Training in Massive MIMO Systems

Wan Amirul Wan Mohd Mahyiddin

A thesis submitted for the degree of

Doctor of Philosophy

in

Electrical and Electronic Engineering

University of Canterbury

New Zealand

2015

Abstract

Massive multiple-input multiple-output (MIMO) systems have been gaining interest recently due to their potential to achieve high spectral efficiency [1]. Despite their potential, they come with certain issues such as pilot contamination. Pilot contamination occurs when cells simultaneously transmit the same pilot sequences, creating interference. Unsynchronizing the pilots can reduce pilot contamination, but it can produce data to pilot interference. This thesis investigates the impact of pilot contamination and other interference, namely data to pilot interference, on the performance of finite massive MIMO systems with synchronized and unsynchronized pilots. Two unsynchronized pilot schemes are considered. The first is based on an existing time-shifted pilot scheme, where pilots overlap with downlink data from nearby cells. The second time-shifted method overlaps pilots with uplink data from nearby cells. Results show that if there are small numbers of users, the first time-shifted method provides the best sum rate performance. However, for higher numbers of users, the second time-shifted method provides better performance than the other methods. We also show that time-synchronized pilots are not necessarily the worst case scenario in terms of sum rate performance when shadowing effects are considered.

The wireless channel can be time and frequency varying due to the Doppler effect from mobile user equipment (UE) and a multipath channel. These variations can be simulated by using a selective channel model, where the channel can vary within the coherence block in both time and frequency domains. The block fading channel model approximates these variations by assuming the channel stays constant within a coherence block, but changes independently between blocks [2]. Due to its simplicity, the block fading model is widely used in massive MIMO studies [3–8]. Our research compares the impact of block fading and time-selective fading channel models in massive MIMO systems. To achieve this, we derive a novel closed form sum rate expression for time-selective channels. Results show that there are significant differences in sum rate performance between these models.

In addition to time variation from Doppler effect, the channel can also experience frequency variation due to delay spread from multipath signal propagation. The combination of time and frequency selective channels can be described as a doubly-selective channel. Hence, the sum rate expression for time-selective channels can also be extended

to doubly-selective channels. We investigate two types of pilot sequences, namely constant amplitude pilots and zero padded pilots in doubly-selective channels. Results show that a zero padded pilot has a better sum rate performance than a constant amplitude pilot for a wide range of antenna numbers and time-frequency correlation values. Two different type of training optimization, namely average optimum training and adaptive optimum training, are investigated. Both methods shows similar sum rate performance. In addition, we also study the effect of increasing frequency reuse and the pilot reuse factor. Even though these methods can reduce intercell interference, they also result to lower sum rate due to inefficient use of time-frequency resources.

Acknowledgements

I would like to express my sincerest appreciation to my supervisors, Assoc. Prof. Philippa A. Martin and Prof. Peter J. Smith, for their relentless support and guidance throughout my doctorate study. The knowledge and expertise that they have given me has definitely improved my research skill, especially in wireless communication field. I am also very grateful to University of Malaya and Malaysian Ministry of Education for providing scholarship and other financial supports for my doctorate study. I would also like to thank Assoc. Prof. Philippa A. Martin along with Department of Electrical and Computer Engineering, University of Canterbury, for funding me to present my research at IEEE Vehicular Technology Conference 2014 in Vancouver, Canada. I extend my gratitude to the editors and reviewers of my journals and conference paper for their insightful commentaries which have helped me to improve my research. Last but not least, I am thankful for the people in communications research group, university staffs and my friends for their help and support.

Contents

Abstract ii

Acknowledgements iv

Contents vi

List of Figures x

List of Tables xiii

Abbreviations xiv

Notations xvi

1 Introduction 1

1.1 Massive MIMO issues 2

1.2 Problem statements 5

1.3 Thesis contributions 5

1.4 Thesis outline 6

1.5 Publications 8

2 Background model and assumptions 9

2.1 Channel model 9

2.1.1 Large scale fading 11

2.1.2 Small scale fading 11

2.1.3 Block fading channel and selective fading channel 12

2.2 MIMO 13

2.2.1 Antenna spacing 14

2.2.2 MU-MIMO 15

2.3 Massive MIMO 16

2.4 Channel estimation 17

2.4.1 Channel reciprocity 18

2.4.2 LMMSE channel estimation 19

2.5 Multicell system 20

2.6 OFDM system 21

2.7 Summary 22

3 Time-shifted pilots in massive MIMO **23**

3.1 Introduction 23

3.2 System model 25

3.3 Time-shifted pilot with downlink data overlap 26

3.3.1 Channel estimation 27

3.3.2 Downlink transmission rate 32

3.3.3 Uplink transmission rate 37

3.3.4 Power optimization 42

3.4 Time-shifted pilot with uplink data overlap 43

3.4.1 Channel estimation 44

3.4.2 Downlink transmission rate 47

3.4.3 Uplink transmission rate 50

3.4.4 Power optimization 53

3.5 Time-synchronized pilot 54

3.6 Numerical results 56

3.7 Summary 61

4 Massive MIMO systems in time-selective channels 65

4.1 Introduction 65

4.2 System model 66

4.3 Channel estimation 67

4.4 Achievable sum rate 70

4.5 Results 74

4.6 Summary 78

5 Massive MIMO systems in time and frequency selective channels 81

5.1 Introduction 81

5.2 System model 83

5.3 Channel estimation 84

5.4 Achievable sum rate 89

5.5 Training sequences and optimizations 94

5.5.1 Constant amplitude pilot 94

5.5.2 Zero padded pilot 95

5.5.3 Training optimization 96

5.5.4 Frequency and pilot reuse 97

5.6 Numerical results 99

5.7 Summary 107

6 Conclusions and future works **109**

6.1 Conclusions 109

6.2 Future works 110

6.2.1 Spatial correlation 110

6.2.2 Millimeter wave 111

6.2.3 Cell radius 112

6.2.4 Coding scheme 113

6.3 Summary 114

A Proof of Lemma 3.2 **115**

B Proof of uncorrelated channel estimation error **117**

C Lemma C.1 **119**

D Power optimization **121**

D.1 Power optimization for TShdown 121

D.2 Power optimization for TShup 122

Bibliography **123**

List of Figures

1.1 Pilot contamination. 3

2.1 Multipath signal propagation from transmitter to receiver in a SISO channel. 10
2.2 Combined path-loss, shadowing (large scale fading) and multipath (small scale fading) versus distance (d) [32]. 10
2.3 Time variation of small-scale fading power for a block fading channel and a selective fading channel with 150 km/h UE speed. 12
2.4 MIMO transmission through the channel $\boldsymbol{H}$. 13
2.5 Hexagonal multicell structure. 20
2.6 OFDM transmission structure in time and frequency domain. 21

3.1 Cell group arrangement for 7 cell system. 26
3.2 The arrangement of time-shifted pilots for TShdown. 27
3.3 Received signal at cell j BS during group A_1 cells uplink pilot transmission phase for TShdown. 27
3.4 Received signal at UE k' in cell j during the downlink data transmission phase for group A_1 cells with TShdown. 33
3.5 Received signal at BS in cell j during group A_1 cells uplink data transmission phase for TShdown. 37
3.6 The arrangement of time-shifted pilots for TShup. 44
3.7 Received signal at BS in cell j during the uplink pilot transmission phase for group A_1 cells with TShup. 45
3.8 Received signal at UE k' in cell j during the downlink data transmission phase for group A_1 cells with TShup. 47
3.9 Received signal at BS in cell j during the uplink data transmission phase for group A_1 cells with TShup. 51

