{[t]}$ constructed from the columns of \mathbf{I}_N with indices not in the set \mathcal{G}_t . Both $\ell \in \{0, \dots, \frac{N}{N_{\text{Dopp}}} - 1\}$ and $\theta[q] \in \mathbb{R}$, are arbitrary. The corresponding data dimension per transmit antenna is $N_s = N - (T+1)N_{\text{Dopp}}N_{\text{delay}} + N_{\text{Dopp}}$.

Example 7 (MIMO-FDKD). Assuming $\frac{N}{N_{\text{delay}}} \in \mathbb{Z}$, consider the (frequency-domain) pilot index sets $\mathcal{P}_f^{[t]}$ and the guard index set \mathcal{G}_f :

$$\begin{aligned}\mathcal{P}_f^{[t]} &= \left\{ \ell + tN_{\text{Dopp}}, \ell + tN_{\text{Dopp}} + \frac{N}{N_{\text{delay}}}, \dots, \ell + tN_{\text{Dopp}} + \frac{(N_{\text{delay}}-1)N}{N_{\text{delay}}} \right\} \\ \mathcal{G}_f &= \bigcup_{t \in \mathcal{T}} \bigcup_{k \in \mathcal{P}_f^{[t]}} \{-N_{\text{Dopp}} + 1 + k, \dots, N_{\text{Dopp}} - 1 + k\}.\end{aligned}$$

An N -block CP MIMO-MMSE-PAT scheme for the CE-BEM DSC is given by $\mathbf{p}^{[t]} = \mathbf{F}_N^H \check{\mathbf{p}}^{[t]}$, with

$$\check{p}^{[t]}[q] = \begin{cases} \sqrt{\frac{E_p}{TN_{\text{delay}}}} e^{j\theta[q]} & q \in \mathcal{P}_f^{[t]} \\ 0 & q \notin \mathcal{P}_f^{[t]} \end{cases} \quad (6.25)$$

and by $\mathbf{B}^{[t]}$ constructed from the columns of inverse DFT matrix \mathbf{F}_N^H with indices not in the set \mathcal{G}_f . Both $\ell \in \{0, \dots, \frac{N}{N_{\text{delay}}} - 1\}$ and $\theta[q] \in \mathbb{R}$, are arbitrary. The corresponding data dimension per transmit antenna is $N_s = N - (T + 1)N_{\text{Dopp}}N_{\text{delay}} + N_{\text{delay}}$.

Example 8 (MIMO-Time-Domain-Chirp). Assuming even N , an N -block CP MIMO-MMSE-PAT scheme for the CE-BEM DSC is given by

$$p^{[t]}[q] = \sqrt{\frac{E_p}{NT}} e^{j\frac{2\pi}{N} \left(\frac{N_{\text{Dopp}}}{2} q^2 + tN_{\text{Dopp}}N_{\text{delay}}q \right)} \quad (6.26)$$

$$[\mathbf{B}^{[t]}]_{q,m} = \frac{1}{\sqrt{N}} e^{j\frac{2\pi}{N} \left(\frac{N_{\text{Dopp}}}{2} q^2 + (m + TN_{\text{Dopp}}N_{\text{delay}})q \right)}, \quad (6.27)$$

for $q \in \{0, \dots, N-1\}$ and $m \in \{0, \dots, N_s-1\}$, where the data dimension per transmit antenna is $N_s = N - (T + 1)N_{\text{Dopp}}N_{\text{delay}} + 1$.

Example 9 (MIMO-Frequency-Domain-Chirp). Assuming even N , an N -block CP MMSE-PAT scheme for the CE-BEM DSC is given by $\mathbf{p}^{[t]} = \mathbf{F}_N^H \check{\mathbf{p}}^{[t]}$ and $\mathbf{B}^{[t]} = \mathbf{F}_N^H \check{\mathbf{B}}^{[t]}$, with

$$\check{p}^{[t]}[q] = \sqrt{\frac{E_p}{NT}} e^{j\frac{2\pi}{N} \left(\frac{N_{\text{delay}}}{2} q^2 + tN_{\text{Dopp}}N_{\text{delay}}q \right)} \quad (6.28)$$

$$[\check{\mathbf{B}}^{[t]}]_{q,m} = \frac{1}{\sqrt{N}} e^{j\frac{2\pi}{N} \left(\frac{N_{\text{delay}}}{2} q^2 + (m + TN_{\text{Dopp}}N_{\text{delay}})q \right)}, \quad (6.29)$$

for $q \in \{0, \dots, N-1\}$ and $m \in \{0, \dots, N_s-1\}$, where the data dimension per transmit antenna is $N_s = N - (T + 1)N_{\text{Dopp}}N_{\text{delay}} + 1$.

It is straightforward to verify that the above examples satisfy the MIMO-MMSE-PAT requirements (6.18)-(6.19). The structure of MIMO-TDKD is similar to the MMSE-PAT scheme from [43], with the difference that [43] focused on zero-padded (ZP) block transmissions. We undertake a detailed comparison of ZP and CP schemes

in Section 6.5. For CP transmissions with (time) superimposed pilots, we see that there exist other MMSE-PAT schemes, e.g., MIMO-FDKD and MIMO-Chirp schemes. For the MIMO-TDKD scheme (illustrated in Fig. 6.1), the impulsive nature of the equi-spaced pilot blocks yields direct estimates of the channel's time varying impulse response. The spacing between the pilot impulses of each transmit antenna (i.e., $\frac{N}{N_{\text{Dopp}}} \approx \frac{1}{2B_{\text{Dopp}}}$) is such that a time-domain Nyquist sampling criterion is satisfied, and the guard zeros in each pilot block are constructed to prevent interference from data and pilot symbols transmitted by other antennas. MIMO-FDKD can be recognized as the frequency-domain dual of MIMO-TDKD, where the pilot-impulse spacing is chosen so that a frequency-domain Nyquist criterion is satisfied, and where the roles of N_{delay} and N_{Dopp} reverse. Though the MIMO-Chirp schemes has pilot and data components that overlap in time and frequency, the pilots and data are still linearly separable at the channel output, as required by (6.14). We note that the MIMO-Chirp schemes may have advantages over MIMO-TDKD and MIMO-FDKD in peak-to-average power ratio.

6.3 Achievable Rates of MIMO-MMSE-PAT

We now calculate bounds on the ergodic achievable rate of MIMO-MMSE-PAT for the CE-BEM DSC from Section 6.2, paying special attention to the high-SNR regime. As in the SISO case, we consider Gaussian codes and mismatched weighted minimum distance decoding. Our unified achievable-rate analysis holds for an arbitrary $\{\mathbf{p}^{[t]}, \mathbf{B}^{[t]}\}_{t \in \mathcal{T}}$ MIMO-MMSE-PAT scheme, where the pilot vectors $\{\mathbf{p}^{[t]}\}_{t \in \mathcal{T}}$ and the data precoding matrices $\{\mathbf{B}^{[t]}\}_{t \in \mathcal{T}}$ satisfy the requirements of Lemma 6.

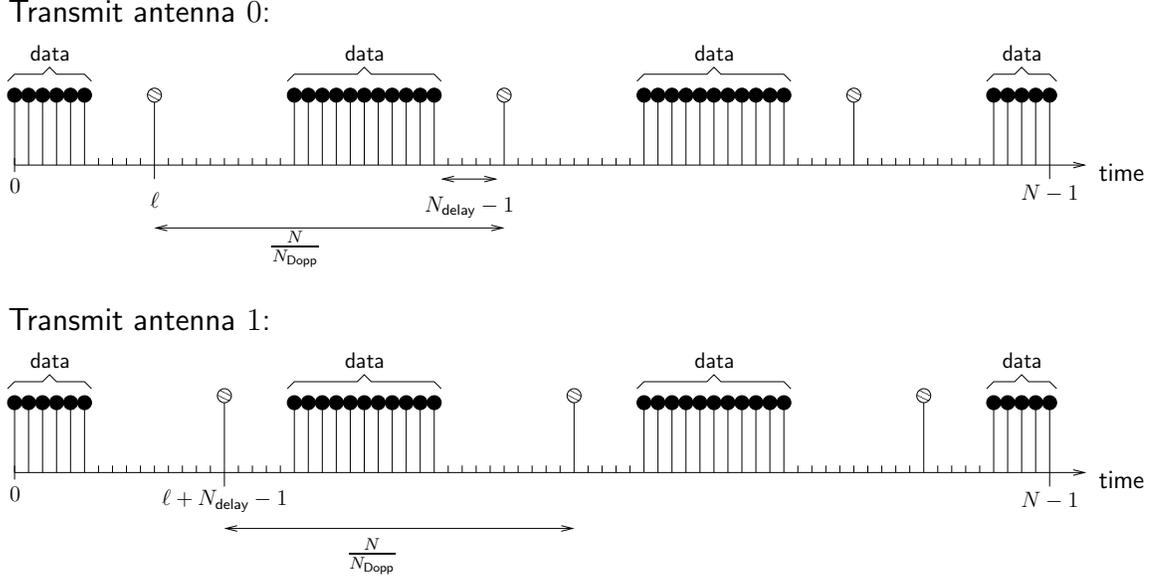


Figure 6.1: Structure of $T = 2$ MIMO-TDKD for DSC with $N_{\text{delay}} = 5$ and $N_{\text{Dopp}} = 3$.

We consider MIMO-MMSE-PAT schemes for the model discussed in Section 6.1. Now, we assume that the columns of precoding matrix $\mathbf{B}^{[t]}$ are orthonormal (whereas earlier $\mathbf{B}^{[t]}$ was specified as full rank). Note that, for the purpose of ergodic achievable-rate analysis, this assumption can be made without loss of generality, since the mutual information between \mathbf{s} and \mathbf{y} remains unaffected by invertible transformations of \mathbf{s} . Suppose that the MIMO-MMSE-PAT scheme $\{\mathbf{p}^{[t]}, \mathbf{B}^{[t]}\}_{t \in \mathcal{T}}$ for the CE-BEM DSC has total pilot energy E_p [recall (6.6)] and yields per-transmit-antenna data dimension N_s . We define the data energy $\mathbb{E}\{\sum_{t=0}^{T-1} \|\mathbf{d}^{[t]}\|^2\} = \mathbb{E}\{\sum_{t=0}^{T-1} \|\mathbf{s}^{[t]}\|^2\} = E_s$, the total transmit energy $E_{\text{tot}} = E_p + E_s$, the average transmit power $\sigma^2 = \frac{1}{N} \sum_{t=0}^{T-1} \sum_{n=0}^{N-1} \mathbb{E}\{|x^{[t]}[n]|^2\} = \frac{E_{\text{tot}}}{N}$. We require that $E_{\text{tot}} \leq N$ and ρ is the signal-to-noise ratio of the system. In addition, we define the normalized signal power $\sigma_s^2 = \frac{E_s}{TN_s}$.

In the sequel, we analyze the ergodic achievable rates of MIMO-MMSE-PAT schemes $\{\mathbf{p}^{[t]}, \mathbf{B}^{[t]}\}_{t \in \mathcal{T}}$ over the CE-BEM-DSC.

It will be convenient to define $\mathbf{H}^{[r,t]} \in \mathbb{C}^{N \times N}$ element-wise as $[\mathbf{H}^{[r,t]}]_{n,m} = h^{[r,t]}[n; \langle n-m \rangle_N]$, so that the input-output relation (6.3) becomes

$$\mathbf{y}^{[r]} = \sqrt{\rho} \underbrace{[\mathbf{H}^{[r,0]} \dots \mathbf{H}^{[r,T-1]}]}_{\mathbf{H}^{[r]}} (\bar{\mathbf{p}} + \bar{\mathbf{B}}\bar{\mathbf{s}}) + \mathbf{v}^{[r]}, \quad (6.30)$$

where $\bar{\mathbf{p}} = [\mathbf{p}^{[0]\top}, \dots, \mathbf{p}^{[T-1]\top}]^\top$, $\bar{\mathbf{s}} = [\mathbf{s}^{[0]\top}, \dots, \mathbf{s}^{[T-1]\top}]^\top$, and $\bar{\mathbf{B}} = \text{diag}(\mathbf{B}^{[0]}, \dots, \mathbf{B}^{[T-1]})$. Then, defining $\bar{\mathbf{H}} = [\mathbf{H}^{[0]\top}, \dots, \mathbf{H}^{[R-1]\top}]^\top$, we collect the observations of all receive antennas into $\bar{\mathbf{y}} = [\mathbf{y}^{[0]\top}, \dots, \mathbf{y}^{[R-1]\top}]^\top$, such that

$$\bar{\mathbf{y}} = \sqrt{\rho} \bar{\mathbf{H}} \bar{\mathbf{p}} + \sqrt{\rho} \bar{\mathbf{H}} \bar{\mathbf{B}} \bar{\mathbf{s}} + \bar{\mathbf{v}}, \quad (6.31)$$

with $\bar{\mathbf{v}} = [\mathbf{v}^{[0]\top}, \dots, \mathbf{v}^{[R-1]\top}]^\top$. Similar to $\bar{\mathbf{H}}$, we construct $\hat{\mathbf{H}}$ and $\check{\mathbf{H}}$ using $\hat{\mathbf{h}}$ and $\check{\mathbf{h}}$ respectively. To present the bounds, we use the normalized channel estimate

$$\check{\mathbf{H}} = \hat{\mathbf{H}} \sqrt{NRT / \text{tr}\{\mathbf{R}_{\hat{\mathbf{h}}}\}} \text{ so that } \mathbb{E}\{\check{\mathbf{H}}^H \check{\mathbf{H}}\} = \mathbb{E}\{\bar{\mathbf{H}}^H \bar{\mathbf{H}}\} = R\mathbf{I}_{NT}.$$

Theorem 8 (Achievable-Rate Bounds). *For the T -transmit R -receive antenna N -block CP MIMO-MMSE-PAT scheme $\{\mathbf{p}^{[t]}, \mathbf{B}^{[t]}\}_{t \in \mathcal{T}}$ with i.i.d. Gaussian $\bar{\mathbf{s}} \in \mathbb{C}^{N_s}$ over the CE-BEM DSC, the per-block ergodic achievable rate $\mathcal{R}_{\text{mmse-blk}}$ obeys*

$\mathcal{R}_{\text{mimo-mmse-blk-lb}} \leq \mathcal{R}_{\text{mimo-mmse-blk}} \leq \mathcal{R}_{\text{mimo-mmse-blk-ub}}$, where

$$\mathcal{R}_{\text{mimo-mmse-blk-lb}} = \mathbb{E}\{\log \det[\mathbf{I} + \rho_{\text{lb}} \bar{\mathbf{B}}^H \check{\mathbf{H}}^H \check{\mathbf{H}} \bar{\mathbf{B}}]\} \text{ bits/block} \quad (6.32)$$

$$\mathcal{R}_{\text{mimo-mmse-blk-ub}} = \mathbb{E}\{\log \det[\mathbf{I} + \rho_{\text{ub}} \bar{\mathbf{B}}^H \bar{\mathbf{H}}^H \bar{\mathbf{H}} \bar{\mathbf{B}}]\} \text{ bits/block} \quad (6.33)$$

$$\rho_{\text{lb}} = \frac{\rho \sigma_s^2 \text{tr}\{\mathbf{R}_{\hat{\mathbf{h}}}\}}{T \rho \sigma_s^2 \text{tr}\{\mathbf{R}_{\hat{\mathbf{h}}}\} + NRT} \quad (6.34)$$

$$\rho_{\text{ub}} = \rho \sigma_s^2. \quad (6.35)$$

Proof. See Appendix E.2. □

The lower bound (6.32) describes the worst case scenario of channel estimation error acting as CWGN. This concept was previously used in, e.g., [41]. The upper bound (6.33) describes the best case scenario of perfect channel estimates. In the next section, we study the time and bandwidth requirements of each block and rewrite these rates in units of bits/sec/Hz.

Now, we characterize the high-SNR asymptotics of the achievable-rate $\mathcal{R}_{\text{mimo-mmse-blk}}$ of MIMO-MMSE-PAT schemes.

Theorem 9 (Asymptotic Achievable Rate). *For an N -block CP MIMO-MMSE-PAT scheme operating over the CE-BEM DSC with T transmit and R receive antennas, and with per-transmit-antenna data dimension N_s , the per-block ergodic achievable rate $\mathcal{R}_{\text{mimo-mmse-blk}}$ obeys*

$$\mathcal{R}_{\text{mimo-mmse-blk}}(\rho) = \min\{R(N - TN_{\text{Dopp}}N_{\text{delay}}), TN_s\} \log \rho + O(1), \text{ bits/block} \quad (6.36)$$

as $\rho \rightarrow \infty$.

Proof. See Appendix E.3. □

The pre-log factor in $\mathcal{R}_{\text{mimo-mmse-blk}}$ can be interpreted as follows. Since each transmit antenna is allocated only N_s data dimensions, the total number of data symbols transmitted per block is TN_s . At each receive antenna, the observations corresponding to MIMO-MMSE-PAT pilot symbol $\mathbf{PU}\boldsymbol{\lambda}^{[r]}$ uses up $TN_{\text{Dopp}}N_{\text{delay}}$ dimensions [recall (6.15)], leaving only $N - TN_{\text{Dopp}}N_{\text{delay}}$ data observations, due to the requirement for pilot-data orthogonality at the channel output [recall (6.14)]. Thus, at most $\min\{R(N - TN_{\text{Dopp}}N_{\text{delay}}), TN_s\}$ data symbols per N -block can be “resolved” at the receiver.

Increasing the number of transmit antennas increases the total number of data symbols that can be transmitted but also increases the number of channel parameters that need to be estimated. For this reason, the asymptotic achievable rate of MIMO-MMSE-PAT does not necessarily increase with the number of antennas. To see this, with $Q = \min\{R(N - TN_{\text{Dopp}}N_{\text{delay}}), TN_s\}$ denoting the pre-log factor, note that $Q \leq TN_s \leq T(N - TN_{\text{Dopp}}N_{\text{delay}})$, using the bound (6.23) on N_s . As a continuous function of t , $t(N - tN_{\text{delay}}N_{\text{Dopp}})$ is a parabola whose maximum attained at $t = \frac{N}{2N_{\text{delay}}N_{\text{Dopp}}}$. It follows that increase of number of receive and transmit antennas beyond $\lceil \frac{N}{2N_{\text{delay}}N_{\text{Dopp}}} \rceil$ does not increase the pre-log factor. In fact, increasing the number of transmit antennas beyond $\lceil \frac{N}{2N_{\text{delay}}N_{\text{Dopp}}} \rceil$ *decreases* the pre-log factor. Hence, it is worthwhile to determine the number of active transmit and receive antennas, $T_\star \in \{1, \dots, T\}$ and $R_\star \in \{1, \dots, R\}$, which maximize the pre-log factor in $\mathcal{R}_{\text{mmse-blk}}$. Note that T_\star and R_\star depend on the choice of MIMO-MMSE-PAT scheme through data dimension N_s . For the MIMO-MMSE-PAT examples from Section 6.2.1, we have

$$R_\star = R \tag{6.37}$$

$$T_\star = \begin{cases} \min\left(T, \left\lceil \frac{\xi}{2N_{\text{Dopp}}N_{\text{delay}}} - \frac{R_\star}{2} \right\rceil\right) & \text{if } \xi^2 < 4R_\star NN_{\text{delay}}N_{\text{Dopp}}, \\ \min\left(T, \left\lceil \frac{\xi - \sqrt{\xi^2 - 4R_\star NN_{\text{Dopp}}N_{\text{delay}}}}{2N_{\text{Dopp}}N_{\text{delay}}} \right\rceil\right) & \text{else.} \end{cases} \tag{6.38}$$

$$\xi = N + (R_\star - 1)N_{\text{Dopp}}N_{\text{delay}} + \kappa \tag{6.39}$$

where $\kappa = 1$, N_{delay} , and N_{Dopp} for Chirp schemes, FDKD, and TDKD, respectively. In (6.38), $\lceil \cdot \rceil$ denotes rounding up or down to the nearest integer depending on the which gives higher pre-log factor. Derivation details are given in Appendix E.4. Note that, in some cases, the transmitter uses strictly less than T antennas. Similar results were obtained for MIMO flat-fading and time-selective channels in [7] and [9], respectively.