3.10 7 cells group arrangement for TSync. 55
3.11 The arrangement of time-shifted pilots for TSync. 55
3.12 Downlink and uplink SINR for the TShdown and TShup method using Monte Carlo (MC) and closed form (CF) simulations. 57
3.13 Average downlink and uplink sum rate for TSync, TShdown and TShup and without power optimization (OP). 58
3.14 Distribution for optimum downlink data power for TShdown generated using (3.50) with 100 BS antennas, 15 UEs and 10^4 random UEs location drops. 62
3.15 Distribution for optimum uplink data power for TShup generated using (3.79) with 100 BS antennas, 15 UEs and 10^4 random UEs location drops. 62
3.16 Power optimization (OP) using exhaustive search and equation-based methods for TShdown and TShup for 15 UEs. 63

4.1 Pilot and data arrangement in one frame. 66
4.2 Optimum frame length, data length and pilot length for the time-selective channels with $\delta = 0.95$ and $\delta = 0.99$. 75
4.3 Optimum frame length, data length and pilot length for the block-fading channels with $\delta = 0.95$ and $\delta = 0.99$. 76
4.4 Performance of block-fading sum rate using optimum training for block-fading (BL-BL), time-selective sum rate using optimum training for block-fading (TS-BL) and time-selective sum rate using optimum training for time-selective (TS-TS). The lines are sum rates from analytical simulations, while each point marked with a '$\times$' is from Monte Carlo simulations. 76
4.5 Optimum frame length, data length and pilot length for the time-selective channels with 100 antennas st BS for various values of δ. 78
4.6 Sum rate with optimum training for the time-selective channels with 100 antennas st BS for various values of δ. 78

5.1 Time-frequency arrangement of the transmission block. T_{f}, T_{p}, T_{d} are the frame, pilot and data length in the time domain (number of symbols) and F is the length of the training/data block in the frequency domain (number of subcarriers). Each small square represents a resource element which is occupied by a pilot/data symbol. 83

5.2 Uplink pilot signal received at BS in cell l from all UEs including those in nearby cells. 85
5.3 Downlink data signal received at UE k' in cell j from all BSs in surrounding cells. 89
5.4 Cell arrangement with frequency reuse factor 3. F_1, F_2 and F_3 represents different frequency band that is used by each cells. 98
5.5 Sum rate performance using CA and ZP pilots. The lines are obtained using closed form expressions in (5.18) while points marked '$\times$' are obtained using Monte Carlo simulation. 100
5.6 Optimum F, T_{p}, and T_{d} values for CA pilots, CA pilots with power control and ZP pilots. 101
5.7 Sum rate performance for CA pilot, CA pilot with the power control and ZP pilot for various number of antennas. 102
5.8 Sum rate performance for CA pilot, ZP pilot and combination of CA and ZP pilots for various number of antennas. 103
5.9 Sum rate performance for CA pilots, CA pilots with power control and ZP pilot for various σ_{c} and μ_{c} values using 100 BS antennas. 104
5.10 Sum rate performance for average optimal training and adaptive optimal training for various number of UEs using 100 BS antennas. 105
5.11 Optimum F, T_{p}, and T_{d} values for various frequency and pilot reuse factor. 106
5.12 Sum rate performance for various frequency and pilot reuse factor. 107

List of Tables

3.1 Percentage change in the average sum rate with respect to TSync for 200 antennas 60

5.1 Transmission variables 99

Abbreviations

AWGN	Additive white Gaussian noise.
BS	Base station.
CA	Constant amplitude.
CF	Closed form.
CSI	Channel state information.
CSCG	Circularly symmetric complex Gaussian.
FDD	Frequency division duplex.
FFT	Fast Fourier transform.
iid	Independent and identically distributed.
LMMSE	Linear minimum mean squared error.
LS	Least square.
LTE	Long Term Evolution.
MC	Monte Carlo.
MF	Matched filter.
MIMO	Multiple-input multiple-output.
MMSE	Minimum mean squared error.
MU-MIMO	Multiuser multiple-input multiple-output.
OFDM	Orthogonal frequency-division multiplexing.
OFDMA	Orthogonal frequency-division multiple access.
OP	Optimum power.
SISO	Single-input single-output.
SDMA	Space-division multiple access.
TDD	Time division duplex.
TDMA	Time division multiple access.
TShdown	Time-shifted pilot with downlink data overlap.

TShup	Time-shifted pilot with uplink data overlap.
TSync	Time-synchronized pilot.
UE	User equipment.
ZF	Zero forcing.
ZP	Zero padded.

Notations

$\boldsymbol{I}_M$	$M \times M$ identity matrix.
$\mathbf{0}_M$	$M \times M$ zero matrix.
$(\cdot)^T$	Matrix transpose.
$(\cdot)^H$	Hermitian transpose.
$\lvert \cdot \rvert$	Absolute value.
$\det\lvert \cdot \rvert$	Determinant.
$\lVert \cdot \rVert$	Two-norm.
$J_0(\cdot)$	Zeroth order Bessel function of the first kind.
$\mathbb{E}[\cdot]$	Expectation.
$(\cdot)^*$	Conjugate value.
$\mathrm{tr}(\cdot)$	Trace of a matrix.
$\gcd(x, y)$	Greatest common denominator of integers x and y.
lim	Limit.
$\log_n$	Logarithm base n.
$\sim$	Is distributed as.
$\mathcal{CN}(a, b)$	CSCG random variable with mean a and variance b.
$\mathcal{N}(a, b)$	Real Gaussian random variable with mean a and variance b.
$\boldsymbol{R_{xy}}$	Cross-covariance between vector $\boldsymbol{x}$ and $\boldsymbol{y}$.
$\Gamma(\cdot)$	Gamma Function.
$\rightarrow$	Approaches.
$\mathrm{vec}(\cdot)$	Vectorized matrix.
$\in$	Is an element of.
$\notin$	Is not an element of.

Notes

- Italic bold upper case letters represents matrices while italic bold lower case letters represent vectors.
- $\sum_{\{a_1,a_2\cdots a_N\}=1}^{\{A_1,A_2\cdots A_N\}} B(a_1,a_2\cdots a_N)$ means the summation of $B(a_1,a_2\cdots a_N)$ from $a_1=1$ to $a_1=A_1$, $a_2=1$ to $a_2=A_2$, and so on until $a_N=1$ to $a_N=A_N$.

Chapter 1

Introduction

The rapid growth of users in wireless cellular networks along with the greater usage of data for multimedia consumption demands a higher rate of data transmission. Current trends show that these demands grow exponentially with time and analysts predict that the traffic size can potentially increase by around 50% per year [9]. Keeping up with these demands will be an uphill task for wireless operators as the spectrum bandwidth becomes a scarce and expensive resource [10]. Wireless issues such as loss of performance from fading and interference further complicate providing higher throughput [11]. To overcome these challenges, various technologies are being investigated [12–16].

Multiple antenna systems, commonly known as MIMO systems, have been shown to increase system capacity [18–22] and small systems are used in the latest wireless standards, such as WiFi [11]. The basic idea of MIMO is to use more than one antenna at the transmitter and receiver in a wireless transmission system [17]. Multipath propagation of radio signals from transmitter to receiver results in diverse spatial characteristics [23]. MIMO exploits this property to enhance the transmission performance using diversity schemes or increases the capacity of the system through spatial multiplexing. Diversity can be produced using space-time coding to improve reliability and quality of wireless transmission [24, 25]. Spatial multiplexing increases transmission rate by using multiple independent streams of data in the same time-frequency resources [11]. MIMO has become an important part of wireless technology standards such as WiFi, Worldwide Interoperability for Microwave Access (WiMAX) and 3GPP Long-Term Evolution (3GPP-LTE) [11].

The improvements provided by conventional MIMO systems are not enough to keep up with the rapid increases of future traffic. Hence, massive MIMO systems have received significant attention recently due to their ability to achieve high spectral efficiency on a much greater scale than conventional MIMO [1]. The concept of massive MIMO comes from the mathematical framework which demonstrates that the total capacity gain from spatial multiplexing can be increased simply by adding antennas to the transmission system [2]. It was shown in [26] that under certain conditions, it is always beneficial to add more antennas at the base station (BS). The additional antennas enable the transmission energy to be focused into much smaller target areas which results in higher throughput and reduces interference in other area [27]. Such findings have encouraged interest in studying the effect of increasing the number of antennas to a massive scale. A field test in [27] has shown that it is possible to use more than 100 antennas at the BS.

1.1 Massive MIMO issues

Since MIMO can transmit multiple streams of independent data using the same frequency resource at the same time, there will be inevitable signal overlapping when the data arrives at the receiver. Channel state information (CSI) is a vital part of MIMO as it enables us to resolve the overlapped signals by performing equalization at the receiver or beamforming at the transmitter [37]. The capacity in (2.12) assumes that there is no CSI at the transmitter, but perfect CSI is available at the receiver. Practically, CSI will not be perfect because it is obtained through channel estimation. For example, in a training-based MIMO system, CSI is obtained by transmitting training sequences or pilots, which will then be used to estimate the CSI. In the multicell scenario, UEs from different cells may use the same pilot which results in pilot contamination [4]. The problem basically arises when the number of orthogonal pilot sequences available for users is limited due to a finite coherence time. The overhead data from pilot training will affect the channel capacity and it has been shown that the optimum number of pilots in one coherence frame is equal to the number of transmit antennas [38].

Since the length of pilot sequences is limited, there will be a need for pilot reuse which results in pilot contamination. Figure 1.1 illustrates the occurrence of pilot contamination where two users from different adjacent cells simultaneously transmit the same

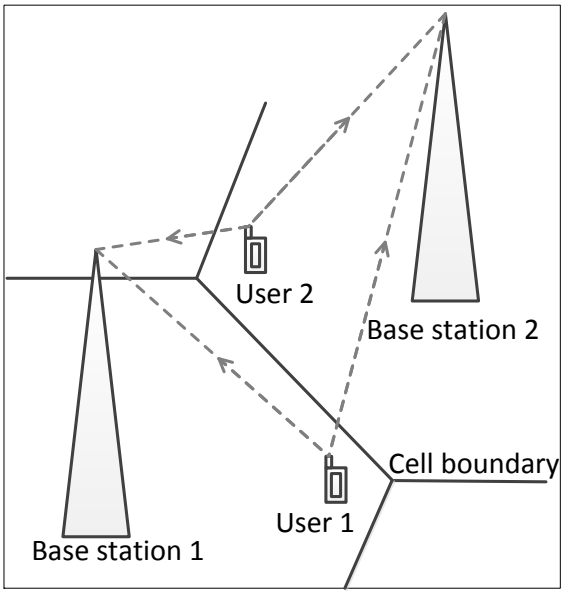

FIGURE 1.1: Pilot contamination.

pilot sequence which is received by both of the BSs. As a result, transmissions from one user in one cell cannot be distinguished from those coming from a different cell. The effective interference from pilot contamination scales linearly with the number of antennas [4] and this limits the performance of massive MIMO. In [39] it is shown that the problem of pilot contamination only exists due to the usage of inappropriate linear channel estimation. Hence, a scheme based on blind pilot decontamination is proposed to eliminate the pilot contamination effect. However, [27] argues that under specific power control assumptions, pilot contamination is still a problem that must be dealt with.

In addition to pilot contamination, there are several other issues related to massive MIMO. Since there can be hundreds of antennas in a massive MIMO system, naturally there will be concerns regarding the practicality of such a system. For example, one of the main concerns is computational complexity. A massive number of antennas means that there are a large number of signals that need to be processed. For an optimal signal detector, the computational complexity grows exponentially with the number of transmitters [40]. For precoding, a non-linear method can provide better performance than linear precoding, but non-linear signal processing is harder to implement in massive MIMO due to computational complexity [1]. However, results in [4] show that linear precoders such as matched filter (MF) and zero forcing (ZF) can achieve near optimal

performance when the number of antennas becomes very large. Therefore, the issue of calculation complexity can be handled by using simple signal processing tools such as linear detectors and precoders which provide reliable performance in massive MIMO [41]. Therefore, this thesis will use a simple processing tool, namely MF, to recover the data signal.

Another issue that may arise when dealing with large numbers of antennas is energy consumption. This comes from the fact that a power supply is needed for each one of the hundreds of antennas to transmit the radio signals. However, results from [5] have shown that as the number of antennas increases, the amount of power per unit transmission rate is reduced. Therefore, massive MIMO can have beneficial consequences in terms of power consumption. A practical issue that may also arise is the construction cost of such a large antenna system. In conventional MIMO, high quality antennas are used in order to avoid antenna failure, which would result in a significant loss of performance. This means that each antenna can be expensive to build. Fortunately, the high quality antenna criteria can be relaxed in massive MIMO systems due to the law of large numbers, which means that massive MIMO has a greater tolerance to antenna imperfection [42]. In addition, energy consumption per antenna is significantly lower than conventional MIMO. This means that we can use cheaper materials to build the low power antennas. Another concern for massive MIMO systems is that the total physical size of the array may become very large. The antennas are required to be spaced at a certain distance in order to avoid mutual coupling and antenna correlation. This requirement can be a problem if there are hundreds of antennas at the BS. However, a field test in [33] has shown that massive MIMO performance can be achieved while maintaining the physical size within a practical limit.

To summarize, despite its potential, massive MIMO also comes with several issues such as pilot contamination, computational complexity and energy consumption. Although various studies have given valid potential solutions to these issues, pilot contamination are still a limitation to massive MIMO systems. Hence, massive MIMO is a work in progress and further improvement can still be made.

1.2 Problem statements

As discussed in 1.1, a major limiting factor with the training schemes in massive MIMO is pilot contamination [1]. Various detailed studies have been done to examine this issue [1, 4–8]. In [1, 4], the worst case scenario is assumed to be when adjacent cells send the same pilot sequence at the same time. To avoid the problem of synchronized pilots between neighboring cells, [3] proposed a time-shifted pilot method, where some cells send downlink data while others transmit pilots. Although the method can significantly improve the transmission rate, the analysis in [3] is based on the assumption of an infinite number of antennas at the BS. As the number of antennas goes to infinity, the interference from different cell groups during data transmission becomes negligible and the rate performance can be overestimated. In the case of a limited number of antennas at the BS, we cannot ignore the impact of the aforementioned interference on the transmission performance. Therefore, we aim to analyze the effect of massive, but finite MIMO systems on the time-shifted performance.

Another issue that arises when designing the channel training is related to the channel models that are used to analyze the performance of massive MIMO. For a moving user equipment (UE), the channel may be time varying due to the Doppler effect. The channel may also experience frequency variation due to different delays in the multipath channel. In order to simulate the time and frequency varying channel, researchers often use a block fading model which simplifies the channel selectivity by assuming the channel to be constant within a coherence block and to vary independently from block to block. The block fading model is also widely used in massive MIMO research [4–7] as it can greatly simplify analysis. Since the block fading model is an approximation of the selective channel environment [2], training optimization using a block fading model may not be accurate, such as in the case of high speed UEs. Therefore, our research considers the channel selectivity in order to design the training sequence.