6.4 Pilot/Data Power Allocation

Until now, the MMSE-PAT schemes were designed using fixed pilot energy E_p . Now we consider the problem of allocating a fixed transmit energy E_{tot} between pilots and data in order to maximize ρ_{lb} . Letting $\alpha \in [0, 1]$ denote the fraction of energy allocated to the data symbols, i.e., $E_s = \alpha E_{\text{tot}}$ and $E_p = (1 - \alpha)E_{\text{tot}}$, we are interested in finding $\alpha_\star = \arg \max_\alpha \rho_{\text{lb}}(\alpha)$. The value of α_\star is obtained by finding the value of α which sets $\partial \rho_{\text{lb}} / \partial \alpha = 0$.

For the general case of $\mathbf{R}_{\bar{\lambda}}$, it is difficult to obtain a closed-form expression for α_\star . For the case of identically distributed coefficients, i.e., $\mathbf{R}_{\bar{\lambda}} = \frac{N}{N_{\text{Dopp}} N_{\text{delay}}} \mathbf{I}$, the desired value of α is given by

$$\alpha_{\star, \text{iid}} = \begin{cases} \beta - \sqrt{\beta^2 - \beta} & \text{if } N_s > TN_{\text{Dopp}} N_{\text{delay}} \\ \beta + \sqrt{\beta^2 - \beta} & \text{if } N_s < TN_{\text{Dopp}} N_{\text{delay}} \\ \frac{1}{2} & \text{if } N_s = TN_{\text{Dopp}} N_{\text{delay}} \end{cases} \quad (6.40)$$

$$\beta = \frac{1 + \frac{TN_{\text{Dopp}} N_{\text{delay}}}{\rho E_{\text{tot}}}}{1 - \frac{TN_{\text{Dopp}} N_{\text{delay}}}{N_s}}. \quad (6.41)$$

Furthermore, it can be shown that $\alpha_{\star, \text{iid}}$ maximizes the achievable rate lower bound (6.32) for the i.i.d. BEM case. To see this, note that $\bar{\mathbf{H}}$ and $\check{\mathbf{H}}$ have same distribution, so that the power allocation fraction α affects $\mathcal{R}_{\text{mimo-mmse-blk-lb}}$ only through ρ_{lb} . Since $\mathcal{R}_{\text{mimo-mmse-blk-lb}}$ is an increasing function of ρ_{lb} , maximizing ρ_{lb} is equivalent to maximizing $\mathcal{R}_{\text{mimo-mmse-blk-lb}}$.

Returning to the case of general $\mathbf{R}_{\bar{\lambda}}$, closed-form solutions of α_\star are tractable for the high-SNR and low-SNR asymptotic cases. In particular, it can be shown that $\arg \max_\alpha \lim_{\rho \rightarrow 0} \rho_{\text{lb}} = \frac{1}{2}$ and $\arg \max_\alpha \lim_{\rho \rightarrow \infty} \rho_{\text{lb}} = \lim_{\rho \rightarrow \infty} \alpha_{\star, \text{iid}}$, which can be calculated from (6.40) with $\beta = (1 - TN_{\text{Dopp}} N_{\text{delay}} / N_s)^{-1}$.

While achievable-rate bounds and pilot/data power allocation for a *particular* MIMO-MMSE-PAT scheme was obtained in [43], the results here hold for *arbitrary* MIMO-MMSE-PAT schemes.

6.5 Streaming MIMO-MMSE-PAT

In this section, we find the spectral efficiency of a system which transmits a *stream* of blocks, where each block is constructed according to the MIMO-MMSE-PAT principles discussed earlier and separated from its neighbors (by guards) to prevent inter-block interference at the channel output. We pay attention to the MMSE-PAT examples from Section 6.2.1 and the zero-padded MMSE-PAT scheme from [43]. The streaming of blocks is motivated from the requirement to code over many blocks to achieve the ergodic rates described in Section 6.3. Streaming also makes clear the inter-block guard requirements.

We assume that the continuous-time baseband pulse used for modulation is (approximately) time-limited to $[-\frac{\mathcal{T}_s}{2}, \frac{\mathcal{T}_s}{2}]$ seconds and band-limited to $[-\frac{1}{2\mathcal{T}_s}, \frac{1}{2\mathcal{T}_s}]$ Hz, where \mathcal{T}_s denotes the system's sampling interval. To compute the time consumed in block transmission, we include the guard overhead. To compute the bandwidth consumed in block transmission, we measure the bandwidth of the signal at the channel *output*, as would be relevant when considering the egress onto adjacent frequency bands. We assume uncorrelated data symbols (e.g., from an i.i.d. Gaussian codebook) and assume that N is large enough to make the approximation $\frac{1}{2B_{\text{Dopp}}} \approx \frac{N_{\text{Dopp}} - 1}{N\mathcal{T}_s}$ accurate.

For the MIMO-TDKD scheme in Section 6.2.1 constructed with $\ell = \frac{N}{N_{\text{Dopp}}} - TN_{\text{delay}}$, the transmitted block (for every antenna) includes built-in zero-padding, eliminating

the need for an explicit guard. (See Fig. 6.2.) This streaming zero-padded (SZP) MIMO-TDKD scheme is identical to the scheme proposed in [43] and its per-block time and bandwidth consumption are $N\mathcal{T}_s$ seconds and $\frac{1}{\mathcal{T}_s} + 2\mathcal{B}_{\text{Dopp}} \approx \frac{N+N_{\text{Dopp}}-1}{N\mathcal{T}_s}$ Hz. (See Fig. 6.2.) The streaming MIMO-FDKD scheme from Section 6.2.1 constructed with $\ell = \frac{N}{N_{\text{delay}}} - TN_{\text{Dopp}}$, referred as SCP-MIMO-FDKD, has per-block time and bandwidth consumption of $(N + N_{\text{delay}} - 1)\mathcal{T}_s$ seconds and $\frac{N-N_{\text{Dopp}}-1}{N\mathcal{T}_s} + 2\mathcal{B}_{\text{Dopp}} \approx \frac{1}{\mathcal{T}_s}$ Hz, where the bandwidth is less than that of SZP-MIMO-TDKD due to null subcarriers. (See Fig. 6.2.) Since the streaming-chirp schemes does not have a built-in time-domain guard nor null subcarriers, it consumes $(N + N_{\text{delay}} - 1)\mathcal{T}_s$ seconds per block and $\frac{N+N_{\text{Dopp}}-1}{N\mathcal{T}_s}$ Hz. (See Fig. 6.2.) See Chapter 4 for more details on the time and bandwidth consumption of streaming block transmissions over the DSC.

We define SNR as *the ratio of signal power to noise power observed at the output of the pulse-shaping filter associated with any receive antenna*. Because the discrete-time channels between all transmit-receive antenna pairs were assumed to be energy preserving, SNR can be equivalently described as the ratio of total transmit power to per-antenna received noise power. It can be easily verified that SZP-MIMO-TDKD, SCP-MIMO-FDKD and SCP-MIMO-Chirp schemes have the same average transmit power of $\frac{\rho E_{\text{tot}}}{N}$ and hence can be fairly compared under the SNR ρ .

Dividing the per-block achievable rates (6.32), (6.33) and (6.36) by the factor ζ (in units of sec Hz per block) gives the streaming MIMO-MMSE-PAT achievable rates in units of bits/sec/Hz, where $\zeta = N + N_{\text{Dopp}} - 1$, $N + N_{\text{delay}} - 1$, $\frac{(N+N_{\text{delay}}-1)(N+N_{\text{Dopp}}-1)}{N}$ for SZP-MIMO-TDKD, SCP-MIMO-FDKD and SCP-MIMO-Chirp schemes respectively. Denoting the pre-log factor of a bits/sec/Hz asymptotic-rate expression by η , it is straightforward to show that for integer N_{delay} and N_{Dopp} , $N_{\text{delay}} \geq N_{\text{Dopp}} \Leftrightarrow$

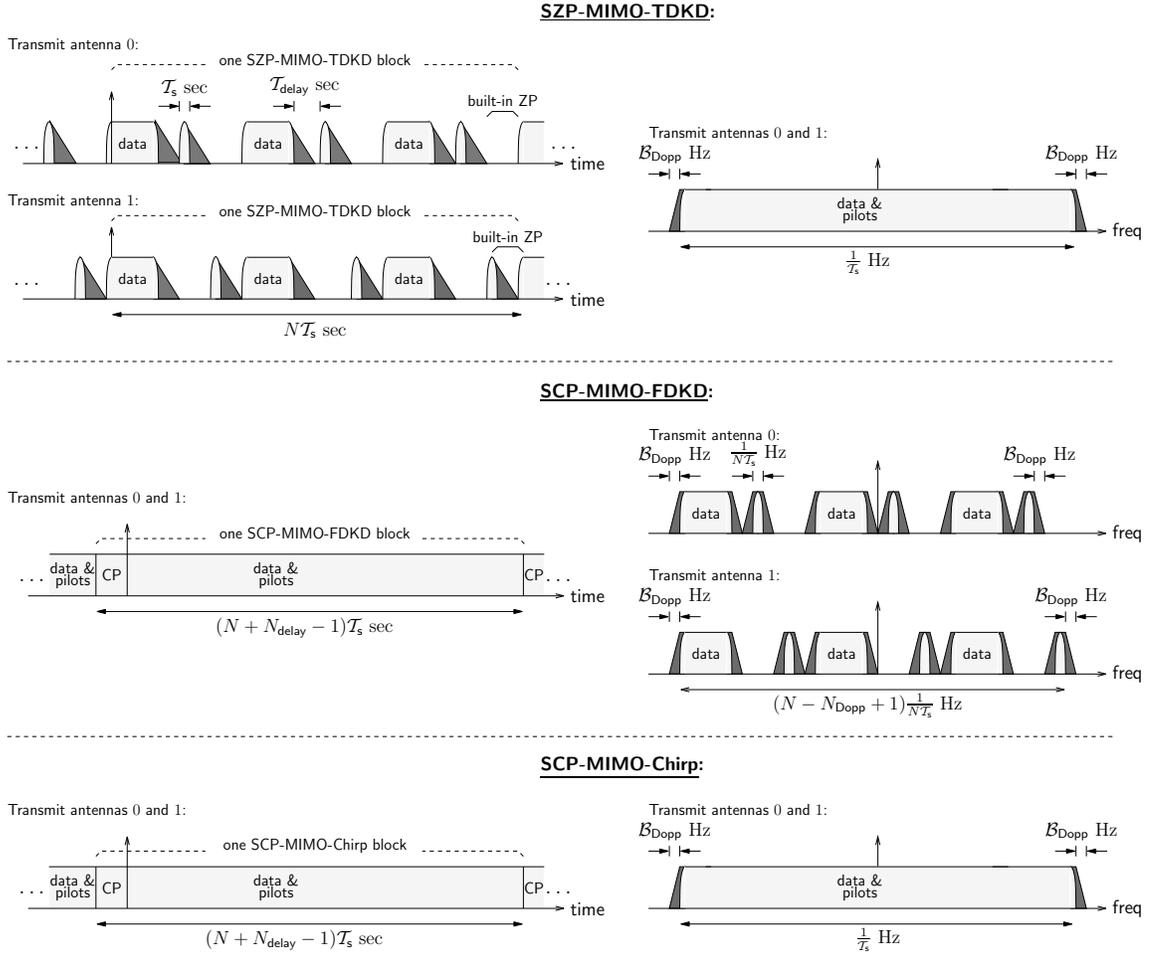


Figure 6.2: Time and bandwidth occupation for several streaming $T = 2$ MIMO-MMSE-PAT systems designed for DSC with $N_{\text{delay}} = 3$ and $N_{\text{Dopp}} = 3$. Lightly shaded shows channel input and darkly shaded shows channel output.

$\eta_{\text{MIMO-FDKD}} \geq \eta_{\text{MIMO-TDKD}}$ and vice versa. In other words, the relative spectral efficiency of SCP-MIMO-FDKD and SZP-MIMO-TDKD (from [43]) depends on the relative time-frequency spread of the channel. It is also easy to see that $\eta_{\text{MIMO-Chirp}}$ is less than both $\eta_{\text{MIMO-TDKD}}$ and $\eta_{\text{MIMO-FDKD}}$ for all integer pairs $(N_{\text{delay}}, N_{\text{Dopp}}) \neq (1, 1)$. Clearly, a higher spectral efficiency translates into a higher achievable rate as SNR grows.

6.6 Numerical Results

In this section, we present numerical examples of the achievable rate of several streaming MIMO-MMSE-PAT schemes. For this purpose, we evaluate the achievable rate bounds derived in Section 6.5 (in units of bits/sec/Hz) for the SZP-MIMO-TDKD, SCP-MIMO-FDKD, and SCP-MIMO-Chirp schemes, using the power allocation procedure described in Section 6.4. In all cases, we consider block size $N = 126$ and $R = 2$ receive antennas, and we plot the bounds over the SNR range of practical interest.

In Figures 6.3 and 6.4, we show results for $T = 2$ and a CE-BEM DSC with $N_{\text{delay}} = 9$ and $N_{\text{Dopp}} = 3$. These discrete spreading parameters correspond to a spreading index of $\gamma \approx 0.07$, as would result from, e.g., carrier frequency $f_c = 80$ GHz, sampling interval $\mathcal{T}_s = 1 \mu\text{sec}$, maximum mobile velocity $v_{\text{max}} = 150$ km/hr, and a channel delay spread of $9 \mu\text{sec}$. Figure 6.3 corresponds to the case of uniform decay and uniform Doppler profile such that $\mathbf{R}_\lambda = \frac{N}{N_{\text{Dopp}}N_{\text{delay}}}\mathbf{I}$. There we employed the power allocation fraction given in (6.40). Figure 6.4 corresponds to exponential decay and Jakes Doppler profile such that $\text{E}\{|\lambda(k; \ell)|^2\} = \chi e^{-0.1\ell}(\mathcal{B}_{\text{Dopp}}^2 - k^2(N\mathcal{T}_s)^{-2})^{-0.5}$, where χ is chosen such that $\text{tr}\{\mathbf{R}_\lambda\} = N$. There we employed equal power allocation

between pilots and data. In both Fig. 6.3 and Fig. 6.4, we observe that SCP-MIMO-FDKD achieves a higher rate than do SZP-MIMO-TDKD and SCP-MIMO-Chirp schemes in the high SNR regime, since $N_{\text{delay}} > N_{\text{Dopp}}$. In Fig. 6.5, for the same channel parameters with i.i.d. BEM coefficients, we study the performance of SCP-MIMO-FDKD with $T \in \{1, 2, 3\}$ transmit antennas. We find that the high-SNR achievable-rate is maximized when $T = 2$, coinciding with T_* from (6.38).

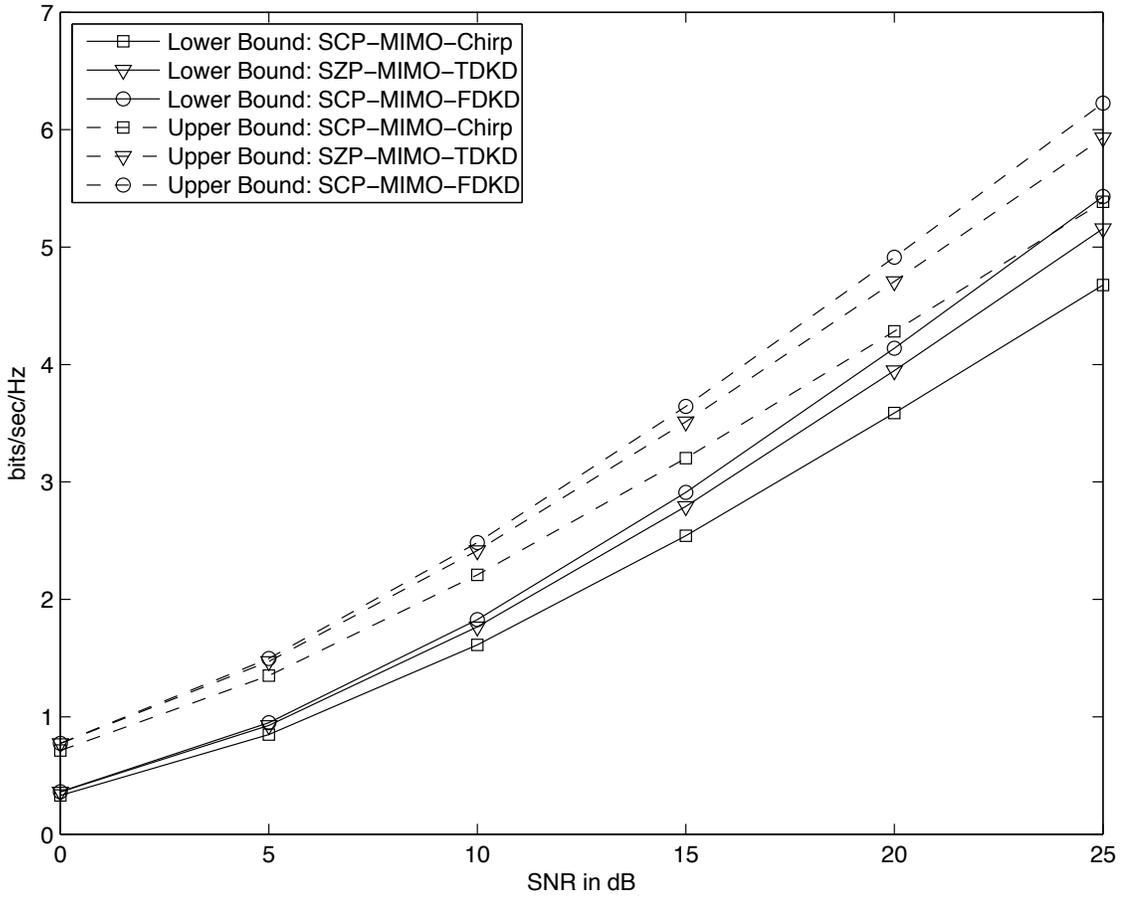


Figure 6.3: Achievable rates for $T = 2$, $R = 2$ MIMO system over a $N_{\text{delay}} = 9$, $N_{\text{Dopp}} = 3$ DSC with uniform decay and Doppler profile.