1.3 Thesis contributions

As discussed in the problem statements, the original time-shifted pilot research [3] only analyzed the performance of the method using infinite number of antennas. Therefore, this thesis aims to investigate the impact of finite number of antennas on the time-shifted

method. To achieve this, we have derived a novel closed form lower bound ergodic sum rate for the time-shifted methods with finite number of antennas. The original time-shifted pilot method overlaps the uplink pilot with downlink data from other cells. In this thesis, we also investigate another variation of time-shifted pilot method where the uplink pilot is overlapped with uplink data from different cells. The closed form lower bound ergodic sum rate is also derived for this time-shifted method. So far, there has been no other research, which provides detailed sum rate expressions and performance comparisons of the two time-shifted methods.

Another issue mentioned in the problem statements is the channel model. Most of recent works on massive MIMO analyze its performance using the block fading model [4–7]. However, as we discussed in the problem statement, the block fading is just an approximation of selective fading and may not provide accurate performance analysis. To provide more accurate analysis, this thesis has derived a novel closed form lower bound ergodic rate for the time selective channel. The key contribution of this investigation is that we have proven there are significant differences between sum rate of that obtained using the block fading model and selective fading model. In addition, we extend the sum rate expression of the time selective model to doubly selective model in multicell scenario. Based on the sum rate derivation, we are able to analyze optimal training size in both time and frequency domain.

1.4 Thesis outline

The outline of the thesis is as follows:

In Chapter 2, we provide a general discussion of the background model and assumptions we use throughout this thesis. These are mainly related to the physical layer aspects of the transmission such as the channel model and multiple antenna systems.

In Chapter 3, we analyze the impact of pilot contamination and other interference, namely data to pilot interference, on the performance of finite massive MIMO systems with synchronized and unsynchronized pilots. Two unsynchronized pilot schemes are considered. The first is based on an existing time-shifted pilot scheme, where pilots overlap with downlink data from nearby cells. The second time-shifted method overlaps pilots with uplink data from nearby cells. Results show that if there are small numbers

of users, the first time-shifted method provides the best sum rate performance. However, for higher numbers of users, the second time-shifted method has advantages compared to other methods. We also show that a time-synchronized pilot is not necessarily the worst case scenario in term of sum rate performance when shadowing effects are considered.

In Chapter 4, we investigate the performance of massive MIMO systems in time-selective channels using the first order Gauss-Markov Rayleigh fading channel model. We derive a closed form achievable rate for time-selective channels and provide a proof that the intracell interference effect in time-selective channels with constant amplitude (CA) pilots does not diminish in the asymptotic case. We show that there is a significant difference between the sum rate obtained using block-fading and time-selective models. We also show that the optimum training for block-fading may not be optimal for a time-selective channel, particularly for large numbers of antennas at the BS.

In Chapter 5, we investigate the performance of massive MIMO systems in time and frequency selective channels. A novel closed form achievable rate is derived for the channel model. We also compare the transmission performance of two different pilot sequences, a constant amplitude pilot and a zero padded pilot. The results show that in general, as the number of antennas increases, the optimum training block size and spatial multiplexing gain increase. Results also show that a zero padded pilot has a better sum rate performance than a constant amplitude pilot for a wide range of antenna numbers and time-frequency correlation values. This chapter also studies two different training optimization methods which are adaptive optimal training and average optimal training. Results show that both methods have a similar sum rate performance. In addition, we study the effect of increasing frequency reuse and pilot reuse factor. Despite their ability to reduce intercell interference, these methods can lower the sum rate due to inefficient use of time-frequency resources.

In Chapter 6, we provide overall conclusions for our work. This chapter also includes possible future research directions. Specifically, we can extend our research to include antennas spatial correlation and milimeter wave. We can also study the impact of cell radius and coding schemes on the transmission performance.

1.5 Publications

Journal:

1. Mahyiddin, W.A.W.M.; Martin, P.A.; Smith, P.J., "Performance of Synchronized and Unsynchronized Pilots in Finite Massive MIMO Systems," in *IEEE Transactions on Wireless Communications*, Early Access, July 2015.

2. Mahyiddin, W.A.W.M.; Martin, P.A.; Smith, P.J., "Massive MIMO Systems in Time-Selective Channel," in *IEEE Communications Letter*, Early Access, September 2015.

3. Mahyiddin, W.A.W.M.; Martin, P.A.; Smith, P.J., "Massive MIMO Systems in Time and Frequency Selective Channel," submitted to *IEEE Transactions on Communications*.

Conference:

1. Mahyiddin, W.A.W.M.; Martin, P.A.; Smith, P.J., "Pilot Contamination Reduction Using Time-Shifted Pilots in Finite Massive MIMO Systems," in *IEEE Vehicular Technology Conference (VTC Fall)*, pp.1-5, 14-17 Sept. 2014.

Chapter 2

Background model and assumptions

Our research relies on mathematical tools to analyze the physical layer of wireless communication systems. Hence, in this chapter, we provide background details on the models and assumptions that will be applied throughout this thesis. Information on wireless standards and parameters that are applied in this research are also provided in this chapter, such as the channel and path loss models, channel estimation, MIMO systems, multicell scenarios and orthogonal frequency-division multiplexing (OFDM) systems.

2.1 Channel model

A wireless channel is a propagation medium linking the transmitter and the receiver [11, 24]. It is a vital part of determining the performance of wireless communication systems in a given environment. A radio wave can be reflected or diffracted from physical obstructions between transmitter and receiver such as vehicles, trees and buildings. As a result, the same signal may arrive at the receiver from a number of different propagation paths. In this condition, the wireless channel undergoes multipath fading, as illustrated in Figure 2.1 for a single-input single-output (SISO) channel. Multipath fading can be represented using an impulse response between transmitter and receiver [32]. Due to the multipath fading, a wireless channel may experience variation across time, frequency and spatial domains. Such channel variation is commonly categorized into two types,

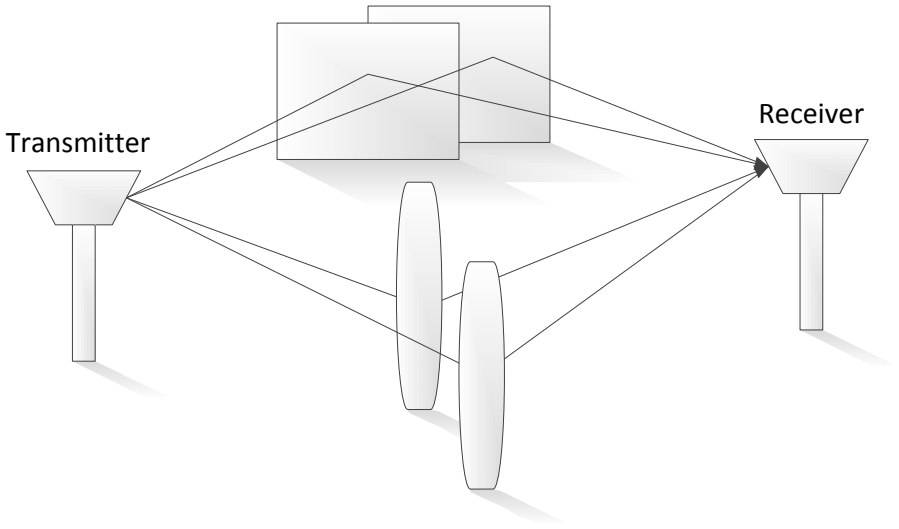

FIGURE 2.1: Multipath signal propagation from transmitter to receiver in a SISO channel.

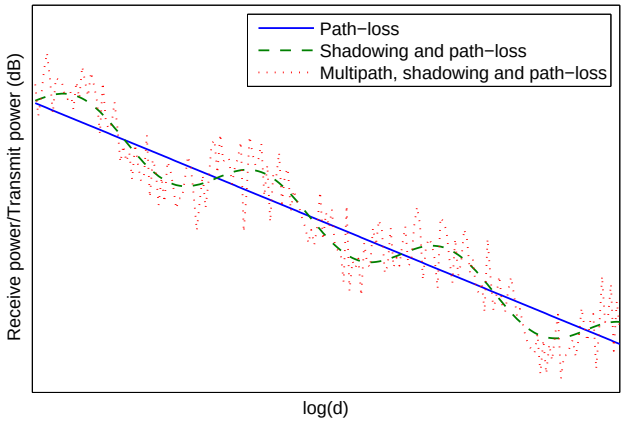

FIGURE 2.2: Combined path-loss, shadowing (large scale fading) and multipath (small scale fading) versus distance (d) [32].

which are large scale fading and small scale fading [24]. An example of the combination of large and small scale fading is shown in Figure 2.2. The general form of channel model which combines these two types of channel can be expressed as [4]

$$g = \sqrt{\beta}h, \tag{2.1}$$

where β is due to large scale fading and h is the small scale fading value.

2.1.1 Large scale fading

Large scale fading is a signal power variation due to distance decay and shadowing [24]. Distance decay occurs when a mobile moves further from the BS which results in the power at the receiver experiencing a geometric decay. Shadowing is when the signal is obstructed by large objects such as buildings. Large scale fading is usually assumed to be time, frequency and space independent within a small scale range (for example, in the range of ten by ten wavelengths [11]). The value of β can be determined using various path loss models [11]. We generate β using the simple model defined by [5]

$$\beta = z\delta^{-\gamma}, \tag{2.2}$$

where δ is the distance between the transmitter and receiver normalized to the inner radius of a cell [6] and γ is the exponential decay factor. z is a shadowing variable which has a log normal distribution, with $z = 10^{\frac{x}{10}}$, where $x \sim \mathcal{N}(0, \sigma_{\mathrm{s}}^2)$.

2.1.2 Small scale fading

Small scale fading occurs when a radio signal from a transmitter arrives at a receiver from multiple paths which results in electromagnetic wave superposition (constructive and destructive interference) [24]. The small scale fading is commonly modeled as Rayleigh fading [23]. This model assumes that there is no dominant line of sight between transmitter and receiver. The surrounding environment acts as multiple scatterers which enable the signal to arrive at the receiver using multiple paths. This condition can occur in a dense urban environment where the signal can be reflected, diffracted and attenuated by many surrounding objects. A Rayleigh fading channel can be expressed as

$$h = \frac{1}{\sqrt{2}}\left(x + iy\right), \tag{2.3}$$

where x and y are independent Gaussian variables with zero mean and variance one. The complex valued h represents the amplitude and phase shift variation experienced by a radio signal when propagating through multiple channel paths. Due to the fading process, the value of h can vary across time, frequency and spatial domains. Time varying small scale fading happens due to the Doppler effect from a moving UE. Frequency variation occurs due to signals arriving at the receiver with different delays due to the

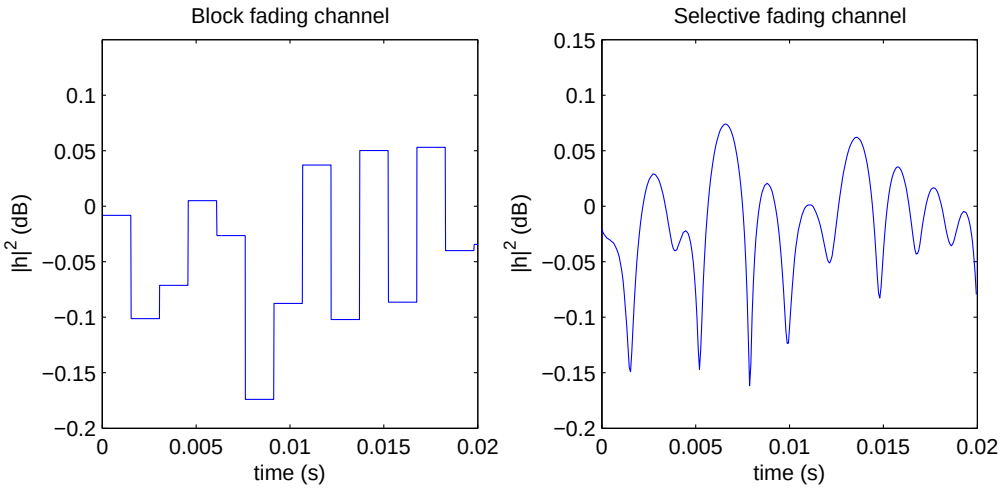

FIGURE 2.3: Time variation of small-scale fading power for a block fading channel and a selective fading channel with 150 km/h UE speed.

multipath channel. Spatial diversity can be realized by using several antenna elements separated in space, which is further discussed in Section 2.2. The variation of h across these dimensions depends on the channel model and parameters.

2.1.3 Block fading channel and selective fading channel

Channel variation across the time and frequency domains can be represented by a continuous fading model such as Jakes' model [28]. In order to simplify analysis, the continuous model is often discretized to a block fading model, which is useful for many time-division multiple access (TDMA), frequency-hopping, or block-interleaved systems [2]. This model approximates continuous fading by assuming that the channel stays constant within a coherence block, but changes its value between blocks. Such an assumption is practical for slow fading and sufficiently narrow channels. A coherence block is the range in time and frequency domains where the channel is approximately constant. The correlation between channel values within the coherence block is defined to be above a certain limit [29]. Due to its simplicity, the block fading model is widely used in massive MIMO studies [4–7]. A more accurate way to represent the channel variation is using the selective fading model. In this model, the channel can vary within the coherence block. Figure 2.3 shows examples of time variation of small scale fading power for a block fading channel and a selective fading channel for a UE with 150 km/h speed. The block fading channel is generated based on 50% coherence time [29], which means that the channel correlation at two ends of the coherence time is 50%. The

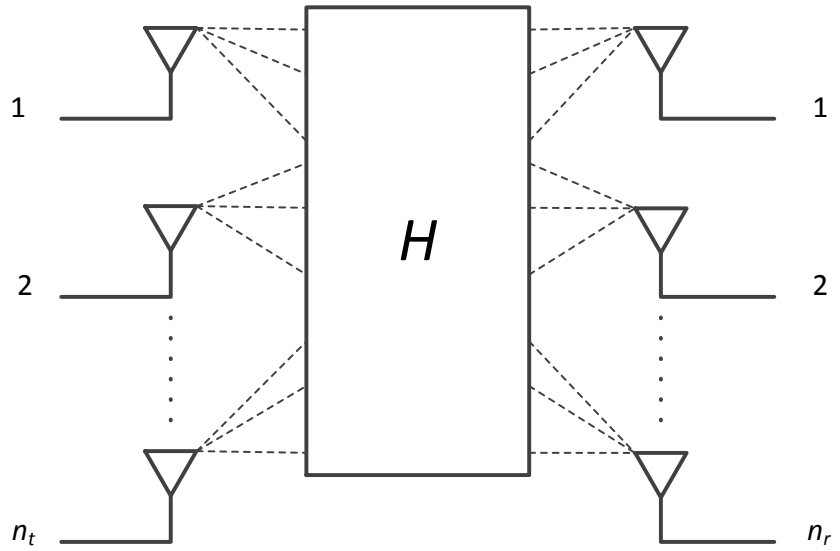

FIGURE 2.4: MIMO transmission through the channel $\boldsymbol{H}$.

coherence time is determined by maximum Doppler shift spread f_{d}, as follows [29]

$$T_{\mathrm{C}} \approx \frac{0.423}{f_{\mathrm{d}}}. \tag{2.4}$$

The maximum Doppler shift depends on the UE speed, as follows

$$f_{\mathrm{d}} = \frac{f_{\mathrm{c}} s}{c}, \tag{2.5}$$

where f_{c} is the carrier frequency, s is the UE speed and c is the speed of light. For the selective fading, the channel is generated based on Jakes model [28].