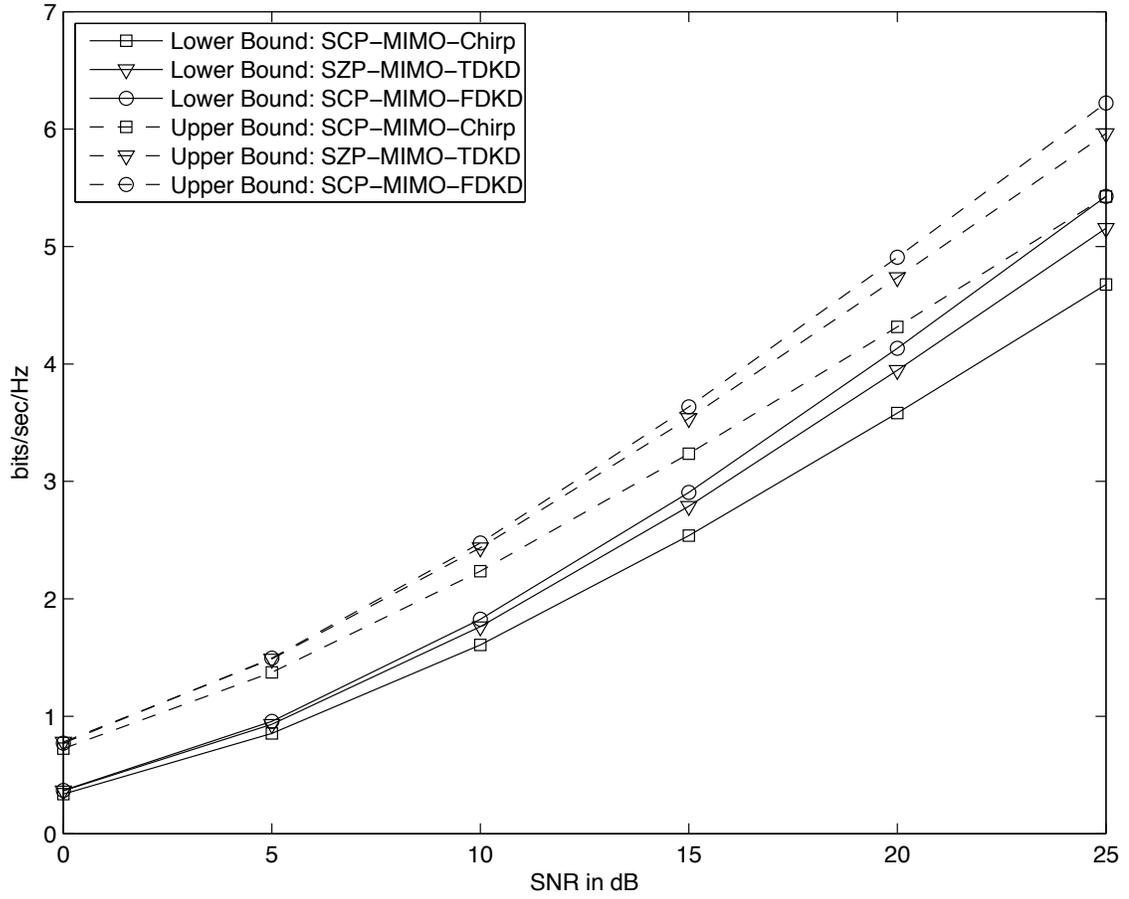


Figure 6.4: Achievable rates for $T = 2$, $R = 2$ MIMO system over a $N_{\text{delay}} = 9$, $N_{\text{Dopp}} = 3$ DSC with exponential decay and Jakes Doppler profile.

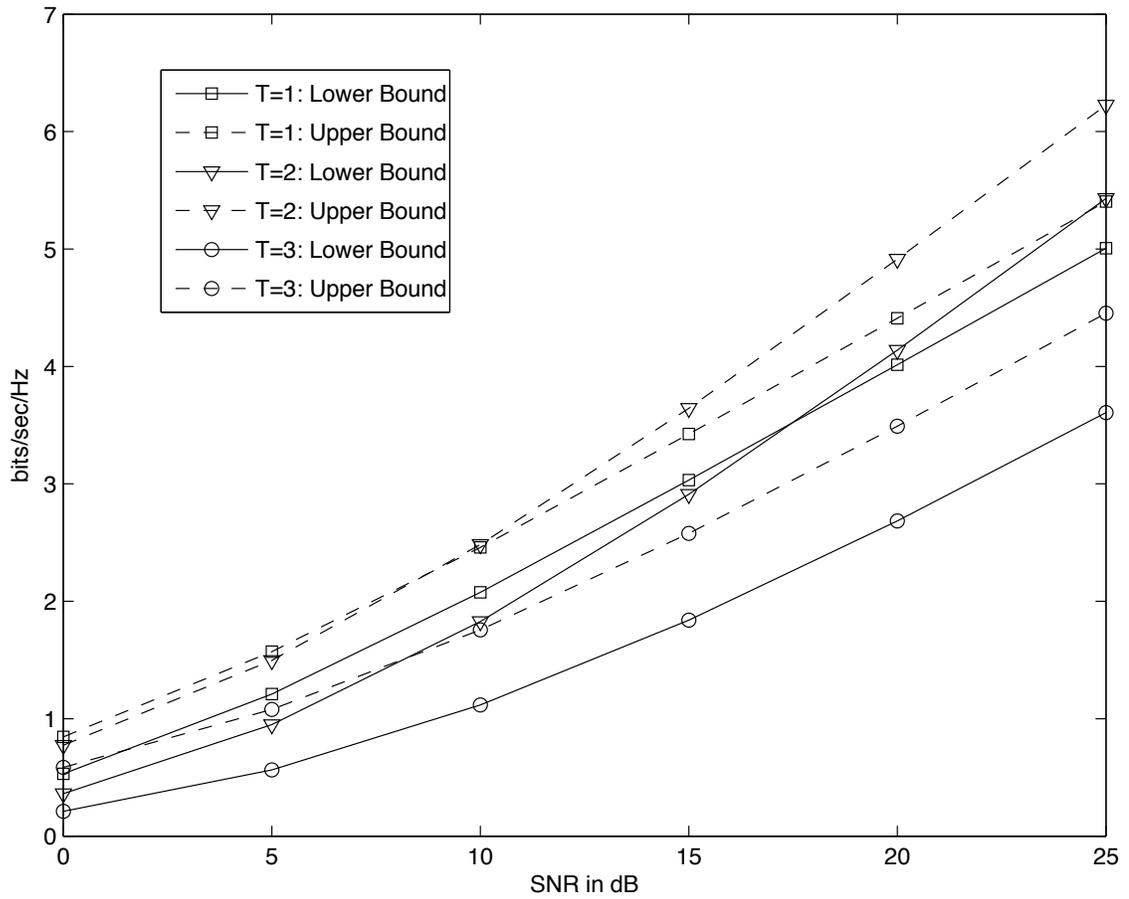


Figure 6.5: Achievable rates of SCP-MIMO-FDKD with $T \in \{1, 2, 3\}$ transmit antennas and $R = 2$ receive antennas.

CHAPTER 7

CONCLUSIONS

In this dissertation, we considered the problem of designing communication systems for noncoherent doubly selective channels and analyzed their performance in the high-SNR regime. In the high-SNR asymptotic regime, we characterized the prelog factor of the channel capacity, when the channel input is constrained to be a *continuous* random vector. We also addressed the design of pilot aided transmissions for this channel based on two different criteria: minimize the MSE of channel estimates and maximize the prelog factor of achievable rates. In this chapter, we present a summary of the original work in this dissertation and also give possible directions for future research.

7.1 Summary of Original Work

We characterized the noncoherent doubly selective channel's capacity in the high SNR regime under the constraint that the channel input is a continuous random vector. For CE-BEM DSC, we establish that the constrained capacity grows logarithmically with SNR with a slope of $1 - \gamma$ where the spreading index $\gamma = N_{\text{Dopp}}N_{\text{delay}}/N$ is related to the delay spread and the Doppler spread of the DSC.

We designed PAT schemes based on minimizing MSE criterion. We obtained necessary and sufficient conditions for the pilot/data pattern to attain the minimal MSE given a constraint on the pilot energy. The conditions require that the pilot and data should not “interfere” at the channel output. Also, the pilot signal has to be chosen to excite all the uncorrelated channel modes in an orthogonal manner. We also established that the MMSE-PAT schemes occur in time-frequency dual pairs. Based on these results, we designed novel MMSE-PAT structures for DSC. Analyzing the ergodic achievable rates of MMSE-PAT schemes, we establish that a frequency domain MMSE-PAT (FDKD) outperforms a time-domain MMSE-PAT (TDKD) from [13], when the DSC’s discrete delay spread N_{delay} dominates its discrete Doppler spread N_{Dopp} . We proved that the prelog factor of the ergodic asymptotic achievable rates of all the MMSE-PAT schemes is strictly less than that of constrained capacity, for strictly doubly selective channels (i.e., $N_{\text{delay}} > 1$, $N_{\text{Dopp}} > 1$).

We considered design of PAT schemes whose rates achieve the optimal prelog factor $1-\gamma$. We presented sufficient conditions on the pilot/data pattern to achieve the optimal prelog factor and gave a novel spectrally efficient PAT scheme satisfying those requirements. We also compared the performance of the SE-PAT and MMSE-PAT schemes through simulations. Agreeing with our theoretical results, our numerical results showed that the SE-PAT scheme achieves higher rates than the MMSE-PAT schemes if the SNR is high enough. We also found that, in the moderate SNR regime, MMSE-PAT schemes may achieve higher rates than SE-PAT. We theoretically computed approximate thresholds on the SNR values to find the regions in which a particular PAT scheme achieves rates higher than others.

We also extended the MMSE-PAT design for MIMO doubly selective channels. In the MIMO case, we showed that, for independent fading between different antenna pairs, the number of receive antennas does not affect the design of MIMO-MMSE-PAT for DSC. We also established that the spectral efficiency of MIMO-MMSE-PAT does not necessarily increase even if the number of transmit and receive antennas are increased simultaneously. We also presented the optimal number of active antennas which maximizes the spectral efficiency, for the given values of system parameters (block size N) and the channel parameters N_{delay} and N_{Dopp} .

7.2 Possible Future Research

Previous works on characterization of the capacity of noncoherent channels focused on low SNR or high SNR regime. The characterization of channel capacity in the intermediate finite SNR regime is still an open problem even for simple flat fading channel models.

Most of the results in multi-user wireless networks are based on the assumption that the channel state information is available at either the transmitter or the receiver. How does the lack of CSI affects the capacity and the achievable rates in those scenarios is an interesting problem to investigate. In the single user case, we saw that there is a “pre-log factor loss” proportional to number of unknown channel parameters. In the multi-user scenario, with perfect apriori CSI, the performance of opportunistic communication techniques (which allocates the time/bandwidth resources to the user having the best channel) improves with the number of users. But, in the noncoherent setting, how to design opportunistic communication schemes and how they perform with the increase in the number of users is a possible research direction.

In the high SNR regime, for SISO channels (flat, time selective, frequency selective or doubly selective) the PAT schemes are required to sacrifice signaling dimensions whose number is at least the number of unknown channel parameters to avoid the error floor in the decoder which treats the channel estimation error as a part of effective noise. But, in the moderate SNR regime, if some of the fading coefficients have very small variance, then it might not be “worth” sacrificing some signaling dimensions to estimate them. Analyzing the effect of SNR and the variance of the unknown channel fading coefficients jointly is a possible future research.

All our results in this dissertation are based on the CE-BEM approximation of doubly selective block fading DSC. In general, when the block size is relatively small, due to the leakage effects of rectangular windowing, there will be some small variance inter-carrier interference coefficients in the frequency domain. How does these small coefficients affect our results is also a research possibility.

APPENDIX A

DERIVATION OF CHANNEL FADING STATISTICS

A.1 Discrete-Time Channel Statistics

Equations (2.6) and (2.2) can be used to obtain

$$\begin{aligned}
& \mathbb{E}\{h[n; l]h^*[n - p; l - q]\} \\
&= \int \int \int \int \psi^*(t)\psi(t - \tau)\psi(t')\psi^*(t' - \tau') \\
&\quad \mathbb{E}\{h(t + n\mathcal{T}_s; \tau + l\mathcal{T}_s)h^*(t' + n\mathcal{T}_s - p\mathcal{T}_s; \tau' + l\mathcal{T}_s - q\mathcal{T}_s)\} dt d\tau dt' d\tau' \\
&= \int \int \int \int \psi^*(t)\psi(t - \tau)\psi(t')\psi^*(t' - \tau') \\
&\quad R_{\text{lag;delay}}(t - t' + p\mathcal{T}_s; \tau' + l\mathcal{T}_s)\delta(\tau - \tau' + q\mathcal{T}_s) dt d\tau dt' d\tau' \\
&= \int \int \int \psi^*(t)\psi(t - \tau)\psi(t')\psi^*(t' - \tau + q\mathcal{T}_s) \\
&\quad R_{\text{lag;delay}}(t - t' + p\mathcal{T}_s; \tau - q\mathcal{T}_s + l\mathcal{T}_s) dt d\tau dt'.
\end{aligned}$$

If we assume that the coherence time is larger than \mathcal{T}_s , then the limited time-support of $\psi(t)$ implies that $\psi(t)R_{\text{lag;delay}}(t_o + t; \tau) \approx \psi(t)R_{\text{lag;delay}}(t_o; \tau)$, in which case

$$\begin{aligned}
& \mathbb{E}\{h[n; l]h^*[n - p; l - q]\} \\
&\approx \int \int \int \psi^*(t)\psi(t - \tau)\psi(t')\psi^*(t' - \tau + q\mathcal{T}_s)R_{\text{lag;delay}}(p\mathcal{T}_s; \tau - q\mathcal{T}_s + l\mathcal{T}_s) dt d\tau dt' \\
&= \int A^*(\tau)A(\tau - q\mathcal{T}_s)R_{\text{lag;delay}}(p\mathcal{T}_s; \tau - q\mathcal{T}_s + l\mathcal{T}_s) d\tau
\end{aligned}$$

for $A(\tau) = \int \psi(t)\psi^*(t - \tau)dt$ such that $A(0) = 1$ and $\int |A(\tau)|^2 d\tau = C_\psi$. Due to the limited time-support of $A(\tau)$, we have

$$\mathbb{E}\{h[n; l]h^*[n - p; l - q]\} \approx \delta[q] \int |A(\tau)|^2 R_{\text{lag;delay}}(p\mathcal{T}_s; \tau + l\mathcal{T}_s) d\tau \quad (\text{A.1})$$

$$\approx R_{\text{lag;delay}}(p\mathcal{T}_s; l\mathcal{T}_s) \delta[q] C_\psi, \quad (\text{A.2})$$

where the second approximation holds when $R_{\text{lag;delay}}(t_o; \tau)$ is a smooth function of τ .

A.2 Basis-Expansion Statistics

From (2.8), we know $\lambda[k; l] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} h[n; l] e^{-j\frac{2\pi}{N}nk}$ for $k \in \{0, \dots, N-1\}$, so that

$$\begin{aligned} & \mathbb{E}\{\lambda[k; l]\lambda^*[k - p; l - q]\} \\ &= \frac{1}{N} \sum_{n=0}^{N-1} \sum_{n'=0}^{N-1} \mathbb{E}\{h[n; l]h^*[n'; l - q]\} e^{-j\frac{2\pi}{N}[nk - n'(k-p)]} \end{aligned} \quad (\text{A.3})$$

$$\approx \delta[q] \frac{1}{N} \sum_{n=0}^{N-1} \sum_{n'=0}^{N-1} C_\psi R_{\text{lag;delay}}((n - n')\mathcal{T}_s; l\mathcal{T}_s) e^{-j\frac{2\pi}{N}[(n - n')k + n'p]} \quad (\text{A.4})$$

$$= C_\psi \delta[q] \sum_{m=-N+1}^{N-1} |N - m| R_{\text{lag;delay}}(m\mathcal{T}_s; l\mathcal{T}_s) e^{-j\frac{2\pi}{N}mk} \frac{1}{N} \sum_{n'=0}^{N-1} e^{-j\frac{2\pi}{N}n'p} \quad (\text{A.5})$$

$$= C_\psi \delta[q] \delta[p] \sum_{m=-N+1}^{N-1} |N - m| R_{\text{lag;delay}}(m\mathcal{T}_s; l\mathcal{T}_s) e^{-j\frac{2\pi}{N}mk} \quad (\text{A.6})$$

where we have used (2.7) for the approximation in (A.4). Recalling that $R_{\text{lag;delay}}(t, \cdot) = \int R_{\text{Dopp;delay}}(f; \cdot) e^{j2\pi ft} df$,

$$\begin{aligned} & \mathbb{E}\{\lambda[k; l]\lambda^*[k - p; l - q]\} \\ &= C_\psi \delta[q] \delta[p] \sum_{m=-N+1}^{N-1} (N - |m|) \int R_{\text{Dopp;delay}}(f; l\mathcal{T}_s) e^{j2\pi m\mathcal{T}_s(f - \frac{k}{N\mathcal{T}_s})} df \end{aligned} \quad (\text{A.7})$$

$$= C_\psi \delta[q] \delta[p] \int R_{\text{Dopp;delay}}(f' + \frac{k}{N\mathcal{T}_s}; l\mathcal{T}_s) \left(\frac{\sin(\pi f' \mathcal{T}_s N)}{\sin(\pi f' \mathcal{T}_s)} \right)^2 df'. \quad (\text{A.8})$$

where for (A.8) we used $\sum_{m=-N+1}^{N-1} (N - |m|) e^{jm\phi} = \left(\frac{\sin(\phi N/2)}{\sin(\phi/2)} \right)^2$.

APPENDIX B

PROOFS OF ASYMPTOTIC CHANNEL CAPACITY RESULTS

B.1 Proof of Theorem 1

Defining the two vectors

$$\mathbf{y}_s = [y[0], \dots, y[N_{\text{Dopp}}N_{\text{delay}} - 1]]^\top \quad (\text{B.1})$$

$$\mathbf{y}_r = [y[N_{\text{Dopp}}N_{\text{delay}}], \dots, y[N - 1]]^\top, \quad (\text{B.2})$$

and using the chain rule for mutual information [51], we have $I(\mathbf{y}; \mathbf{x}^\rho) = I(\mathbf{y}_s; \mathbf{x}^\rho) + I(\mathbf{y}_r; \mathbf{x}^\rho | \mathbf{y}_s)$. Using $h(\cdot)$ to denote the differential entropy, we have

$$I(\mathbf{y}_r; \mathbf{x}^\rho | \mathbf{y}_s) = h(\mathbf{y}_r | \mathbf{y}_s) - h(\mathbf{y}_r | \mathbf{y}_s, \mathbf{x}^\rho) \quad (\text{B.3})$$

$$\leq h(\mathbf{y}_r) - h(\mathbf{y}_r | \mathbf{y}_s, \mathbf{x}^\rho, \mathbf{H}) \quad (\text{B.4})$$

since conditioning reduces entropy. Because of the power constraint on the input, the covariance of \mathbf{y}_r , $\mathbf{R}_{\mathbf{y}_r}$ can be bounded as $\mathbf{R}_{\mathbf{y}_r} \leq (k\rho + 1)\mathbf{I}_{N - N_{\text{delay}}N_{\text{Dopp}}}$ for some constant $k \in \mathbb{R}$ and k is independent of ρ . Since Gaussian distribution maximizes the entropy for a given covariance matrix, we have $h(\mathbf{y}_r) \leq \log \det[(k\rho + 1)\mathbf{I}_{N - N_{\text{delay}}N_{\text{Dopp}}}]$. Now, $h(\mathbf{y}_r | \mathbf{y}_s, \mathbf{x}^\rho, \mathbf{H})$ is equal to the entropy of the unit variance white noise term

in \mathbf{y}_r , which is bounded and independent of ρ . So, we have $\lim_{\rho \rightarrow \infty} \frac{I(\mathbf{y}_r; \mathbf{x}^\rho | \mathbf{y}_s)}{\log \rho} = N - N_{\text{Dopp}} N_{\text{delay}}$.