In this thesis, we are using both types of channel models. In Chapter 3, we use a block fading model in the work on time-shifted pilots. In Chapter 4, we use both block and time-selective fading models and provide comparison between these two models. In Chapter 5, we use channel selectivity in time and frequency domains to study the performance of two types of pilot sequence.

2.2 MIMO

An example of MIMO transmission is given in Figure 2.4. If there are n_t transmit

antennas and n_r receive antennas, then the signals at the receiver can be written as

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{v}, \tag{2.6}$$

where $\boldsymbol{y}$ is the received $n_r \times 1$ signal vector, $\boldsymbol{x}$ is the transmitted $n_t \times 1$ signal vector, $\boldsymbol{H}$ is $n_r \times n_t$ channel matrix and $\boldsymbol{v}$ is the $n_r \times 1$ noise vector at the receiver. We assume each element in $\boldsymbol{H}$, $\boldsymbol{x}$ and $\boldsymbol{v}$ are independent and identically distributed (iid) complex Gaussian.

To mathematically analyze the performance of this wireless system, researchers often use channel capacity, which is an information-theoretical approach that was originally developed by Claude E. Shannon for the additive white Gaussian noise (AWGN) channel [30]. Channel capacity is a theoretical upper limit on data rate that can be transmitted without error. The benefit of using channel capacity is that it enables us to obtain performance analysis of a wireless system without having to go into the details of transceiver design, or the coding and modulation schemes employed. This is done by assuming that the code rate is capable of being adapted to its upper bound rate. Research shows that near channel capacity performance is achievable using turbo coding schemes [31]. Channel capacity was originally applied to study a single link system [30] and has been extended to analyze MIMO [18]. If we assume channel state information (CSI) is available at the receiver, but not at the transmitter. It can be shown that the instantaneous channel capacity of the transmission in (2.6) can be expressed as

$$C = \log_2 \det\left(\boldsymbol{I}_{n_r} + \frac{\rho \boldsymbol{H}\boldsymbol{H}^H}{n_t}\right) \text{ bits/s/Hz}, \tag{2.7}$$

where ρ is defined as average SNR [18]. Following the derivation in [1], (2.7) is upper bounded by $\min(n_t, n_r) \cdot \log\left(1 + \rho \frac{\max(n_t, n_r)}{n_t}\right)$. This means the transmission capacity scales linearly with $\min(n_t, n_r)$. This proves that multiple antennas can increase the wireless transmission capacity.

2.2.1 Antenna spacing

MIMO systems depend on spatial diversity at the transmitters and the receivers to improve the spectral efficiency. Ideally, the spatial signature between different antennas is uncorrelated. However, practically, the performance of closely spaced antennas will

be degraded by spatial correlation. A well known result for spatial correlation is given by [28]

$$\rho = J_0 \left(2\pi d/\lambda\right), \tag{2.8}$$

where d is distance and λ is carrier wavelength. This is based on Jakes' model where there are an infinite number of scatterers circularly surrounding the antennas. The Bessel function property suggests that to achieve zero antenna correlation, the minimum spacing distance between antennas must be set around 0.4 λ. Since appropriate antenna spacing is required in order to ensure the correlation between antennas can be reduced, it is important to make sure that the number of antennas is not so large that the physical size of the antenna system is impractical. In this thesis, we set the number of antennas at the BS between 50 and 500 while each UE has one antenna. If the carrier frequency is set to 2 GHz, then the size of a 50/500 element BS array distributed in a square form with 0.4 λ equal spacing will be approximately 0.5/1.5 meters in width, which is quite large. The physical size can further be reduced by distributing the antennas in a cylindrical shape such as in [27]. We can also reduce the antenna spacing by increasing the carrier frequency. However, radio signals with extremely high frequency may encounter various issues such as high attenuation [34]. We assume that all the antennas are within a coherence length. The coherence length can be approximated as $L = c/B$, where c is speed of light and B is the bandwidth. For a 20 MHz bandwidth, the coherence length is 15 meters. Since the wavelength for a 2 GHz carrier frequency is 0.15 meters, the coherence length is much larger than the wavelength and this allows massive MIMO to be implemented.

2.2.2 MU-MIMO

Conventional point-to-point MIMO only uses one user for the same time and frequency. The specific type of MIMO that we use is a multiuser MIMO (MU-MIMO), which enables spatial multiplexing transmission with more than one UE [35]. This method can also be termed as space-division multiple access (SDMA). In this thesis, we define the number of spatial multiplexed UEs as the number of UEs that use the same time-frequency resources in a cell to perform parallel data transmission, where each UE has one antenna. Keep in mind that the number of spatial multiplexed UEs is not necessarily the same as the total number of active UEs in a cell. This is because there can be different groups

of UEs that use different time-frequency resources (such as in the case of orthogonal frequency-division multiple access (OFDMA)), which means that the total number of UEs can be higher than the number of spatial multiplexed UEs.

2.3 Massive MIMO

Massive MIMO, which is also known as very large MIMO or large scale antenna system, can use up to hundreds of antennas in the transmission. The rationale behind this idea is that as the number of BS antennas increases, the spatial signature between different UEs becomes less correlated [1]. In the case of Rayleigh fading [23], the channel between different users of the same cell will become asymptotically orthogonal when the number of BS antennas approaches infinity. This is due to the asymptotic of random matrix theory starting to take place when the number of antennas grows large. The channel orthogonality enables the separation of multiple parallel streams of transmitted data with minimal error.