What remains to be shown is that $\lim_{\rho \rightarrow \infty} \frac{I(\mathbf{y}_s; \mathbf{x}^\rho)}{\log \rho} = 0$. Using the chain rule for mutual information again, we have

$$I(\mathbf{y}_s; \mathbf{x}^\rho) = I(y[0]; \mathbf{x}^\rho) + \sum_{i=1}^{N_{\text{Dopp}} N_{\text{delay}} - 1} I(y[i]; \mathbf{x}^\rho | y[0], \dots, y[i-1]) \quad (\text{B.5})$$

$$\leq I(y[0]; \mathbf{x}^\rho) + \sum_{i=1}^{N_{\text{Dopp}} N_{\text{delay}} - 1} I(y[i]; \mathbf{x}^\rho, y[0], \dots, y[i-1]). \quad (\text{B.6})$$

We shall analyze each term in (B.6) separately. We define the vectors $\mathbf{x}_i^\rho = [x^\rho[i], \dots, x^\rho[i - N_{\text{delay}} + 1]]^\top$ and their ‘‘complements’’ $\bar{\mathbf{x}}_i^\rho$, which are composed of elements of \mathbf{x}^ρ not in \mathbf{x}_i^ρ . With these definitions, the first term in (B.6) can be written $I(y[0]; \mathbf{x}^\rho) = I(y[0]; \mathbf{x}_0^\rho) + I(y[0]; \bar{\mathbf{x}}_0^\rho | \mathbf{x}_0^\rho)$. Conditioned on \mathbf{x}_0^ρ , the uncertainty in $y[0]$ is due to channel coefficients and additive noise which are independent of $\bar{\mathbf{x}}_0^\rho$. Hence, $I(y[0]; \bar{\mathbf{x}}_0^\rho | \mathbf{x}_0^\rho) = 0$. Now, $I(y[0]; \mathbf{x}_0^\rho)$ corresponds to a overspread channel (i.e., one observation with N_{delay} unknown channel coefficients) and, using the result from [52], we have $I(y[0]; \mathbf{x}_0^\rho) \leq \log \log \rho + O(1)$. Hence $\lim_{\rho \rightarrow \infty} \frac{I(y[0]; \mathbf{x}^\rho)}{\log \rho} = 0$. Now considering the general term inside the summation of (B.6),

$$\begin{aligned} & I(y[i]; \mathbf{x}^\rho, y[0], \dots, y[i-1]) \\ &= \underbrace{I(y[i]; \mathbf{x}_i^\rho)}_{\leq \log \log \rho + O(1)} + \underbrace{I(y[i]; \bar{\mathbf{x}}_i^\rho | \mathbf{x}_i^\rho)}_{=0} + \underbrace{I(y[i]; y[0], \dots, y[i-1] | \mathbf{x}^\rho)}_{T_i}, \end{aligned}$$

it remains to be shown that $\lim_{\rho \rightarrow \infty} \frac{T_i}{\log \rho} = 0$.

Recall that \mathbf{y} and \mathbf{h} are jointly Gaussian conditioned on \mathbf{x}^ρ . In terms of differential entropies, $I(y[i]; y[0], \dots, y[i-1] | \mathbf{x}^\rho) = h(y[i] | \mathbf{x}^\rho) - h(y[i] | \mathbf{x}^\rho, y[0], \dots, y[i-1])$. It easily follows that

$$h(y[i] | \mathbf{x}^\rho) = \mathbb{E} \left\{ \log(1 + \rho \sum_{\ell=0}^{N_{\text{delay}} - 1} \mathbb{E} \{ |h[i; \ell]|^2 \} | x^\rho[i - \ell]|^2) \right\}, \quad (\text{B.7})$$

where the expectation is with respect to \mathbf{x}^ρ . Now, given $\{y[0], \dots, y[i-1]\}$, we split $y[i]$ into MMSE estimate and error as $y[i] = \mathbb{E}\{y[i]|y[0], \dots, y[i-1]\} + \tilde{y}[i]$, and we have $\mathbb{h}(y[i]|y[0], \dots, y[i-1], \mathbf{x}^\rho) = \mathbb{E} \log(\mathbb{E}|\tilde{y}[i]|^2)$. Defining $\mathbf{h}_i = [h[i; 0], \dots, h[i; N_{\text{delay}} - 1]]^\top$ and denoting the covariance of $\mathbf{h}_i - \mathbb{E}\{\mathbf{h}_i|y[0], \dots, y[i-1]\}$ by $\tilde{\mathbf{R}}_i$, we have $\mathbb{E}|\tilde{y}[i]|^2 = 1 + \rho \mathbf{x}_i^{\rho \text{H}} \tilde{\mathbf{R}}_i \mathbf{x}_i^\rho$. Let $\mu_{\max, i}$ denote the maximum eigenvalue of $\tilde{\mathbf{R}}_i$ and \mathbf{q}_i denote the corresponding eigenvector. Now define $\kappa_{\max, i} = \inf_{\mathbf{x}^\rho \in \mathbb{C}^N} \mu_{\max, i}$. For $i \in \{1, \dots, N_{\text{Dopp}} N_{\text{delay}} - 1\}$, all the elements of \mathbf{h}_i can not be estimated perfectly, even in the absence of noise ($\rho = \infty$), since $\{y[0], \dots, y[i-1]\}$ correspond to a projection of $\boldsymbol{\lambda}$ onto a subspace of smaller dimension, and hence $\kappa_{\max, i} > 0$. Now, $\mathbb{E}|\tilde{y}[i]|^2 \geq 1 + \rho \kappa_{\max, i} |\sum_{k=0}^{N_{\text{delay}}-1} q_i[k] x^\rho[i-k]|^2$, and hence

$$\mathbb{h}(y[i]|\mathbf{x}^\rho, y[0], \dots, y[i-1]) \geq \mathbb{E}\{\log(1 + \rho \kappa_{\max, i} |\sum_{k=0}^{N_{\text{delay}}-1} q_i[k] x^\rho[i-k]|^2)\}. \quad (\text{B.8})$$

Combining (B.7) and (B.8), we have $T_i \leq \mathbb{E} \log \frac{1 + \rho \sum_{\ell=0}^{N_{\text{delay}}-1} \mathbb{E}\{|h[i; \ell]|^2\} |x^\rho[i-\ell]|^2}{1 + \rho \kappa_{\max, i} |\sum_{k=0}^{N_{\text{delay}}-1} q_i[k] x^\rho[i-k]|^2}$. Since \mathbf{x}^ρ is a sequence of continuous random vectors converging to a continuous random vector, $\lim_{\rho \rightarrow \infty} |\sum_{k=0}^{N_{\text{delay}}-1} q_i[k] x^\rho[i-k]|^2 > 0$ with probability 1, and $\lim_{\rho \rightarrow \infty} \frac{T_i}{\log \rho} = 0$.

B.2 Proof of Lemma 1

Using the chain rule for mutual information, we have

$$I(\mathbf{y}; \mathbf{x}) = I(\mathbf{y}; \mathbf{x}, \mathbf{H}) - I(\mathbf{y}; \mathbf{H}|\mathbf{x}) \quad (\text{B.9})$$

$$\geq I(\mathbf{y}; \mathbf{x}|\mathbf{H}) - I(\mathbf{y}; \mathbf{H}|\mathbf{x}). \quad (\text{B.10})$$

Now, $I(\mathbf{y}; \mathbf{x}|\mathbf{H})$ corresponds to coherent case of perfect receiver CSI and since \mathbf{x} is i.i.d. Gaussian, we have

$$I(\mathbf{y}; \mathbf{x}|\mathbf{H}) = \mathbb{E}\{\log \det[\mathbf{I}_N + \rho \mathbf{H} \mathbf{H} \mathbf{H}^\text{H}]\}. \quad (\text{B.11})$$

Since $\mathbf{H}\mathbf{H}^H$ is full rank with probability 1 with continuously distributed eigen values, we have

$$\lim_{\rho \rightarrow \infty} \frac{I(\mathbf{y}; \mathbf{x}|\mathbf{H})}{\log \rho} = N. \quad (\text{B.12})$$

Recall that from (2.21), the output is written as $\mathbf{y} = \sqrt{\rho}\mathbf{X}\mathbf{U}\boldsymbol{\lambda} + \mathbf{v}$. Since $\boldsymbol{\lambda}$ captures all the degrees of freedom of DSC over a block, we have $I(\mathbf{y}; \mathbf{H}|\mathbf{x}) = I(\mathbf{y}; \boldsymbol{\lambda}|\mathbf{x}) = I(\mathbf{y}; \boldsymbol{\lambda}|\mathbf{X})$. Conditioned on \mathbf{X} , \mathbf{y} and $\boldsymbol{\lambda}$ are jointly Gaussian and using Jensen's inequality, we have

$$I(\mathbf{y}; \boldsymbol{\lambda}|\mathbf{X}) = \mathbb{E}\{\log \det[\mathbf{I} + \rho(\mathbf{X}\mathbf{U})^H \mathbf{X}\mathbf{U}\mathbf{R}_\lambda]\} \quad (\text{B.13})$$

$$\leq \log \det \mathbb{E}\{[\mathbf{I} + \rho(\mathbf{X}\mathbf{U})^H \mathbf{X}\mathbf{U}\mathbf{R}_\lambda]\} \quad (\text{B.14})$$

$$\leq N_{\text{Dopp}}N_{\text{delay}} \log \rho + O(1), \quad (\text{B.15})$$

and hence

$$\lim_{\rho \rightarrow \infty} \frac{I(\mathbf{y}; \mathbf{H}|\mathbf{x})}{\log \rho} \leq N_{\text{Dopp}}N_{\text{delay}}. \quad (\text{B.16})$$

The desired result follows from (B.10), (B.12) and (B.16).

APPENDIX C

PROOFS OF MMSE-PAT RESULTS

C.1 Proof of Theorem 2

We begin with the singular value decompositions $\mathbf{P}\mathbf{U} = \mathbf{V}_p \boldsymbol{\Sigma}_p \mathbf{Q}_p^H$ and $\mathbf{D}\mathbf{U} = \mathbf{V}_d \boldsymbol{\Sigma}_d \mathbf{Q}_d^H$, where $\boldsymbol{\Sigma}_p$ and $\boldsymbol{\Sigma}_d$ are diagonal and full-rank [66]. Let $K \leq N_{\text{Dopp}} N_{\text{delay}}$ denote the rank of $\boldsymbol{\Sigma}_p$. Defining $\mathbf{z} = \mathbf{V}_p^H \mathbf{y}$ and using (4.2),

$$\mathbf{z} = \underbrace{\sqrt{\rho} \boldsymbol{\Sigma}_p \mathbf{Q}_p^H}_{\mathbf{A}_p} \boldsymbol{\lambda} + \underbrace{\sqrt{\rho} \mathbf{V}_p^H \mathbf{V}_d \boldsymbol{\Sigma}_d \mathbf{Q}_d^H}_{\mathbf{A}_d} \boldsymbol{\lambda} + \underbrace{\mathbf{V}_p^H \mathbf{v}}_n. \quad (\text{C.1})$$

Since projection onto $\text{col}(\mathbf{V}_p)$ does not attenuate the pilot component of \mathbf{y} , the pilot-aided MMSE channel estimate given $\{\mathbf{y}, \mathbf{P}\}$ is equal to that given $\{\mathbf{z}, \mathbf{P}\}$. With $\mathbf{R}_{z,\lambda} = E\{\mathbf{z}\boldsymbol{\lambda}^H\}$ and $\mathbf{R}_z = E\{\mathbf{z}\mathbf{z}^H\}$, the MMSE estimate of $\boldsymbol{\lambda}$ given $\{\mathbf{z}, \mathbf{P}\}$ is

$$\hat{\boldsymbol{\lambda}} = \mathbf{R}_{z,\lambda}^H \mathbf{R}_z^{-1} \mathbf{z}, \quad (\text{C.2})$$

$$\mathbf{R}_{z,\lambda} = \sqrt{\rho} \mathbf{A}_p \mathbf{R}_\lambda, \quad (\text{C.3})$$

$$\mathbf{R}_z = \underbrace{\rho \mathbf{A}_p \mathbf{R}_\lambda \mathbf{A}_p^H}_{\boldsymbol{\Delta}} + \underbrace{\rho E\{\mathbf{A}_d \mathbf{R}_\lambda \mathbf{A}_d^H\}}_{\mathbf{U}_d \boldsymbol{\Lambda}_d \mathbf{U}_d^H}, \quad (\text{C.4})$$

with diagonal $\boldsymbol{\Lambda}_d \geq 0$ and $\mathbf{U}_d^H \mathbf{U}_d = \mathbf{I}$. Note that the MMSE estimate of \mathbf{h} is $\hat{\mathbf{h}} = \mathbf{U} \hat{\boldsymbol{\lambda}}$ and that $\sigma_e^2 = E\{\|\mathbf{h} - \hat{\mathbf{h}}\|^2\} = E\{\|\boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}}\|^2\}$. We have $\mathbf{R}_{\hat{\boldsymbol{\lambda}}} = E\{(\boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}})(\boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}})^H\} =$

$\mathbf{R}_\lambda - \mathbf{R}_{z,\lambda}^H \mathbf{R}_z^{-1} \mathbf{R}_{z,\lambda}$ and hence

$$\begin{aligned}
\sigma_e^2 &= \text{tr}\{\mathbf{R}_\lambda - \mathbf{R}_{z,\lambda}^H \mathbf{R}_z^{-1} \mathbf{R}_{z,\lambda}\} \\
&= \text{tr}\{\mathbf{R}_\lambda - \mathbf{R}_\lambda^H \mathbf{A}_p^H (\mathbf{\Delta} + \mathbf{U}_d \mathbf{\Lambda}_d \mathbf{U}_d^H)^{-1} \mathbf{A}_p \mathbf{R}_\lambda\} \\
&= \text{tr}\{\mathbf{R}_\lambda - \mathbf{R}_\lambda^H \mathbf{A}_p^H [\mathbf{\Delta}^{-1} - \mathbf{\Delta}^{-1} \mathbf{U}_d (\mathbf{\Lambda}_d^{-1} \\
&\quad + \mathbf{U}_d^H \mathbf{\Delta}^{-1} \mathbf{U}_d)^{-1} \mathbf{U}_d^H \mathbf{\Delta}^{-1}] \mathbf{A}_p \mathbf{R}_\lambda\}, \tag{C.5}
\end{aligned}$$

$$\geq \text{tr}\{\mathbf{R}_\lambda - \mathbf{R}_\lambda^H \mathbf{A}_p^H \mathbf{\Delta}^{-1} \mathbf{A}_p \mathbf{R}_\lambda\}, \tag{C.6}$$

where we used the matrix inversion lemma in (C.5). The inequality (C.6) follows since $\mathbf{\Delta} > 0$ and $\mathbf{\Lambda}_d \geq 0$. Since $\mathbf{\Sigma}_p$ is full rank, \mathbf{A}_p has full row rank, and so equality in (C.6) is achieved if and only if

$$\mathbf{U}_d \mathbf{\Lambda}_d \mathbf{U}_d^H = \mathbf{0} \Leftrightarrow E\{\mathbf{A}_d \mathbf{R}_\lambda \mathbf{A}_d^H\} = \mathbf{0}. \tag{C.7}$$

Since $\mathbf{R}_\lambda > 0$, (C.7) is satisfied if and only if $\mathbf{A}_d = \mathbf{0}$, which is equivalent to (4.7), since $\mathbf{\Sigma}_p$ and $\mathbf{\Sigma}_d$ are full rank square matrices. We proceed further assuming that (4.7) is satisfied. With $\mathbf{A}_d = \mathbf{0}$,

$$\begin{aligned}
\sigma_e^2 &= \text{tr}\{\mathbf{R}_\lambda - \mathbf{R}_\lambda^H \mathbf{A}_p^H (\mathbf{A}_p \mathbf{R}_\lambda \mathbf{A}_p^H + \frac{1}{\rho} \mathbf{I}_K)^{-1} \mathbf{A}_p \mathbf{R}_\lambda\}, \\
&= \text{tr}\{(\mathbf{R}_\lambda^{-1} + \rho \mathbf{A}_p^H \mathbf{A}_p)^{-1}\} \tag{C.8}
\end{aligned}$$

using the matrix inversion lemma [67]. With $\sigma_{\lambda_m}^2 = [\mathbf{R}_\lambda]_{m,m}$, diagonal \mathbf{R}_λ implies

$$\sigma_e^2 \geq \sum_{m=0}^{N_{\text{Dopp}} N_{\text{delay}} - 1} \left(\frac{1}{\sigma_{\lambda_m}^2} + \rho \alpha_m \right)^{-1}, \tag{C.9}$$

where $\alpha_m = [\mathbf{A}_p^H \mathbf{A}_p]_{m,m}$ and equality in (C.9) is achieved if and only if $\mathbf{A}_p^H \mathbf{A}_p = (\mathbf{P}\mathbf{U})^H \mathbf{P}\mathbf{U}$ is diagonal. (See [66, 68] for additional details.) Since all the columns of $\mathbf{P}\mathbf{U}$ have norm $\frac{E_p}{N}$, the bound (C.9) is written as (4.6). When (4.7) and (4.8) are met, it follows straightforwardly that $\mathbf{R}_{\hat{h}}$ and $\mathbf{R}_{\hat{h}}$ are given by (4.9) and (4.10), respectively.

C.2 Proof of Lemma 2

$(\mathbf{P}\mathbf{U})^H\mathbf{P}\mathbf{U}$ is composed of $N_{\text{Dopp}} \times N_{\text{Dopp}}$ blocks $\bar{\mathbf{P}}_{k_2, k_1} = \bar{\mathbf{F}}^H \mathbf{P}_{-k_2}^H \mathbf{P}_{-k_1} \bar{\mathbf{F}}$ for $k_1, k_2 \in \{0, \dots, N_{\text{delay}} - 1\}$. For these k_1, k_2 and for $m_1, m_2 \in \{0, \dots, N_{\text{Dopp}} - 1\}$, (4.8) becomes

$$[\bar{\mathbf{P}}_{k_2, k_1}]_{m_1, m_2} = \frac{E_p}{N} \delta[k_1 - k_2] \delta[m_1 - m_2]. \quad (\text{C.10})$$

The definitions of $\bar{\mathbf{F}}$ and \mathbf{P}_{-i} imply

$$[\bar{\mathbf{P}}_{k_2, k_1}]_{m_1, m_2} = \frac{1}{N} \sum_{i=0}^{N-1} p[i - k_1] p^*[i - k_2] e^{-j \frac{2\pi}{N} (m_1 - m_2) i} \quad (\text{C.11})$$

Setting $k = k_2 - k_1$ and $m = m_1 - m_2$, so that $k \in N_{\text{delay}}$ and $m \in N_{\text{Dopp}}$, (C.11) becomes

$$\begin{aligned} [\bar{\mathbf{P}}_{k_2, k_1}]_{m_1, m_2} &= \frac{1}{N} \sum_{q=-k_1}^{N-1-k_1} p[q] p^*[q - k] e^{-j \frac{2\pi}{N} m (q + k_1)}, \\ &= \frac{e^{-j \frac{2\pi}{N} m k_1}}{N} \sum_{q=0}^{N-1} p[q] p^*[q - k] e^{-j \frac{2\pi}{N} m q} \end{aligned} \quad (\text{C.12})$$

where in (C.12) we exploited the fact that $p[-q] = p[N - q]$ for $1 \leq q < N_{\text{delay}}$. Combining (C.10) and (C.12), we obtain (4.11). Similarly, we can show that (4.7) is equivalent to

$$\sum_{i=0}^{N-1} d[i] p^*[i - k] e^{-j \frac{2\pi}{N} m i} = 0 \quad \forall k \in \mathcal{N}_{\text{delay}}, \forall m \in \mathcal{N}_{\text{Dopp}}. \quad (\text{C.13})$$

Since *any* data vector \mathbf{d} should satisfy (C.13), and since the information symbols \mathbf{s} are arbitrary, the requirements (4.12) and (C.13) are equivalent.

C.3 Design of Optimal Pilot Sequence \mathbf{p}

Say $L = \frac{N}{N_{\text{delay}}} \in \mathbb{Z}$ and let $\mathbf{q} \in \mathbb{C}^L$ be arbitrary. Then if $\mathbf{p} = \mathbf{q} \otimes [1 \ 0 \ \dots \ 0]^\top$, we have $p[i]p^*[i-k] = 0$ for $k \in \mathcal{N}_{\text{delay}} \setminus 0$, and hence

$$\sum_{i=0}^{N-1} p[i]p^*[i-k]e^{-j\frac{2\pi}{N}mi} = 0 \quad \forall m, \quad \forall k \in \mathcal{N}_{\text{delay}} \setminus 0. \quad (\text{C.14})$$

Note that (C.14) partially satisfies (4.11); \mathbf{q} can be chosen to satisfy the remaining conditions:

$$\forall m \in \mathcal{N}_{\text{Dopp}}, \quad E_p \delta[m] = \sum_{i=0}^{N-1} p[i]p^*[i]e^{-j\frac{2\pi}{N}mi} \quad (\text{C.15})$$

$$= \sum_{i=0}^{L-1} |q[i]|^2 e^{-j\frac{2\pi}{L}mi} \quad (\text{C.16})$$

With $r[i] = |q[i]|^2$, $\mathbf{r} = [r[0], \dots, r[L-1]]^\top$, and the $(2N_{\text{Dopp}} - 1) \times L$ matrix $\bar{\mathbf{F}}_L$ with the elements $[\bar{\mathbf{F}}_L]_{n,m} = e^{j\frac{2\pi}{N}(n-N_{\text{Dopp}}+1)m}$, (C.16) becomes

$$\bar{\mathbf{F}}_L \mathbf{r} = E_p \mathbf{e}_{N_{\text{Dopp}}-1}, \quad (\text{C.17})$$

where \mathbf{e}_p is the p^{th} standard basis vector of $\mathbb{C}^{2N_{\text{Dopp}}-1}$. To find *real non-negative* \mathbf{r} that satisfies (C.17), we consider three cases.

- I) $L < N_{\text{Dopp}}$: This indicates an over-spread channel, previously excluded from consideration; no solution to (C.17) exists.
- II) $N_{\text{Dopp}} \leq L \leq 2N_{\text{Dopp}} - 1$: Here it is easy to show that $\mathbf{r} = \frac{E_p}{L}[1 \ \dots \ 1]^\top$ is the unique solution to (C.17).
- III) $2N_{\text{Dopp}} - 1 < L$: Here there exist many valid solutions to (C.17). Specifically, the real non-negative \mathbf{r} satisfying (C.17) are those that satisfy $\mathbf{F}_L^H \mathbf{r} = \mathbf{g}$ with

$\mathbf{g} = [g[0], \dots, g[L-1]]^\top$ such that $g[0] = \frac{E_p}{\sqrt{L}}$, $g[1] = \dots = g[N_{\text{Dopp}} - 1] = 0$, and $g[i] = g^*[L-i]$. Here \mathbf{F}_L denotes $L \times L$ unitary DFT matrix. Such \mathbf{r} can be generated via the following steps.

- 1) Choose any $\mathbf{g} \in \mathbb{C}^L$ with $g[0] \in \mathbb{R}^+$, $g[i] = g^*[L-i]$ and $g[1] = \dots = g[N_{\text{Dopp}} - 1] = 0$.
- 2) Compute the DFT $\mathbf{r}_1 = \mathbf{F}_L \mathbf{g}$.
- 3) Generate \mathbf{r}_2 element-wise as $r_2[i] = r_1[i] + c$ for any $c \in \mathbb{R}$ such that $r_2[i] \geq 0$.
- 4) Set $\mathbf{r} = \alpha \mathbf{r}_2$, where $\alpha = \frac{E_p}{\sum_{i=0}^{L-1} r_2[i]}$.

Given \mathbf{r} satisfying (C.17), we choose \mathbf{q} element-wise by setting $q(i) = \sqrt{r[i]} e^{j\theta[i]}$ for arbitrary $\theta[i] \in [0, 2\pi)$. Finally, we set $\mathbf{p} = \mathbf{q} \otimes [1 \ 0 \ \dots \ 0]^\top$.

C.4 Proof of Optimality of MMSE-PAT examples

C.4.1 TDKD

First, we show that the pilot sequence given in (4.18) satisfies the excitation criterion (4.11). Since the spacing between the non-zeros samples of the pilot pattern $\frac{N}{N_{\text{Dopp}}}$ is greater than N_{delay} (from the underspread assumption $N_{\text{Dopp}} N_{\text{delay}} < N$, we have $p[i] p^*[i-k] = 0$ for $k \in \mathcal{N}_{\text{delay}} \setminus 0$, and hence

$$\sum_{q=0}^{N-1} p[q] p^*[q-k] e^{-j \frac{2\pi}{N} m q} = 0 \quad \forall m \in \mathcal{N}_{\text{Dopp}}, \quad \forall k \in \mathcal{N}_{\text{delay}} \setminus 0. \quad (\text{C.18})$$

Now, for the case $k = 0$,

$$\begin{aligned}
\sum_{i=0}^{N-1} |p[i]|^2 e^{-j\frac{2\pi}{N}mi} &= \sum_{q=0}^{N_{\text{Dopp}}-1} \frac{E_p}{N_{\text{Dopp}}} e^{-j\frac{2\pi}{N}m(q\frac{N}{N_{\text{Dopp}}}+l)} \\
&= \frac{E_p}{N_{\text{Dopp}}} e^{-j\frac{2\pi}{N}ml} \underbrace{\sum_{q=0}^{N_{\text{Dopp}}-1} e^{-j\frac{2\pi}{N_{\text{Dopp}}}mq}}_{N_{\text{Dopp}}\delta[m]} \\
&= E_p\delta[m].
\end{aligned} \tag{C.19}$$

The equations (C.18) and (C.19) verify the optimal excitation requirement.

To verify the pilot data orthogonality requirement, we notice that the location of non-zero elements in each column of \mathbf{B} lies outside the set $\mathcal{G}_f^{[\ell]}$. Because of this, the nonzero entries in $\{p[i]\}$ never overlap with nonzero entries in $\{b_q[i]\}$, nor with nonzero entries of k -shifted $\{p[q-k]\}$ for shifts $k \in \mathcal{N}_{\text{delay}}$. In other words, $\{i : p[i-k] \neq 0 \forall k \in \mathcal{N}_{\text{delay}}\} \cap \{i : b_q[i] \neq 0\} = \emptyset$ and hence (4.7) follows.

C.4.2 FDKD

Note that $(\check{\mathbf{p}}, \check{\mathbf{B}})$ in FDKD coincide with the TDKD pattern but the roles of discrete delay spread N_{delay} and the discrete Doppler spread N_{Dopp} interchanged. So, $(\check{\mathbf{p}}, \check{\mathbf{B}})$ correspond to the MMSE-PAT for CE-BEM DSC with discrete delay spread N_{Dopp} and discrete Doppler spread N_{delay} . Now, from MMSE-PAT duality (Lemma 4), $(\mathbf{p} = \mathbf{F}_N^H \check{\mathbf{p}}, \mathbf{B} = \mathbf{F}_N^H \check{\mathbf{B}})$ is an MMSE-PAT for DSC with the discrete delay spread N_{delay} and the discrete Doppler spread N_{Dopp} .

C.4.3 Time domain Chirps

First, we shall establish that the chirp sequence (4.20) satisfies the optimal excitation criterion. For $i \in \{0, \dots, N_{\text{delay}} - 1\}$, we have

$$\begin{aligned}
 p[-i] &= p[N - i] = \sqrt{\frac{E_p}{N}} e^{j \frac{2\pi}{N} \frac{N_{\text{Dopp}}}{2} (N-i)^2} \\
 &= \sqrt{\frac{E_p}{N}} \underbrace{e^{j 2\pi \frac{N_{\text{Dopp}}}{2} N}}_{=1, \text{ for } N \text{ even}} e^{j \frac{2\pi}{N} \frac{N_{\text{Dopp}}}{2} i^2} \underbrace{e^{j 2\pi N_{\text{Dopp}} i}}_{=1} \\
 &= \sqrt{\frac{E_p}{N}} e^{j \frac{2\pi}{N} \frac{N_{\text{Dopp}}}{2} (-i)^2}.
 \end{aligned}$$

So, for $k \in \mathcal{N}_{\text{delay}}$ and $m \in \mathcal{N}_{\text{Dopp}}$, we have

$$\sum_{i=0}^{N-1} p[i] p^*[i - k] e^{-j \frac{2\pi}{N} m i} = \sum_{i=0}^{N-1} \sqrt{\frac{E_p}{N}} e^{j \frac{2\pi}{N} \frac{N_{\text{Dopp}}}{2} i^2} \sqrt{\frac{E_p}{N}} e^{j \frac{2\pi}{N} \frac{N_{\text{Dopp}}}{2} (i-k)^2} e^{j \frac{2\pi}{N} m i}, \quad (\text{C.20})$$

$$= \frac{E_p}{N} \sum_{i=0}^{N-1} e^{j \frac{2\pi}{N} \frac{N_{\text{Dopp}}}{2} (2ik - k^2)} e^{-j \frac{2\pi}{N} m i}, \quad (\text{C.21})$$

$$= \frac{E_p}{N} e^{-j \frac{2\pi}{N} \frac{N_{\text{Dopp}}}{2} k^2} \underbrace{\sum_{i=0}^{N-1} e^{j \frac{2\pi}{N} (k N_{\text{Dopp}} - m) i}}_{N \delta[k N_{\text{Dopp}} - m]}, \quad (\text{C.22})$$

$$= E_p \delta[k] \delta[m], \quad (\text{C.23})$$

where (C.23) follows from the fact that $k N_{\text{Dopp}} - m \neq 0 \pmod{N}$ as long as $k \in \mathcal{N}_{\text{delay}}$, $m \in \mathcal{N}_{\text{Dopp}}$ and $N_{\text{Dopp}} N_{\text{delay}} < N$.

Now, considering the inner product between the columns of chirp data modulation matrix,

$$\begin{aligned}
\sum_{i=0}^{N-1} b_{q_1}[i] b_{q_2}^*[i] &= \sum_{i=0}^{N-1} \left(\sqrt{\frac{1}{N}} e^{j \frac{2\pi}{N} (q_1 + N_{\text{Dopp}} N_{\text{delay}} i)} e^{j \frac{2\pi}{N} \frac{N_{\text{Dopp}}}{2} i^2} \right) \\
&\quad \times \left(\sqrt{\frac{1}{N}} e^{j \frac{2\pi}{N} (q_2 + N_{\text{Dopp}} N_{\text{delay}} i)} e^{j \frac{2\pi}{N} \frac{N_{\text{Dopp}}}{2} i^2} \right)^* \\
&= \frac{1}{N} \underbrace{\sum_{i=0}^{N-1} e^{j \frac{2\pi}{N} (q_1 - q_2) i}}_{N \delta[q_1 - q_2]} \\
&= \delta[q_1 - q_2],
\end{aligned}$$

we see that they are orthonormal and linearly independent. To verify the pilot data orthogonality, note that

$$\begin{aligned}
\sum_{i=0}^{N-1} b_q[i] p^*[i - k] e^{-j \frac{2\pi}{N} m i} &= \frac{\sqrt{E_p}}{N} \sum_{i=0}^{N-1} e^{j \frac{2\pi}{N} (q + N_{\text{Dopp}} N_{\text{delay}}) i} e^{j \frac{2\pi}{N} \frac{N_{\text{Dopp}}}{2} i^2} e^{-j \frac{2\pi}{N} \frac{N_{\text{Dopp}}}{2} (i - k)^2} \\
&\quad \times e^{-j \frac{2\pi}{N} m i}, \\
&= \frac{\sqrt{E_p}}{N} e^{-j \frac{2\pi}{N} \frac{N_{\text{Dopp}}}{2} k^2} \underbrace{\sum_{i=0}^{N-1} e^{j \frac{2\pi}{N} (k N_{\text{Dopp}} + q + N_{\text{Dopp}} N_{\text{delay}} - m) i}}_{N \delta[k N_{\text{Dopp}} + N_{\text{Dopp}} N_{\text{delay}} + q - m]},
\end{aligned}$$

which is zero for $k \in \mathcal{N}_{\text{delay}}$, $m \in \mathcal{N}_{\text{Dopp}}$ and the range of q given in (4.21).

C.4.4 Frequency domain Chirps

The frequency domain chirps is the dual of time domain chirps from Example 3. The proof follows along the same lines as that of FDKD in which the role of TDKD is replaced by the time domain MMSE chirp.

C.5 Proof of Theorem 3

Let the columns of $\mathbf{B}_p \in \mathbb{C}^{N \times N_{\text{delay}} N_{\text{Dopp}}}$ form an orthonormal basis for $\text{col}(\mathbf{P}\mathbf{U})$ and the columns of $\mathbf{B}_d \in \mathbb{C}^{N \times (N - N_{\text{delay}} N_{\text{Dopp}})}$ form an orthonormal basis for the left null

space of $\mathbf{P}U$. Notice that, with $\bar{\mathbf{B}} = [\mathbf{B}_p \mathbf{B}_d]$, we have $\bar{\mathbf{B}}^H \bar{\mathbf{B}} = \bar{\mathbf{B}} \bar{\mathbf{B}}^H = \mathbf{I}_N$. The pilot-data orthogonality of MMSE-PAT [recall (4.7)] implies that $\mathbf{B}_p^H \hat{\mathbf{H}} \mathbf{B} = \mathbf{0}$ and $\mathbf{B}_d^H \hat{\mathbf{H}} \mathbf{p} = \mathbf{0}$. Projecting the observed vector \mathbf{y} onto the pilot and data subspaces, we obtain $\mathbf{y}_p = \mathbf{B}_p^H \mathbf{y}$ and $\mathbf{y}_d = \mathbf{B}_d^H \mathbf{y}$, respectively, where

$$\mathbf{y}_p = \sqrt{\rho} \mathbf{B}_p^H \hat{\mathbf{H}} \mathbf{p} + \underbrace{\mathbf{B}_p^H \mathbf{v}}_{\mathbf{v}_p}, \quad (\text{C.24})$$

$$\mathbf{y}_d = \sqrt{\rho} \mathbf{B}_d^H \hat{\mathbf{H}} \mathbf{B} \mathbf{s} + \underbrace{\mathbf{B}_d^H \mathbf{v}}_{\mathbf{v}_d}. \quad (\text{C.25})$$

Clearly, \mathbf{v}_p and \mathbf{v}_d are CWGN with unit variance. Since the projection (C.24) does not compromise pilot energy, the LMMSE estimate of \mathbf{h} given \mathbf{y} equals the LMMSE estimate given \mathbf{y}_p .

Splitting the channel matrix $\hat{\mathbf{H}}$ into the estimate and error component as $\hat{\mathbf{H}} = \tilde{\mathbf{H}} + \mathbf{H}$, we have

$$\mathbf{y}_d = \sqrt{\rho} \mathbf{B}_d^H \tilde{\mathbf{H}} \mathbf{B} \mathbf{s} + \underbrace{\sqrt{\rho} \mathbf{B}_d^H \mathbf{H} \mathbf{B} \mathbf{s}}_{\mathbf{n}} + \mathbf{v}_d. \quad (\text{C.26})$$

Let $\mathbf{R}_n = E\{\mathbf{n} \mathbf{n}^H\}$ denote the covariance of the effective noise \mathbf{n} . With the application of weighting factor $\mathbf{Q} = \mathbf{R}_n^{-1/2} \mathbf{B}_d^H$ in the minimum distance decoding (Section 2.3.2), which has the interpretation of pilot-data separation followed by effective-noise whitening-filter, the achievable rates are given by [30]

$$\mathcal{R}_{\text{mmse-blk}} \geq E\{\log \det[\mathbf{I} + \rho \sigma_s^2 \mathbf{R}_n^{-1} \mathbf{B}_d^H \hat{\mathbf{H}} \mathbf{B} (\mathbf{B}_d^H \hat{\mathbf{H}} \mathbf{B})^H]\}. \quad (\text{C.27})$$

The achievable rate result in [30] is derived under the assumption that MMSE channel estimator is employed. But, with the pilot-data orthogonality, our LMMSE estimator (2.30) coincides with the MMSE channel estimator, since the pilot observations and the channel coefficients are jointly Gaussian. Note that the above lemma is

applicable to the CP-MMSE-PAT schemes since they satisfy the pilot-data orthogonality requirement Theorem 2. The above rate expression resembles that of coherent case [33] with \mathbf{n} acting as “effective” Gaussian noise. Furthermore, we have

$$\mathbf{R}_n = \mathbf{I} + \rho\sigma_s^2 E\{\mathbf{B}_d^H \tilde{\mathbf{H}} \mathbf{B} (\mathbf{B}_d^H \tilde{\mathbf{H}} \mathbf{B})^H\} \quad (\text{C.28})$$

$$\leq \mathbf{I} + \rho\sigma_s^2 \mathbf{B}_d^H E\{\tilde{\mathbf{H}} \tilde{\mathbf{H}}^H\} \mathbf{B}_d, \quad (\text{C.29})$$

in the positive definite sense, since the columns of \mathbf{B} are orthonormal.

Since $E\{\tilde{\mathbf{H}} \tilde{\mathbf{H}}^H\} = \frac{\text{tr}\{\mathbf{R}_{\tilde{h}}\}}{N} \mathbf{I}$ and $\mathbf{B}_d^H \mathbf{B}_d = \mathbf{I}$, we have

$$\mathbf{R}_n \leq \left(1 + \rho\sigma_s^2 \frac{\text{tr}\{\mathbf{R}_{\tilde{h}}\}}{N}\right) \mathbf{I}. \quad (\text{C.30})$$

Substituting (C.30) into (C.27), we have

$$\mathcal{R}_{\text{mmse-blk}} \geq E\left\{\log \det\left[\mathbf{I} + \frac{\rho N \sigma_s^2}{\rho\sigma_s^2 \text{tr}\{\mathbf{R}_{\tilde{h}}\} + N} \mathbf{B}_d^H \hat{\mathbf{H}} \mathbf{B} (\mathbf{B}_d^H \hat{\mathbf{H}} \mathbf{B})^H\right]\right\}. \quad (\text{C.31})$$

It follows that $\mathbf{B}_p^H \hat{\mathbf{H}} \mathbf{B} = \mathbf{0}$ due to the similarity in the structure of $\hat{\mathbf{H}}$ and $\tilde{\mathbf{H}}$. Using the determinant identity $\det(\mathbf{I} + \mathbf{G}_1 \mathbf{G}_2) = \det(\mathbf{I} + \mathbf{G}_2 \mathbf{G}_1)$ and the fact that $\mathbf{B}_d \mathbf{B}_d^H = \mathbf{I}_N - \mathbf{B}_p \mathbf{B}_p^H$, the lower bound on the mutual information (C.31) can be rewritten as (4.27) after the normalization of $\hat{\mathbf{H}}$.