Referring to (2.6), each transmit antenna transmits an independent stream of data and the receiver task is to recover all the transmitted data. The data from each transmit antenna can be recovered at the receiver using a linear combiner which correlates a vector with the received signal, $\boldsymbol{w}^H\boldsymbol{y}$ [36]. To recover transmitted data from n-th transmit antenna (n-th element in vector $\boldsymbol{x}$ or x_n), we can use matched filter (MF) equalization, which is $\frac{\boldsymbol{h}_n^H}{n_r}\boldsymbol{y}$. $\boldsymbol{h}_n$ is the n-th column vector of $\boldsymbol{H}$ and we assume the channel is known at the receiver. Using (2.6), we can expand this operation as

$$\frac{\boldsymbol{h}_n^H\boldsymbol{y}}{n_r} = \underbrace{\frac{\boldsymbol{h}_n^H\boldsymbol{h}_n x_n}{n_r}}_{\text{Desired signal}} + \underbrace{\sum_{i\neq n}^{n_t}\frac{\boldsymbol{h}_n^H\boldsymbol{h}_i x_i}{n_r} + \frac{\boldsymbol{h}_n^H\boldsymbol{v}}{n_r}}_{\text{Interference and noise}}. \tag{2.9}$$

$\boldsymbol{h}_i$ is the channel vector between the i-th transmitter and the receiver. Assuming no correlation between antennas, it can be shown that as n_r approaches infinity, then $\frac{\boldsymbol{h}_n^H\boldsymbol{h}_i}{n_r} \to 0$ for $i \neq n$. This also applies to the noise term $\boldsymbol{v}$. As the interferences vanish due to the law of large numbers, the channel has achieved a state of favorable propagation [1]. Consequently, the MF equalization in (2.9) becomes

$$\lim_{n_r\to\infty}\frac{\boldsymbol{h}_n^H\boldsymbol{y}}{n_r} = x_n. \tag{2.10}$$

Obtaining x_n means that the transmitted data can be recovered without any error in the asymptotic case. This proves that as the number of antennas at the receiver becomes very large, the noise and interference diminish.

The potential of massive MIMO is also supported by channel capacity analysis. In [1] it was shown that capacity bounds for MIMO systems can be simplified to

$$\log_2 (1 + \rho n_r) \leq C \leq \min (n_t, n_r) \cdot \log_2 \left(1 + \frac{\rho \max (n_t, n_r)}{n_t} \right). \tag{2.11}$$

The lower bound on capacity occurs when the channel matrix has rank 1 (such as in the case of line of sight transmission) while the upper bound on capacity occurs when the channel matrix is full rank and experiences favorable propagation. In [1], it is further shown that when the number of receive antennas becomes much greater than the number of transmit antennas, the instantaneous capacity when there is perfect knowledge of the channel matrix, $\boldsymbol{H}$, at the transmitter can be represented as

$$\begin{aligned} C_{n_r \gg n_t} &= \log_2 \det \left(\boldsymbol{I}_{n_t} + \frac{\rho}{n_t} \boldsymbol{H}^H \boldsymbol{H} \right) \\ &\approx n_t \cdot \log_2 \left(1 + \frac{\rho n_r}{n_t} \right). \end{aligned} \tag{2.12}$$

This matches the upper bound capacity in (2.11), which shows that having an excess of receive antennas is a desirable condition.

2.4 Channel estimation

As discussed in Section 1.1, spatial multiplexing transmission requires the use of CSI. This thesis uses a linear channel estimator where the received pilot signal is used to obtain the CSI estimate. The pilot is transmitted periodically in order to update the CSI estimate. If there are n_t transmit antennas, n_r receive antennas and a pilot length of N, then the received signal during pilot transmission can be expressed as

$$\boldsymbol{Y} = \boldsymbol{\Psi} \boldsymbol{H} + \boldsymbol{V}, \tag{2.13}$$

where Y is the $N \times n_r$ received pilot matrix, $\boldsymbol{\Psi}$ is the $N \times n_t$ concatenated pilot matrix from n_t transmitters, $\boldsymbol{H}$ is the $n_t \times n_r$ channel matrix and $\boldsymbol{V}$ is an $N \times n_r$ noise matrix

at the receiver. A common way to obtain the channel estimate from the received pilot is to use a least squared (LS) estimate, as follows

$$\hat{\boldsymbol{H}}_{\mathrm{LS}} = \left(\boldsymbol{\Psi}^H \boldsymbol{\Psi}\right)^{-1} \boldsymbol{\Psi}^H \boldsymbol{Y}, \tag{2.14}$$

where the value of $\boldsymbol{\Psi}$ is set so that $\boldsymbol{\Psi}^H \boldsymbol{\Psi} = \boldsymbol{I}_{n_r}$. This can be achieved if pilot sequences from each transmitter are orthogonal to each other and $N \geq n_t$. In the case of massive MIMO systems with a large number of BS antennas, but a small number of total UE antennas, estimation of the uplink channel will be straightforward because the length of uplink pilot should be greater than or equal to the number of total UE antennas. For downlink transmission, the BS will be the transmitter, which means that the length of the downlink pilot needs to be greater than or equal to the number of antennas at the BS to estimate the downlink channel. Since there can be hundreds of BS antennas in massive MIMO systems, this means that a large pilot overhead is needed to estimate the downlink channel. This is a problem since the large overhead means there will be fewer time-frequency resources than can be used to transmit the data within the coherence block. Even if we assume that the UEs manage to estimate the CSI, large information feedback would be needed so that the BS could perform beamforming. This consumes even more time-frequency resources.

2.4.1 Channel reciprocity

The problem of acquiring downlink CSI can be solved by using channel reciprocity, which is a condition when the uplink channel is the same as the downlink channel. If a radio signal is transmitted along a certain channel path, its reverse direction will also follow the same path if it has the same frequency, thus creating the reciprocity condition [43]. This means that channel reciprocity can occur in time division duplex systems (TDD), where the uplink and the downlink signal use the same frequency. In the case of frequency division duplex (FDD), downlink and uplink use different frequency bands which means that channel reciprocity cannot be achieved in FDD. Using channel reciprocity, we can make use of the estimated CSI from the uplink pilot to form a precoding matrix to transmit the downlink data. Therefore, the overhead pilot length will remain the same as the number of spatial multiplexed UEs. In addition, channel reciprocity makes the CSI feedback unnecessary.

An issue that arises in the channel reciprocity assumption is that the signal distortion that originates from the front end hardware of the antennas will cause a channel mismatch between uplink and downlink transmission [44, 45]. In [33], they propose to use the indirect calibration method to achieve reciprocity in massive antenna systems. Field tests in [33] show that this method is a practical solution for channel mismatch. Hence, channel reciprocity is an applicable assumption in massive MIMO.

2.4.2 LMMSE channel estimation

In this thesis, we use linear minimum mean squared error (LMMSE) channel estimation. LMMSE derives its concept from the minimum mean squared error (MMSE) approach which aims to minimize the channel estimation error [46]. Since LMMSE is a linear estimator, it is considered as a sub-optimal estimator compared to the MMSE estimator. However, LMMSE provides a practical solution for channel estimation due to its low computational complexity. In addition, LMMSE has near-optimal performance in massive MIMO [41]. LMMSE and MMSE, both provide a channel estimate that is uncorrelated with the channel estimation error. This property is important to obtain the lower bound ergodic rate [38]. If the initial channel information is obtained from the LS channel estimate in (2.14), then the LMMSE estimate can be defined as $\hat{\boldsymbol{H}}_{\mathrm{LMMSE}} = \boldsymbol{W}\hat{\boldsymbol{H}}_{\mathrm{LS}}$, where $\boldsymbol{W}$ is a linear multiplier that minimizes the error. The channel estimation error is $\boldsymbol{E} = \boldsymbol{H} - \boldsymbol{W}\hat{\boldsymbol{H}}_{\mathrm{LS}}$. This is minimized when it becomes orthogonal to $\hat{\boldsymbol{H}}_{\mathrm{LS}}$, meaning $\mathbb{E}\left[\boldsymbol{E}\hat{\boldsymbol{H}}_{\mathrm{LS}}^{H}\right] = 0$. This equality can be expanded as follows [47]

$$\begin{aligned}
\mathbb{E}\left[\boldsymbol{E}\hat{\boldsymbol{H}}_{\mathrm{LS}}^{H}\right] &= 0 \\
\mathbb{E}\left[\left(\boldsymbol{H} - \boldsymbol{W}\hat{\boldsymbol{H}}_{\mathrm{LS}}\right)\hat{\boldsymbol{H}}_{\mathrm{LS}}^{H}\right] &= 0 \\
\mathbb{E}\left[\boldsymbol{H}\hat{\boldsymbol{H}}_{\mathrm{LS}}^{H}\right] - \boldsymbol{W}\mathbb{E}\left[\hat{\boldsymbol{H}}_{\mathrm{LS}}\hat{\boldsymbol{H}}_{\mathrm{LS}}^{H}\right] &= 0 \\
\boldsymbol{R}_{\boldsymbol{H}\hat{\boldsymbol{H}}_{\mathrm{LS}}} - \boldsymbol{W}\boldsymbol{R}_{\hat{\boldsymbol{H}}_{\mathrm{LS}}\hat{\boldsymbol{H}}_{\mathrm{LS}}} &= 0.
\end{aligned} \tag{2.15}$$