For the upper bound on the mutual information, we consider the “best case” scenario of perfect channel estimates, i.e., $\hat{\mathbf{H}} = \tilde{\mathbf{H}}$ and $\tilde{\mathbf{H}} = \mathbf{0}$. For this scenario, with i.i.d. Gaussian input distribution, it is straightforward [33] to obtain the bound (4.28).

C.6 Proof of Theorem 4

We bound the asymptotic achievable rates from both below and above and get the complete characterization. To start, $\mathcal{R}_{\text{mmse-blk-lb}} \leq \mathcal{R}_{\text{mmse-blk}}$. First, let us characterize

the asymptotic behavior of $\mathcal{R}_{\text{mmse-blk-lb}}$. With any *fixed* power allocation fraction $\alpha \in (0, 1)$, $E_s = \alpha E_{\text{tot}}$ and $E_p = (1 - \alpha)E_{\text{tot}}$, as $\rho \rightarrow \infty$, estimation error variance goes to zero, i.e., $\lim_{\rho \rightarrow \infty} \text{tr}\{\mathbf{R}_{\hat{h}}\} = 0$ and hence it follows that $\lim_{\rho \rightarrow \infty} \text{tr}\{\mathbf{R}_{\hat{h}}\} = N$. Also, as $\rho \rightarrow \infty$, the normalized channel estimates are equal to the actual channel almost surely (a.s.),

$$\lim_{\rho \rightarrow \infty} \bar{\mathbf{H}} = \hat{\mathbf{H}} \text{ a.s.} \quad (\text{C.32})$$

Using Fatou's lemma [69,70], we find the lower bound on the asymptotics of $\mathcal{R}_{\text{mmse-blk-lb}}$, by interchanging the limit and the expectation. We have

$$\lim_{\rho \rightarrow \infty} \mathcal{R}_{\text{mmse-blk-lb}} \geq E\{\lim_{\rho \rightarrow \infty} \log \det[\mathbf{I} + \rho_{\text{lb}} \mathbf{B}^H \bar{\mathbf{H}}^H \bar{\mathbf{H}} \mathbf{B}]\}, \quad (\text{C.33})$$

$$= E\{\log \det \lim_{\rho \rightarrow \infty} [\mathbf{I} + \rho_{\text{lb}} \mathbf{B}^H \bar{\mathbf{H}}^H \bar{\mathbf{H}} \mathbf{B}]\}, \quad (\text{C.34})$$

which follows from the continuity of $\log \det(\cdot)$. It easily follows that $\exists k$ such that $\rho_{\text{lb}} \geq k\rho$, $\forall \rho \geq 1$, we have

$$\lim_{\rho \rightarrow \infty} \mathcal{R}_{\text{mmse-blk-lb}} \geq E\{\log \det \lim_{\rho \rightarrow \infty} [\mathbf{I} + k\rho \mathbf{B}^H \hat{\mathbf{H}}^H \hat{\mathbf{H}} \mathbf{B}]\}, \quad (\text{C.35})$$

$$= E\{\log \prod_{i=0}^{N_s-1} (1 + k\rho\mu_i)\}, \quad (\text{C.36})$$

where μ_i are the eigen values of $\mathbf{B}^H \hat{\mathbf{H}}^H \hat{\mathbf{H}} \mathbf{B}$. Since $\hat{\mathbf{H}} \mathbf{B}$ is full rank (N_s) with probability 1, we have the following lower bound on $\mathcal{R}_{\text{mmse-blk}}$,

$$\mathcal{R}_{\text{mmse-blk}}(\rho) \geq \mathcal{R}_{\text{mmse-blk-lb}}(\rho) = N_s \log(\rho) + O(1), \quad (\text{C.37})$$

as $\rho \rightarrow \infty$.

Now, to bound the asymptotic achievable rates from above, we consider the ‘‘coherent’’ case of zero channel estimation error. Since the pilot component $\hat{\mathbf{H}} \mathbf{p}$ can be

subtracted from \mathbf{y} (4.24) without affecting the mutual information, we start with

$$\mathbf{y}_c = \sqrt{\rho} \dot{\mathbf{H}} \mathbf{B} \mathbf{s} + \mathbf{v}. \quad (\text{C.38})$$

Using \mathcal{C}_{coh} to denote the capacity of (C.38) with power constraint $E\{\|\mathbf{s}\|^2\} \leq N$, it is evident that $\mathcal{R}_{\text{mmse-blk}} \leq \mathcal{C}_{\text{coh}}$. In fact, for (C.38), the capacity maximizing input distribution is zero-mean Gaussian [33], so that

$$\mathcal{R}_{\text{mmse-blk}} \leq \mathcal{C}_{\text{coh}} = \sup_{\text{tr}\{\mathbf{R}_s\} \leq N} E\{\log \det[\mathbf{I}_{N_s} + \rho(\dot{\mathbf{H}} \mathbf{B})^H \mathbf{R}_s \dot{\mathbf{H}} \mathbf{B}]\}, \quad (\text{C.39})$$

where \mathbf{R}_s denotes the covariance of \mathbf{s} . Note that, for any \mathbf{R}_s satisfying the power constraint in (C.39), we have

$$\mathbf{R}_s \leq N \mathbf{I}_{N_s}, \quad (\text{C.40})$$

in the positive semi-definite sense. Using (C.40) in (C.39), we find

$$\mathcal{R}_{\text{mmse-blk}} \leq E\{\log \det[\mathbf{I}_{N_s} + N\rho(\dot{\mathbf{H}} \mathbf{B})^H \dot{\mathbf{H}} \mathbf{B}]\} \quad (\text{C.41})$$

$$= N_s \log \rho + O(1), \quad (\text{C.42})$$

as $\rho \rightarrow \infty$. Theorem 4 follows from (C.37) and (C.42).

APPENDIX D

PROOFS OF SPECTRALLY EFFICIENT PAT RESULTS

D.1 Proof of Theorem 5

We use modulo- N indexing throughout this proof. First define

$$\mathbf{e}_{(k,m)} = \frac{1}{\sqrt{E_p}} [p[k]e^{j\frac{2\pi}{N}m \cdot 0}, p[k+1]e^{j\frac{2\pi}{N}m \cdot 1}, \dots, p[k+N-1]e^{j\frac{2\pi}{N}m(N-1)}]^\top,$$

which are normalized to have unit norm, for convenience. We also define the sets, $\mathcal{N}_{\text{delay}} = \{-N_{\text{delay}}+1, \dots, N_{\text{delay}}-1\}$ and $\mathcal{N}_{\text{Dopp}} = \{-N_{\text{Dopp}}+1, \dots, N_{\text{Dopp}}-1\}$. Let \mathbf{W} be a matrix whose columns are constructed from the set $\{\mathbf{e}_{(k,m)}, k \in \mathcal{N}_{\text{delay}}, m \in \mathcal{N}_{\text{Dopp}}\}$. Now, the orthogonality requirement (4.7) can be written as $\mathbf{W}^H \mathbf{B} = \mathbf{0}$ and hence the number of information symbols in each block $N_s = \text{rank}(\mathbf{B})$ is at most equal to the dimension of the null space of \mathbf{W}^H .

We proof the theorem by contradiction. We assume there are MMSE-PAT schemes for which $\text{rank}(\mathbf{B}) = N - N_{\text{Dopp}}N_{\text{delay}}$ and find the necessary requirements on their pilot vectors. Then we establish that the pilot vectors satisfying these requirements does not yield $\text{rank}(\mathbf{B}) = N - N_{\text{Dopp}}N_{\text{delay}}$.

Let (\mathbf{p}, \mathbf{B}) correspond to a MMSE-PAT with $\text{rank}(\mathbf{B}) = N - N_{\text{Dopp}}N_{\text{delay}}$. Based on the bounds of N_s (4.15), we have N_s is equal to the dimension of the null space of

Now we have,

$$\langle \mathbf{e}_{(k_1, m_1)}, \mathbf{e}_{(k_2, m_2)} \rangle = \frac{1}{E_p} \sum_{i=0}^{N-1} p[i+k_1] p^*[i+k_2] e^{j\frac{2\pi}{N} m_1 i} e^{-j\frac{2\pi}{N} m_2 i} \quad (\text{D.4})$$

$$= \frac{1}{E_p} \sum_{q=k_1}^{N-1+k_1} p[q] p^*[q+(k_2-k_1)] e^{-j\frac{2\pi}{N} (m_2-m_1)(q-k_1)} \quad (\text{D.5})$$

$$= \frac{e^{j\frac{2\pi}{N} (m_2-m_1)k_1}}{E_p} \left(\sum_{q=k_1}^{N-1} p[q] p^*[q+(k_2-k_1)] e^{-j\frac{2\pi}{N} (m_2-m_1)q} \right. \\ \left. + \sum_{q=N}^{N-1+k_1} p[q] p^*[q+(k_2-k_1)] e^{-j\frac{2\pi}{N} (m_2-m_1)q} \right) \quad (\text{D.6})$$

$$= \frac{e^{j\frac{2\pi}{N} (m_2-m_1)k_1}}{E_p} \sum_{q=0}^{N-1} p[q] p^*[q+(k_2-k_1)] e^{-j\frac{2\pi}{N} (m_2-m_1)q} \quad (\text{D.7})$$

$$= e^{j\frac{2\pi}{N} (m_2-m_1)k_1} r_{(k_2-k_1, m_2-m_1)}. \quad (\text{D.8})$$

From (D.3) and (D.8), all the elements in Fig. D.1 *within any rectangle of height N_{delay} and width N_{Dopp} are orthonormal.* We also have,

$$r_{(k, m)}^* = \langle \mathbf{e}_{(0,0)}, \mathbf{e}_{(k, m)} \rangle^* = \langle \mathbf{e}_{(k, m)}, \mathbf{e}_{(0,0)} \rangle = e^{-j\frac{2\pi}{N} mk} r_{(-k, -m)}. \quad (\text{D.9})$$

We use the following intermediate result.

Lemma 9. *If \mathbf{B} is of rank $N - N_{\text{Dopp}}N_{\text{delay}}$, then $|r_{(0, N_{\text{Dopp}})}| = 1$ or $|r_{(N_{\text{delay}}, 0)}| = 1$.*

Proof. Since $\text{rank}(\mathbf{B}) = N - N_{\text{Dopp}}N_{\text{delay}}$ is the nullity of \mathbf{W}^H , it follows that $\text{rank}(\mathbf{W}) = N_{\text{Dopp}}N_{\text{delay}}$. Let \mathbf{W}_o be a matrix whose columns are from the set $\{\mathbf{e}_{(k, m)} : k \in \{0, -1, \dots, -N_{\text{delay}} + 1\}, m \in \{-D, \dots, D\}\}$, $\mathbf{W}_e^1 = [\mathbf{e}_{(-1, -D-1)}, \dots, \mathbf{e}_{(-N_{\text{delay}}+2, -D-1)}]$ and $\mathbf{W}_e^2 = [\mathbf{e}_{(1, -D)}, \dots, \mathbf{e}_{(1, D-1)}]$. (See Fig. D.1.) From (4.8), (D.3) and (D.8), the $N_{\text{Dopp}}N_{\text{delay}}$ columns of \mathbf{W}_o are orthonormal and, since \mathbf{W} is of rank $N_{\text{Dopp}}N_{\text{delay}}$, we

have the following basis expansion:

$$\mathbf{e}_{(k,m)} = \sum_{i=0}^{N_{\text{delay}}-1} \sum_{j=-D}^D \langle \mathbf{e}_{(k,m)}, \mathbf{e}_{(-i,j)} \rangle \mathbf{e}_{(-i,j)}, \forall k \in \mathcal{N}_{\text{delay}}, \forall m \in \mathcal{N}_{\text{Dopp}}, \quad (\text{D.10})$$

$$= \sum_{i=0}^{N_{\text{delay}}-1} \sum_{j=-D}^D e^{j \frac{2\pi}{N} (j-m)k} r_{(-i-k,j-m)} \mathbf{e}_{(-i,j)}, \forall k \in \mathcal{N}_{\text{delay}}, \forall m \in \mathcal{N}_{\text{Dopp}}. \quad (\text{D.11})$$

Since any two elements inside the rectangle of height N_{delay} and width N_{Dopp} are orthogonal ((D.3),(D.8)), for the columns of \mathbf{W}_e^1 , we have

$$[\mathbf{e}_{(0,-D-1)}, \mathbf{e}_{(-1,-D-1)}, \dots, \mathbf{e}_{(-N_{\text{delay}}+2,-D-1)}] = [\mathbf{e}_{(0,D)}, \mathbf{e}_{(-1,D)}, \dots, \mathbf{e}_{(-N_{\text{delay}}+1,D)}] \mathbf{M}_1$$

where $\mathbf{M}_1 \in \mathbb{C}^{N_{\text{delay}} \times N_{\text{delay}}-1}$ is given by

$$\begin{bmatrix} r(0, N_{\text{Dopp}}) & e^{-j \frac{2\pi}{N} N_{\text{Dopp}}} r(1, N_{\text{Dopp}}) & \dots & e^{-j \frac{2\pi}{N} N_{\text{Dopp}} (N_{\text{delay}}-2)} r(N_{\text{delay}}-2, N_{\text{Dopp}}) \\ r(-1, N_{\text{Dopp}}) & e^{-j \frac{2\pi}{N} N_{\text{Dopp}}} r(0, N_{\text{Dopp}}) & \dots & e^{-j \frac{2\pi}{N} N_{\text{Dopp}} (N_{\text{delay}}-2)} r(N_{\text{delay}}-3, N_{\text{Dopp}}) \\ \vdots & \vdots & \dots & \vdots \\ r(-N_{\text{delay}}+1, N_{\text{Dopp}}) & e^{-j \frac{2\pi}{N} N_{\text{Dopp}}} r(-N_{\text{delay}}+2, N_{\text{Dopp}}) & \dots & e^{-j \frac{2\pi}{N} N_{\text{Dopp}} (N_{\text{delay}}-2)} r(-1, N_{\text{Dopp}}) \end{bmatrix}.$$

Similarly, we have the following expansion for the columns of \mathbf{W}_e^2 ,

$$[\mathbf{e}_{(1,-D)}, \mathbf{e}_{(1,-D+1)}, \dots, \mathbf{e}_{(1,D-1)}] = [\mathbf{e}_{(-N_{\text{delay}}+1,-D)}, \mathbf{e}_{(-N_{\text{delay}}+1,-D+1)}, \dots, \mathbf{e}_{(-N_{\text{delay}}+1,D)}] \mathbf{M}_2$$

where $\mathbf{M}_2 \in \mathbb{C}^{N_{\text{Dopp}} \times N_{\text{Dopp}}-1}$ is given by

$$\begin{bmatrix} r(-N_{\text{delay}}, 0) & e^{-j \frac{2\pi}{N}} r(-N_{\text{delay}}, -1) & \dots & e^{-j \frac{2\pi}{N} (N_{\text{Dopp}}-2)} r(-N_{\text{delay}}, -N_{\text{Dopp}}+2) \\ e^{j \frac{2\pi}{N}} r(-N_{\text{delay}}, 1) & r(-N_{\text{delay}}, 0) & \dots & e^{-j \frac{2\pi}{N} (N_{\text{Dopp}}-3)} r(-N_{\text{delay}}, -N_{\text{Dopp}}+3) \\ \vdots & \vdots & \dots & \vdots \\ e^{j \frac{2\pi}{N} (N_{\text{Dopp}}-1)} r(-N_{\text{delay}}, N_{\text{Dopp}}-1) & e^{j \frac{2\pi}{N} (N_{\text{Dopp}}-2)} r(-N_{\text{delay}}, N_{\text{Dopp}}-2) & \dots & e^{j \frac{2\pi}{N}} r(-N_{\text{delay}}, 1) \end{bmatrix}.$$

Since each column of \mathbf{W}_e^1 is orthogonal to each column of \mathbf{W}_e^2 , from their basis expansions, we see that they have only one common basis vector $\mathbf{e}_{(-N_{\text{delay}}+1,D)}$. So, to meet the orthogonality requirement, we have

$$r(-1, N_{\text{Dopp}}) = r(-2, N_{\text{Dopp}}) = \dots = r(-N_{\text{delay}}+1, N_{\text{Dopp}}) = 0 \quad (\text{D.12})$$

or

$$r_{(-N_{\text{delay}},1)} = r_{(-N_{\text{delay}},2)} = \cdots = r_{(-N_{\text{delay}},N_{\text{Dopp}}-1)} = 0. \quad (\text{D.13})$$

Using (D.12) in the basis expansion of $\mathbf{e}_{(0,-D-1)}$, we have

$$\mathbf{e}_{(0,-D-1)} = r_{(0,N_{\text{Dopp}})}\mathbf{e}_{(0,D)}. \quad (\text{D.14})$$

Since both $\mathbf{e}_{(0,-D-1)}$ and $\mathbf{e}_{(0,D)}$ have unit norm, we have

$$|r_{(0,N_{\text{Dopp}})}| = 1 \Rightarrow r_{(0,N_{\text{Dopp}})} = e^{j\theta} \text{ for some } \theta \in \mathbb{R}. \quad (\text{D.15})$$

Similarly, when the condition (D.13) is met, we have

$$\mathbf{e}_{(1,-D)} = r_{(-N_{\text{delay}},0)}\mathbf{e}_{(-N_{\text{delay}}+1,-D)} \quad (\text{D.16})$$

and $|r_{(-N_{\text{delay}},0)}| = 1$. So, from (D.9) we have

$$r_{(N_{\text{delay}},0)} = e^{j\bar{\theta}} \text{ for some } \bar{\theta} \in \mathbb{R}. \quad (\text{D.17})$$

□

Now we study the pilot vectors \mathbf{p} which satisfy (4.8) with the additional constraint that $|r_{(N_{\text{delay}},0)}| = 1$ or $|r_{(0,N_{\text{Dopp}})}| = 1$. Considering these two cases separately, we establish that there is no such \mathbf{p} for which $\text{rank}(\mathbf{B}) = N - N_{\text{Dopp}}N_{\text{delay}}$.

Case I: $r_{(0,N_{\text{Dopp}})} = e^{j\theta}$

From (D.14), we have

$$p[i](e^{-j\frac{2\pi}{N}N_{\text{Dopp}}i} - e^{j\theta}) = 0. \quad (\text{D.18})$$

Now, if $\theta \neq \frac{2\pi}{N}N_{\text{Dopp}}q$ for some $q \in \mathbb{Z}$, then $p[i] = 0 \forall i$, which clearly does not satisfy the MMSE-PAT requirement (4.8), and hence ruled out from consideration. Now, if

$\theta = \frac{2\pi}{N}N_{\text{Dopp}}q$ for some $q \in \mathbb{Z}$, from (D.18), $p[i]$ may be non-zero only if $i = q + \frac{kN}{N_{\text{Dopp}}}$ for $k \in \mathbb{Z}$ such that $\frac{kN}{N_{\text{Dopp}}} \in \mathbb{Z}$. Now, for $k \in \mathbb{Z}$, defining

$$a_q[k] = \begin{cases} |p[q + \frac{kN}{N_{\text{Dopp}}}]|^2 & \text{if } \frac{kN}{N_{\text{Dopp}}} \in \mathbb{Z} \\ 0 & \text{else} \end{cases} \quad (\text{D.19})$$

then from the requirement (D.3), it follows that $\sum_{i=0}^{N_{\text{Dopp}}-1} a_q[i] e^{j\frac{2\pi}{N_{\text{Dopp}}}mi} = E_p \delta[m]$, $\forall m \in \mathcal{N}_{\text{Dopp}}$, which can be met if and only if

$$a_q[i] = \frac{E_p}{N_{\text{Dopp}}}, \forall i \in \{0, \dots, N_{\text{Dopp}} - 1\}. \quad (\text{D.20})$$

From the definition (D.19), it follows that, the above requirement can be met if and only if $\frac{N}{N_{\text{Dopp}}} \in \mathbb{Z}$. If $\frac{N}{N_{\text{Dopp}}} \notin \mathbb{Z}$, there is no training sequence which satisfies both (4.8) and (D.15). Now, if $\frac{N}{N_{\text{Dopp}}} \in \mathbb{Z}$, from (D.19) and (D.20), the sequence $p[i]$ is of the form given in Example 1. For Example 1, as noted earlier, $\text{rank}(\mathbf{B}) = N_s = N - (2N_{\text{delay}} - 1)N_{\text{Dopp}} < N - N_{\text{Dopp}}N_{\text{delay}}$. This contradicts the initial assumption that \mathbf{B} is of rank $N_{\text{Dopp}}N_{\text{delay}}$.

Case II: $r_{(N_{\text{delay}}, 0)} = e^{j\bar{\theta}}$

From (D.9), (D.16) and (D.17), it follows that

$$p[i] = e^{j\bar{\theta}} p[i + Nt]. \quad (\text{D.21})$$

Because of the circular symmetry $p[i + N] = p[i]$, using (D.21), we can find the largest integer $L \in \{1, \dots, N_{\text{delay}}\}$ so that $\frac{N}{L} \in \mathbb{Z}$ and $p[i] = e^{j\theta} p[i + L]$ for some $\theta \in \mathbb{R}$. Note that, if $\frac{N}{N_{\text{delay}}} \in \mathbb{Z}$ then $L = N_{\text{delay}}$ else $L < N_{\text{delay}}$. Again from the circular symmetry, $\theta = \frac{2\pi}{N}Lq$ for some $q \in \mathbb{Z}$. Let $\check{\mathbf{p}}$ denote the N -point unitary discrete Fourier transform

(DFT) of \mathbf{p} . For the sequence \mathbf{p} with the given ‘‘periodic’’ structure, we have

$$\check{p}[k] = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} p[i] e^{-j\frac{2\pi}{N}ik} \quad (\text{D.22})$$

$$= \frac{1}{\sqrt{N}} \sum_{i=0}^{L-1} p[i] e^{-j\frac{2\pi}{N}ik} \sum_{m=0}^{\frac{N}{L}-1} e^{-j\frac{2\pi}{N}L(k-q)m} \quad (\text{D.23})$$

and hence $\check{p}[k] = 0 \forall k \notin \{q, q + \frac{N}{L}, \dots, q + \frac{N(L-1)}{L}\}$. Now, the optimal excitation requirement (4.8) can be written in terms of $\check{\mathbf{p}}$ as (Recall Lemma 3)

$$\sum_{i=0}^{N-1} \check{p}[i] \check{p}^*[i-k] e^{-j\frac{2\pi}{N}mi} = E_p \delta[k] \delta[m] \quad \forall k \in \mathcal{N}_{\text{Dopp}}, \forall m \in \mathcal{N}_{\text{delay}}. \quad (\text{D.24})$$

Defining $|\check{p}[q + \frac{iN}{L}]|^2 = \check{a}_q[i], i \in \{0, \dots, L-1\}$ and using the MMSE-PAT requirements in the frequency domain, we require

$$g[m] = \sum_{i=0}^{L-1} \check{a}_q[i] e^{-j\frac{2\pi}{N}(i\frac{N}{L}+q)m} = E_p \delta[m], \quad \forall m \in \mathcal{N}_{\text{delay}}. \quad (\text{D.25})$$

If $L < N_{\text{delay}}$, then (D.25) can not be satisfied since $g[L] = g[0] e^{j\frac{2\pi}{N}qL}$. So, if $\frac{N}{N_{\text{delay}}} \notin \mathbb{Z}$, there is no MMSE-PAT with $\dim(\mathbf{W}) = N_{\text{Dopp}} N_{\text{delay}}$. Now, if $\frac{N}{N_{\text{delay}}} \in \mathbb{Z}$, then $L = N_{\text{delay}}$ and the only sequence $\{\check{a}_q[i]\}$ satisfying the requirement (D.25) is $\check{a}_q[i] = c$ for some constant $c, \forall i$. This corresponds to the equi-spaced, equi-powered frequency domain pilot sequence of FDKD in Example 2. Again, for this pilot sequence, from Example 2, we have $\text{rank}(\mathbf{B}) < N - N_{\text{Dopp}} N_{\text{delay}}$. Again, we reach a contradiction on the initial assumption that \mathbf{B} is of rank $N_{\text{Dopp}} N_{\text{delay}}$.

D.2 Proof of Theorem 6

First, we establish that, for the PAT schemes satisfying the hypothesis, the total estimation error satisfies $\sigma_e^2 \leq \frac{\kappa}{\rho}, \forall \rho$. Since the PAT is linearly separable, the projection \mathbf{y}_p in (5.8) captures all the pilot energy and denoting $\mathbf{G} = \mathbf{B}_p^H \mathbf{P} \mathbf{U}$, we

have

$$\mathbf{y}_p = \sqrt{\rho} \mathbf{G} \boldsymbol{\lambda} + \mathbf{v}_p. \quad (\text{D.26})$$

Since $\mathbf{P}\mathbf{U}$ is full rank, it easily follows that the matrix \mathbf{G} is full rank. Note that $\sigma_e^2 = \text{E}\{\|\boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}}\|^2\}$ where $\hat{\boldsymbol{\lambda}}$ denotes the LMMSE estimate of $\boldsymbol{\lambda}$. Using the zero forcing estimate from (D.26) to upper bound σ_e^2 , we have

$$\sigma_e^2 \leq \frac{1}{\rho} \text{tr}\{(\mathbf{G}^H \mathbf{G})^{-1}\}. \quad (\text{D.27})$$

Since \mathbf{G} is full rank, we have $\text{tr}\{(\mathbf{G}^H \mathbf{G})^{-1}\} \leq \kappa$ for some $\kappa \in \mathbb{R}$. Now, for a PAT scheme which satisfies pilot-data orthogonality (4.7), with the choice of weighting matrix \mathbf{Q} given in (5.6), the achievable rates are given by (5.9). First, we obtain a lower bound on the prelog factor of the achievable rates. Now, \mathbf{R}_n , the covariance of $\sqrt{\rho} \mathbf{B}_d^H \tilde{\mathbf{H}} \mathbf{B} \mathbf{s} + \mathbf{v}_d$ is bounded as $\mathbf{R}_n \leq (1 + \rho \sigma_e^2 E_s \|\mathbf{M}\|_F^2) \mathbf{I}$. Recall that \mathbf{M} is the ZP or CP mapping matrix. Since $\sigma_e^2 \leq \frac{\kappa}{\rho}$, we have $\mathbf{R}_n \leq C \mathbf{I}$ for some constant C , $\forall \rho > 1$. So the achievable rates are bounded as

$$\mathcal{R}(\rho) \geq \frac{1}{N} \text{E}\{\log \det[\mathbf{I} + \frac{\rho}{C} \hat{\mathbf{H}}_d \mathbf{R}_s \hat{\mathbf{H}}_d^H]\}. \quad (\text{D.28})$$

Since $\sigma_e^2 \rightarrow 0$ as $\rho \rightarrow \infty$, the channel estimates converge almost everywhere to the true channel, i.e., $\lim_{\rho \rightarrow \infty} \hat{\mathbf{H}} = \mathbf{H}$. Also, $\hat{\mathbf{H}}_d \mathbf{R}_s \hat{\mathbf{H}}_d^H$ has rank equal to $\text{rank}(\mathbf{B})$ with continuously distributed eigen values, we have $\lim_{\rho \rightarrow \infty} \frac{\mathcal{R}(\rho)}{\log \rho} \geq \frac{\text{rank}(\mathbf{B})}{N}$. From the hypothesis, $\text{rank}(\mathbf{B}) = N - N_{\text{Dopp}} N_{\text{delay}}$. The upper bound on pre-log factor follows by applying Jensen's inequality and taking the expectation inside $\log \det(\cdot)$ in the rate expression (5.9).

APPENDIX E

PROOFS OF MIMO-MMSE-PAT RESULTS

E.1 Proof of Theorem 7

Using the techniques similar to the derivation of MMSE-PAT requirements for SISO DSC (Chapter 4), we obtain that the pilot-data orthogonality $(\bar{\mathbf{P}}\bar{\mathbf{U}})^H \bar{\mathbf{D}}\bar{\mathbf{U}} = \mathbf{0} \forall \bar{\mathbf{D}}$ which is equivalent to (6.14), is a necessary requirement for MIMO-MMSE-PAT. When (6.14) is satisfied, we have

$$\sigma_e^2 = \text{tr}\{(\mathbf{R}_{\bar{\lambda}}^{-1} + \rho(\bar{\mathbf{P}}\bar{\mathbf{U}})^H \bar{\mathbf{P}}\bar{\mathbf{U}})^{-1}\} \quad (\text{E.1})$$

$$\geq \sum_{m=0}^{N_{\text{Dopp}} N_{\text{delay}} RT - 1} \left(\frac{1}{[\mathbf{R}_{\bar{\lambda}}]_{m,m}} + \rho [(\bar{\mathbf{P}}\bar{\mathbf{U}})^H \bar{\mathbf{P}}\bar{\mathbf{U}}]_{m,m} \right)^{-1} \quad (\text{E.2})$$

$$= R \sum_{t=0}^{T-1} \sum_{m=0}^{N_{\text{Dopp}} N_{\text{delay}} - 1} \left(\frac{1}{[\mathbf{R}_{\lambda}]_{m,m}} + \frac{\rho \|\mathbf{p}^{[t]}\|^2}{N} \right)^{-1} \quad (\text{E.3})$$

and the equality in (E.2) is achieved if and only if the columns of $\bar{\mathbf{P}}\bar{\mathbf{U}}$ are orthogonal. Since $\sum_{t=0}^{T-1} \|\mathbf{p}^{[t]}\|^2 = E_p$, from the method of Lagrange multipliers

$$\sigma_e^2 \geq RT \sum_{m=0}^{N_{\text{Dopp}} N_{\text{delay}} - 1} \left(\frac{1}{[\mathbf{R}_{\lambda}]_{m,m}} + \frac{\rho E_p}{NT} \right)^{-1} \quad (\text{E.4})$$

and the equality is achieved if and only if (6.15) is met.

E.2 Proof of Theorem 8

Let the columns of $\mathbf{B}_p \in \mathbb{C}^{N \times TN_{\text{Dopp}} N_{\text{delay}}}$ form an orthonormal basis for the column space of $\mathbf{P}\mathbf{U}$ and the columns of $\mathbf{B}_d \in \mathbb{C}^{N \times (N - TN_{\text{Dopp}} N_{\text{delay}})}$ form an orthonormal basis for the left null space of $\mathbf{P}\mathbf{U}$. The pilot-data orthogonality of MMSE-PAT [recall (6.14)] implies that $\mathbf{B}_p^H \mathbf{H}^{[r]} \bar{\mathbf{B}} = \mathbf{0}$ and $\mathbf{B}_d^H \mathbf{H}^{[r]} \bar{\mathbf{p}} = \mathbf{0}$, $\forall r$. Projecting the r^{th} antenna observation vector $\mathbf{y}^{[r]}$ onto the data subspace, we obtain

$$\mathbf{y}_d^{[r]} = \sqrt{\rho} \mathbf{B}_d^H \mathbf{H}^{[r]} \bar{\mathbf{B}} \bar{\mathbf{s}} + \mathbf{B}_d^H \mathbf{v}^{[r]}.$$

Defining $\mathbf{v}_d^{[r]} = \mathbf{B}_d^H \mathbf{v}^{[r]}$ and $\bar{\mathbf{y}}_d = [\mathbf{y}_d^{[0]\top}, \dots, \mathbf{y}_d^{[R-1]\top}]^\top$, we have

$$\bar{\mathbf{y}}_d = \sqrt{\rho} \bar{\mathbf{B}}_d^H \bar{\mathbf{H}} \bar{\mathbf{B}} \bar{\mathbf{s}} + \bar{\mathbf{v}}_d, \quad (\text{E.5})$$

where $\bar{\mathbf{B}}_d = \text{diag}(\mathbf{B}_d, \dots, \mathbf{B}_d)$ and $\bar{\mathbf{v}}_d = [\mathbf{v}_d^{[0]\top}, \dots, \mathbf{v}_d^{[R-1]\top}]^\top$. Now, splitting $\bar{\mathbf{H}}$ into estimate and error component as $\bar{\mathbf{H}} = \hat{\mathbf{H}} + \tilde{\mathbf{H}}$, we have

$$\bar{\mathbf{y}}_d = \sqrt{\rho} \bar{\mathbf{B}}_d^H \hat{\mathbf{H}} \bar{\mathbf{B}} \bar{\mathbf{s}} + \underbrace{\sqrt{\rho} \bar{\mathbf{B}}_d^H \tilde{\mathbf{H}} \bar{\mathbf{B}} \bar{\mathbf{s}}}_{\bar{\mathbf{n}}} + \bar{\mathbf{v}}_d. \quad (\text{E.6})$$

Similar to the SISO case, with the mismatched decoder, $\bar{\mathbf{n}}$ acts as AWGN in the achievable rate expression [30]. With an i.i.d. Gaussian distribution on \mathbf{s} , we have the following mutual information bound [33]:

$$I(\bar{\mathbf{y}}_d; \bar{\mathbf{s}}) \geq \mathbb{E}\{\log \det[\mathbf{I} + \rho \sigma_s^2 \bar{\mathbf{R}}_n^{-1} \bar{\mathbf{B}}_d^H \hat{\mathbf{H}} \bar{\mathbf{B}} (\bar{\mathbf{B}}_d^H \hat{\mathbf{H}} \bar{\mathbf{B}})^H]\}, \quad (\text{E.7})$$

where $\bar{\mathbf{R}}_n = \mathbb{E}\{\bar{\mathbf{n}} \bar{\mathbf{n}}^H\}$. Since columns of $\bar{\mathbf{B}}$ are orthonormal, from the statistics of $\tilde{\mathbf{H}}$, it follows that $\bar{\mathbf{R}}_n \leq (1 + \rho \sigma_s^2 \frac{\text{tr}\{\bar{\mathbf{R}}_n\}}{NR}) \mathbf{I}$. Substituting this into (E.7), and normalizing $\hat{\mathbf{H}}$, we have

$$I(\bar{\mathbf{y}}_d; \bar{\mathbf{s}}) \geq \mathbb{E}\{\log \det[\mathbf{I} + \rho_{\text{lb}} \bar{\mathbf{B}}_d^H \check{\mathbf{H}} \bar{\mathbf{B}} (\bar{\mathbf{B}}_d^H \check{\mathbf{H}} \bar{\mathbf{B}})^H]\}. \quad (\text{E.8})$$

It follows that $\bar{\mathbf{B}}_p^H \check{\mathbf{H}} \bar{\mathbf{B}} = \mathbf{0}$ due to similarity in the structure of $\bar{\mathbf{H}}$ and $\check{\mathbf{H}}$. Using the fact that $\mathbf{B}_d \mathbf{B}_d^H = \mathbf{I}_N - \mathbf{B}_p \mathbf{B}_p^H$ and the determinant identity $\det(\mathbf{I} + \mathbf{G}_1 \mathbf{G}_2) = \det(\mathbf{I} + \mathbf{G}_2 \mathbf{G}_1)$, the lower bound on the mutual information (E.8) can be rewritten as (6.32).

For the upper bound on the mutual information, we consider the best case scenario of perfect channel estimates, i.e., $\hat{\mathbf{H}} = \mathbf{H}$ and $\text{tr}\{\mathbf{R}_{\hat{h}}\} = 0$. For this scenario, with i.i.d. Gaussian input distribution, it is straightforward [33] to obtain the bound (6.33).

E.3 Proof of Theorem 9

Since $\mathcal{R}_{\text{mimo-mmse-blk}} \geq \mathcal{R}_{\text{mimo-mmse-blk-lb}}$, we first characterize the asymptotic behavior of $\mathcal{R}_{\text{mimo-mmse-blk-lb}}$. Since (6.32) and (E.8) are equivalent, using Fatou's lemma [70] and the continuity of $\log \det(\cdot)$,

$$\lim_{\text{SNR} \rightarrow \infty} \mathcal{R}_{\text{mimo-mmse-blk-lb}} \geq \mathbb{E}\{\log \det \lim_{\rho \rightarrow \infty} [\mathbf{I} + \rho_{\text{lb}} (\bar{\mathbf{B}}_d^H \check{\mathbf{H}} \bar{\mathbf{B}})^H \bar{\mathbf{B}}_d^H \check{\mathbf{H}} \bar{\mathbf{B}}]\}, \quad (\text{E.9})$$

$$\geq \mathbb{E}\{\log \det \lim_{\rho \rightarrow \infty} [\mathbf{I} + \rho k (\bar{\mathbf{B}}_d^H \bar{\mathbf{H}} \bar{\mathbf{B}})^H \bar{\mathbf{B}}_d^H \bar{\mathbf{H}} \bar{\mathbf{B}}]\}, \quad (\text{E.10})$$

for some $k \in \mathbb{R}$, since $\lim_{\rho \rightarrow \infty} \text{tr}\{\mathbf{R}_{\hat{h}}\} = 0$ and $\lim_{\rho \rightarrow \infty} \check{\mathbf{H}} = \bar{\mathbf{H}}$ almost surely, with any *fixed* power allocation fraction $\alpha \in (0, 1)$. Since the “effective channel” $\mathbf{H}_e = \bar{\mathbf{B}}_d^H \bar{\mathbf{H}} \bar{\mathbf{B}} \in \mathbb{C}^{R(N - TN_{\text{Dopp}} N_{\text{delay}}) \times TN_s}$ is full rank with probability 1 and it follows that

$$\mathcal{R}_{\text{mimo-mmse-blk}}(\rho) \geq \mathcal{R}_{\text{mimo-mmse-blk-lb}}(\rho) \quad (\text{E.11})$$

$$= \min\{R(N - TN_{\text{Dopp}} N_{\text{delay}}), TN_s\} \log \rho + O(1). \quad (\text{E.12})$$

Now, to bound the asymptotic achievable rate from above, we consider the “coherent” case of zero channel estimation error, i.e., $\bar{\mathbf{H}}$ is perfectly known at the receiver. Since $I(\bar{\mathbf{y}}; \bar{\mathbf{s}}) = I(\bar{\mathbf{y}}_d; \bar{\mathbf{s}})$, using \mathcal{C}_{coh} to denote the capacity of (E.5) with the same power constraint, it is evident that $\mathcal{R}_{\text{mimo-mmse-blk}} \leq \mathcal{C}_{\text{coh}}$. Using the coherent capacity

result from [33], we have $\mathcal{C}_{\text{coh}} = \max_{\text{tr}\{\bar{\mathbf{R}}_s\} \leq N} E\{\log \det[\mathbf{I} + \rho \mathbf{H}_e \bar{\mathbf{R}}_s \mathbf{H}_e^H]\}$, where $\bar{\mathbf{R}}_s$ denotes the covariance of $\bar{\mathbf{s}}$. Note that, for any $\bar{\mathbf{R}}_s$ satisfying the power constraint, we have $\bar{\mathbf{R}}_s \leq N\mathbf{I}$, in the positive semi-definite sense. So, as $\rho \rightarrow \infty$,

$$\mathcal{R}_{\text{mimo-mmse-blk}}(\rho) \leq E\{\log \det[\mathbf{I} + \rho N \mathbf{H}_e^H \mathbf{H}_e]\} \quad (\text{E.13})$$

$$= \min\{R(N - TN_{\text{Dopp}}N_{\text{delay}}), TN_s\} \log \rho + O(1). \quad (\text{E.14})$$

E.4 Derivation of T_\star and R_\star

Notice that with optimal T_\star , we have $N - T_\star N_{\text{delay}} N_{\text{Dopp}} > 0$ since we consider underspread channel which satisfies $N - N_{\text{delay}} N_{\text{Dopp}} > 0$. So, we have $R(N - T_\star N_{\text{delay}} N_{\text{Dopp}}) \geq r(N - T_\star N_{\text{delay}} N_{\text{Dopp}}) \forall r \in \{0, \dots, R\}$ and hence an optimal value $R_\star = R$. Finding T_\star follows from geometrical arguments. It is clear that the pre-log factor is non-negative. As a (continuous) function of t , we plot the non-negative values $R_\star(N - tN_{\text{delay}}N_{\text{Dopp}})$ and tN_s in Fig. E.1. Notice that $R_\star(N - tN_{\text{delay}}N_{\text{Dopp}})$ is a straight line and $tN_s = t(N - (t+1)N_{\text{delay}}N_{\text{Dopp}} + \kappa)$ is a parabola as a function of t where $\kappa = 1, N_{\text{delay}}, N_{\text{Dopp}}$ for MIMO- Chirp, FDKD and TDKD schemes respectively. T_\star is an integer which maximizes the minimum of the two curves. There arises two possibilities depending on whether the two curves intersect or not. When the two curves intersect Fig. E.1 (a), it is clear that the pre-log factor increases until the first intersection point (at $t = \frac{\xi - \sqrt{\xi^2 - 4R_\star N N_{\text{Dopp}} N_{\text{delay}}}}{2N_{\text{Dopp}} N_{\text{delay}}}$) after which it decreases. Since T_\star is an integer which can not be more than the number of available antennas T , the optimal T_\star is given by $\min\left(T, \left\lfloor \frac{\xi - \sqrt{\xi^2 - 4R_\star N N_{\text{Dopp}} N_{\text{delay}}}}{2N_{\text{Dopp}} N_{\text{delay}}} \right\rfloor\right)$. When the two curves do not intersect Fig. E.1 (b), the pre-log factor increases until $t = \frac{\xi}{2N_{\text{delay}} N_{\text{Dopp}}} - \frac{R_\star}{2}$ after which it decreases, and T_\star is given by $\min\left(T, \left\lfloor \frac{\xi}{2N_{\text{delay}} N_{\text{Dopp}}} - \frac{R_\star}{2} \right\rfloor\right)$.

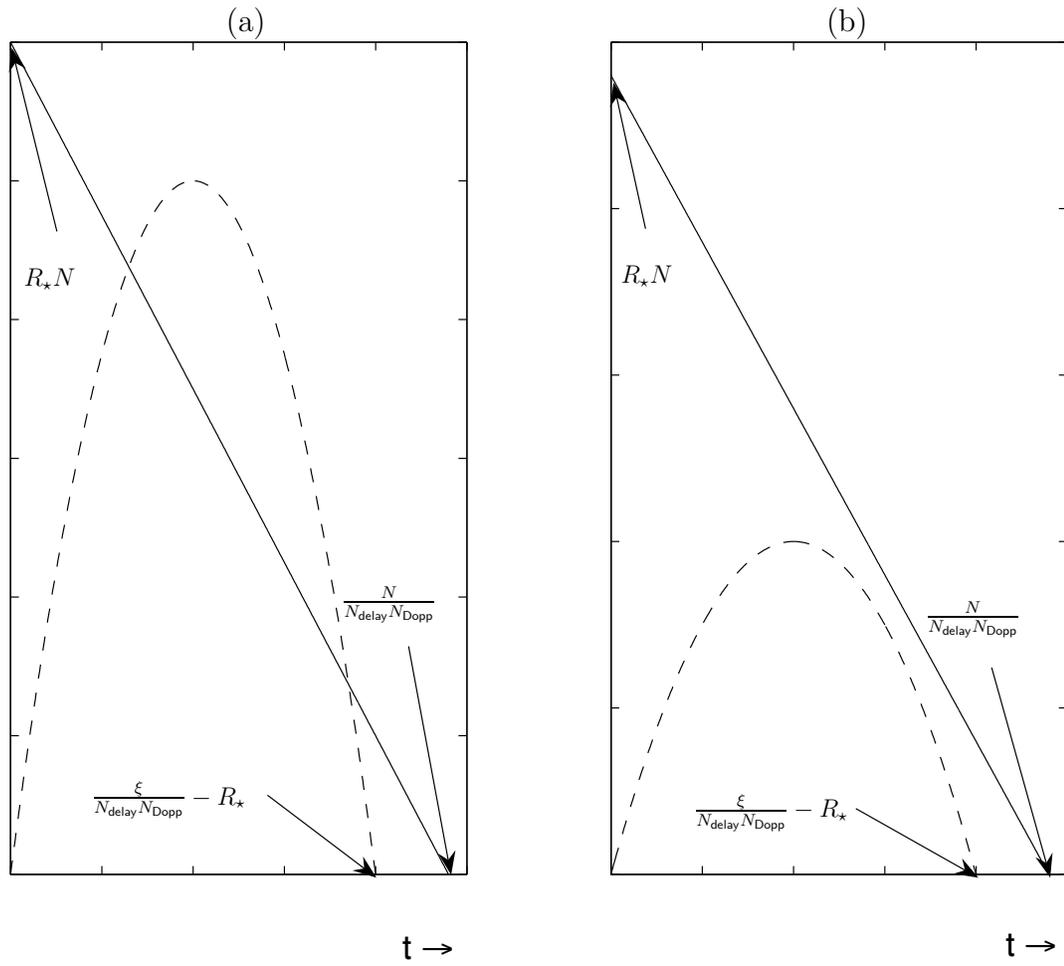


Figure E.1: Pre-log factor with number of active transmit antennas t .

BIBLIOGRAPHY

- [1] C. E. Shannon, “A mathematical theory of communication,” *Bell Labs Technical Journal*, vol. 27, pp. 379–423, 623–656, July, Oct. 1948.
- [2] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 4th ed., 2001.
- [3] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.
- [4] E. Biglieri, J. Proakis, and S. Shamai, “Fading channels: information-theoretic and communication aspects,” *IEEE Trans. on Information Theory*, vol. 44, pp. 2619–2692, Oct. 1998.
- [5] J. Cavers, “An analysis of pilot symbol assisted modulation for Rayleigh fading channels,” *IEEE Trans. on Vehicular Technology*, vol. 40, pp. 686–693, Nov. 1991.
- [6] L. Tong, B. M. Sadler, and M. Dong, “Pilot assisted wireless transmissions: general model, design criteria and signal processing,” *IEEE Signal Processing Magazine*, vol. 21, pp. 12–25, Nov. 2004.
- [7] L. Zheng and D. Tse, “Communication on the grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel,” *IEEE Trans. on Information Theory*, vol. 48, pp. 359–383, Feb. 2002.
- [8] P. A. Bello, “Characterization of randomly time-variant linear channels,” *IEEE Trans. on Communication Systems*, vol. 11, pp. 360–393, Dec. 1963.
- [9] Y. Liang and V. V. Veeravalli, “Capacity of noncoherent time-selective Rayleigh-fading channels,” *IEEE Trans. on Information Theory*, vol. 50, pp. 3095–3110, Dec. 2004.
- [10] W. C. Jakes, *Microwave Mobile Communications*. Wiley, 1974.
- [11] M. K. Tsatsanis and G. B. Giannakis, “Modeling and equalization of rapidly fading channels,” *Internat. Journal of Adaptive Control & Signal Processing*, vol. 10, pp. 156–176, Mar. 1996.

- [12] A. Sayeed and B. Aazhang, “Joint multipath-Doppler diversity in mobile wireless communications,” *IEEE Trans. on Communications*, vol. 47, pp. 123–132, Jan. 1999.
- [13] X. Ma, G. B. Giannakis, and S. Ohno, “Optimal training for block transmissions over doubly-selective wireless fading channels,” *IEEE Trans. on Signal Processing*, vol. 51, pp. 1351–1366, May 2003.
- [14] D. K. Borah and B. D. Hart, “Frequency-selective fading channel estimation with a polynomial time-varying channel model,” *IEEE Trans. on Communications*, vol. 47, pp. 862–873, June 1999.
- [15] T. Zemen and C. F. Mecklenbräuker, “Time-variant channel estimation using discrete prolate spheroidal sequences,” *IEEE Trans. on Signal Processing*, vol. 53, pp. 3597–3607, Sept. 2005.
- [16] T. A. Thomas and F. W. Vook, “Quasi-static antenna array processing for rapidly time-varying channels,” in *Proc. IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 2977–2980, May 2001.
- [17] P. Bello, “Time-frequency duality,” *IEEE Trans. on Information Theory*, vol. 10, pp. 18–33, Jan. 1964.
- [18] D. B. Kilfoyle and A. B. Baggeroer, “The state of the art in underwater acoustic telemetry,” *IEEE Journal of Oceanic Engineering*, vol. 25, pp. 4–27, Jan. 2000.
- [19] T. Marzetta and B. Hochwald, “Capacity of a mobile multiple-antenna communication link in Rayleigh-fading channels,” *IEEE Trans. on Information Theory*, vol. 45, pp. 139–157, Jan. 1999.
- [20] H. Vikalo, B. Hassibi, B. Hochwald, and T. Kailath, “On the capacity of frequency-selective channels in training-based transmission schemes,” *IEEE Trans. on Signal Processing*, vol. 52, pp. 2572–2583, Sep. 2004.
- [21] R. S. Kennedy, *Fading Dispersive Communication Channels*. New York: Wiley-Interscience, 1969.
- [22] E. Telatar and D. Tse, “Capacity and mutual information of wideband multipath fading channels,” *IEEE Trans. on Information Theory*, vol. 46, pp. 1384–1400, July 2000.
- [23] S. Verdú, “Spectral efficiency in the wideband regime,” *IEEE Trans. on Information Theory*, vol. 48, pp. 1319–1343, June 2002.
- [24] M. Medard and R. Gallager, “Bandwidth scaling for fading multipath channels,” *IEEE Trans. on Information Theory*, vol. 48, pp. 840–852, Apr. 2002.

- [25] V. G. Subramanian and B. Hajek, "Broadband fading channels: signal burstiness and capacity," *IEEE Trans. on Information Theory*, vol. 48, pp. 908–927, Apr. 2002.
- [26] V. Sethuraman and B. Hajek, "Capacity per unit energy of fading channels with a peak constraint," *IEEE Trans. on Information Theory*, vol. 51, pp. 3102–3120, Sep. 2005.
- [27] G. Durisi, H. Bölcskei, and S. Shamai, "Capacity of underspread WSSUS channels in the wideband regime," in *Proc. IEEE Internat. Symposium on Information Theory*, July 2006.
- [28] R. H. Etkin and D. N. C. Tse, "Degrees of freedom of some underspread MIMO fading channels," *IEEE Trans. on Information Theory*, vol. 52, pp. 1576–1608, Apr. 2006.
- [29] J. Manton, I. Mareels, and Y. Hua, "Affine precoders for reliable communications," in *Proc. IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 2749–2752, June 2000.
- [30] H. Weingarten, Y. Steinberg, and S. Shamai(Schitz), "Gaussian codes and weighted nearest neighbor decoding in fading multiple antenna channels," *IEEE Trans. on Information Theory*, vol. 50, pp. 1665–1686, Aug. 2004.
- [31] N. Merhav, G. Kaplan, A. Lapidoth, and S. S. (Shitz), "On information rates of mismatched decoders," *IEEE Trans. on Information Theory*, vol. 40, pp. 1953–1967, Nov. 1994.
- [32] A. Lapidoth and S. Shamai, "Fading channels: how perfect need 'perfect side information' be," *IEEE Trans. on Information Theory*, vol. 48, pp. 1118–1134, May 2002.
- [33] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. on Telecommunications*, vol. 10, pp. 585–595, Nov. 1999.
- [34] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM system," in *IEEE Trans. on Consumer Electronics*, vol. 44, pp. 1122–1128, Aug. 1998.
- [35] S. Ohno and G. B. Giannakis, "Optimal training and redundant precoding for block transmissions with applications to wireless OFDM," *IEEE Trans. on Communications*, vol. 50, pp. 2113–2123, Dec. 2002.
- [36] H. Minn and N. Al-Dhahir, "Optimal training signals for MIMO OFDM channel estimation," *IEEE Trans. on Wireless Communication*, vol. 5, pp. 1158–1168, May 2006.

- [37] M. Dong, L. Tong, and B. M. Sadler, "Optimal insertion of pilot symbols for transmission over time-varying flat fading channels," *IEEE Trans. on Signal Processing*, vol. 52, pp. 1403–1418, May 2004.
- [38] M. Dong, L. Tong, and B. M. Sadler, "Optimal pilot placement for channel tracking in OFDM," in *Proc. IEEE Military Communication Conf.*, Oct. 2002.
- [39] J.-W. Choi and Y.-H. Lee, "Optimal pilot pattern for channel estimation in OFDM systems," *IEEE Trans. on Wireless Communication*, vol. 4, pp. 2083–2088, Sep. 2005.
- [40] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. on Signal Processing*, vol. 51, pp. 1615–1624, June 2003.
- [41] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links," *IEEE Trans. on Information Theory*, vol. 49, pp. 951–963, Apr. 2003.
- [42] X. Ma, L. Yang, and G. B. Giannakis, "Optimal training for MIMO frequency-selective fading channels," *IEEE Trans. on Wireless Communication*, vol. 4, pp. 453–466, Mar. 2005.
- [43] L. Yang, X. Ma, and G. B. Giannakis, "Optimal training for MIMO fading channels with time- and frequency- selectivity," in *Proc. IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 821–824, May 2004.
- [44] S. Misra, A. Swami, and L. Tong, "Cutoff rate analysis of the Gauss-Markov fading channel with adaptive energy allocation," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communication*, pp. 388–392, June 2003.
- [45] S. Ohno and G. B. Giannakis, "Capacity maximizing MMSE-optimal pilots for wireless OFDM over frequency-selective block rayleigh-fading channels," *IEEE Trans. on Information Theory*, vol. 50, pp. 2138–2145, Sep. 2004.
- [46] A. P. Kannu and P. Schniter, "Capacity analysis of MMSE pilot-aided transmission for doubly selective channels," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communication*, 2005.
- [47] A. P. Kannu and P. Schniter, "Design and analysis of MMSE pilot-aided cyclic-prefixed block transmissions for doubly selective channels," *IEEE Trans. on Signal Processing*. (submitted Oct. 2005).
- [48] A. P. Kannu and P. Schniter, "MSE-optimal training for linear time-varying channels," in *Proc. IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing*, Mar. 2005.

- [49] A. P. Kannu and P. Schniter, “On the spectral efficiency of noncoherent doubly selective channels,” in *Proc. Allerton Conf. on Communication, Control, and Computing*, 2006.
- [50] L. Zheng, D. Tse, and M. Medard, “Channel coherence in the low-SNR regime,” *IEEE Trans. on Information Theory*, vol. 53, pp. 976–997, Mar. 2007.
- [51] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [52] A. Lapidoth and S. M. Moser, “Capacity bounds via duality with applications to multiple antenna systems on flat fading channels,” *IEEE Trans. on Information Theory*, vol. 49, pp. 2426–2467, Oct. 2003.
- [53] L. L. Scharf, *Statistical Signal Processing*. Reading, MA: Addison-Wesley, 1991.
- [54] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer, 2nd ed., 1994.
- [55] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.
- [56] X. Ma, G. Leus, and G. B. Giannakis, “Space-time-Doppler block coding for correlated time-selective fading channels,” *IEEE Trans. on Signal Processing*, vol. 53, pp. 2167–2182, June 2005.
- [57] S. Ohno and G. B. Giannakis, “Average rate optimal PSAM transmissions over time selective fading channels,” *IEEE Trans. on Wireless Communication*, vol. 1, pp. 712–720, Oct. 2002.
- [58] B. Hart and D. P. Taylor, “Extended MLSE diversity receiver for the time and frequency selective channel,” *IEEE Trans. on Communications*, vol. 45, pp. 322–333, Mar. 1997.
- [59] A. Stamoulis, S. N. Diggavi, and N. Al-Dhahir, “Estimation of fast fading channels in OFDM,” in *Proc. IEEE Wireless Communication and Networking Conf.*, vol. 1, pp. 465–470, Mar. 2002.
- [60] S. Barbarossa and A. Scaglione, “Theoretical bounds on the estimation and prediction of multipath time-varying channels,” in *Proc. IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 2545–2548, June 2000.
- [61] M. Medard, “The effect upon channel capacity in wireless communication of perfect and imperfect knowledge of the channel,” *IEEE Trans. on Information Theory*, vol. 46, pp. 933–946, May 2000.

- [62] A. P. Kannu and P. Schniter, “Minimum mean squared error pilot aided transmission for MIMO doubly selective channels,” in *Proc. Conf. on Information Science and Systems*, 2006.
- [63] P. Schramm, “Analysis and optimization of pilot-channel-assisted BPSK for DS-CDMA systems,” *IEEE Trans. on Communications*, vol. 46, pp. 1122–1124, Sep. 1998.
- [64] P. Schramm and R. Muller, “Pilot symbol assisted BPSK or Rayleigh fading channels with diversity: Performance analysis and parameter optimization,” *IEEE Trans. on Communications*, vol. 46, pp. 1560–1563, Dec. 1998.
- [65] J. Evans, “Optimal resource allocation for pilot symbol aided multiuser receivers in Rayleigh faded CDMA channels,” *IEEE Trans. on Communications*, vol. 50, pp. 1316–1325, Aug. 2002.
- [66] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York, NY: Cambridge University Press, 1985.
- [67] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 4th ed., 2001.
- [68] J. Kotecha and A. Sayeed, “Transmit signal design for optimal estimation of correlated MIMO channels,” *IEEE Trans. on Signal Processing*, vol. 52, pp. 514–557, Feb. 2004.
- [69] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*. New York: Wiley-Interscience, John Wiley and Sons, 2 ed., 1999.
- [70] J. C. Taylor, *An introduction to Probability and Measure*. New York: Springer, 1998.