Using (2.15) and $\hat{\boldsymbol{H}}_{\mathrm{LMMSE}} = \boldsymbol{W}\hat{\boldsymbol{H}}_{\mathrm{LS}}$, we can obtain the LMMSE channel estimate in terms of $\hat{\boldsymbol{H}}_{\mathrm{LS}}$ as [47]

$$\hat{\boldsymbol{H}}_{\mathrm{LMMSE}} = \boldsymbol{R}_{\boldsymbol{H}\hat{\boldsymbol{H}}_{\mathrm{LS}}}\left(\boldsymbol{R}_{\hat{\boldsymbol{H}}_{\mathrm{LS}}\hat{\boldsymbol{H}}_{\mathrm{LS}}}\right)^{-1}\hat{\boldsymbol{H}}_{\mathrm{LS}}. \tag{2.16}$$

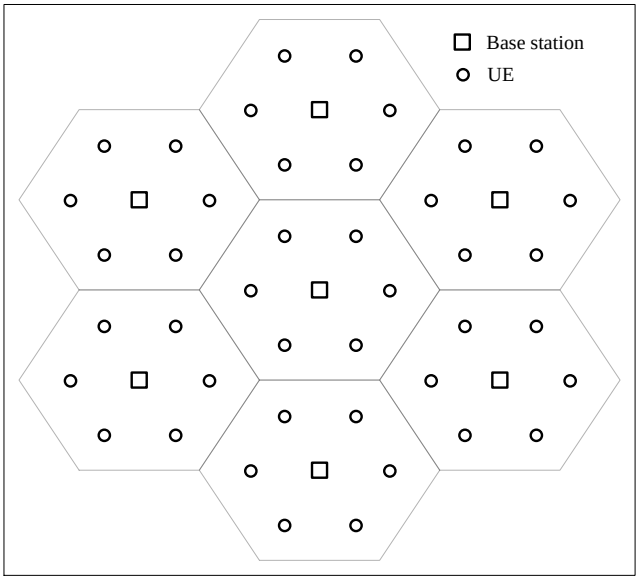

FIGURE 2.5: Hexagonal multicell structure.

This result will be used throughout the thesis. Note that in the case of a block fading channels, the LMMSE estimate is equal to the MMSE estimate when the signal used to estimate the channel at the receiver (i.e. the received pilot) has a Gaussian distribution [5, 8].

2.5 Multicell system

We consider a multicell structure to analyze the impact of intercell interference on massive MIMO systems. We use hexagonal cells with each cell surrounded by 6 adjacent cells, as shown in Figure 2.5. Each cell is set to have the same number of UEs and same number of BS antennas. The UEs can be located at uniform random locations within the cell [3] or at equal distance from the BS [6]. Each UE will belong to the cell which has the biggest link gain (large scale fading) between the UE and the cell BS. This is because UE cell allocation is determined from the power level between BS and UE rather than the physical distance between them [11]. We use frequency reuse 1 (all cells use the same bandwidth), which means there can be interference from all nearby cells. We assume that the interference from outside the 7 cells has a negligible impact on the middle cell in Figure 2.5. This is because the performance degradation mainly

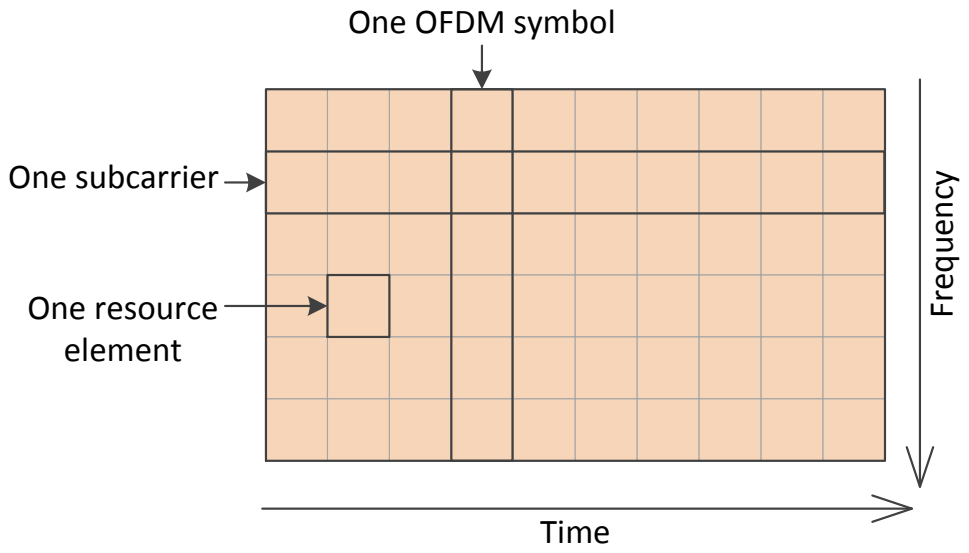

FIGURE 2.6: OFDM transmission structure in time and frequency domain.

comes from interfering signals from the cell edge UEs. Since we use a pathloss exponent decay of 3.8, the power from cell edge UEs outside the 7 cells is only 4% of the power of cell edge UEs from inside the 7 cells. This is calculated based on the radius of the 7 cells being 2 times the radius of one cell. This means that the power ratio between the two distances in the pathloss model without shadowing is $(2/1)^{-3.8} \approx 0.04$. Hence, the interference from outside the 7 cells is negligible compared to the interference from adjacent cells.

2.6 OFDM system

In OFDM, the carrier bandwidth is divided into multiple narrow band subcarriers for parallel streams of data transmission [11, 32, 48]. The narrow band allows the transmission to handle frequency selectivity more effectively. OFDM achieves high rate data by combining large numbers of subcarriers. For example, the number of subcarriers in an LTE system can range from 128 to 2048. To avoid inter carrier interference, each subcarrier is set to be orthogonal to each other. The orthogonality between subcarriers is ensured by setting the subcarrier frequency to be a multiple of the lowest subcarrier frequency. Using the efficient FFT algorithm, the signals from the subcarriers are summed to produce an OFDM symbol. Our research will analyze the transmission from a time-frequency point of view, which is represented in Figure 2.6. Signals from each

resource element are transmitted through a time-frequency channel model before arriving at the receiver. Each resource element is occupied with either pilot or data symbols. Pilot symbols are a set of orthogonal sequences which are used to estimate the channel. Data symbols are modulated signal which are represented as complex numbers. We assumed that the data has a Gaussian distribution, which is an assumption needed to obtain a lower bound on ergodic rate [8]. The arrangement and values of pilot and data in OFDM depends on the training model to be used, which will be discussed in detail in the following chapters. Note that in a standard wireless system, a cyclic prefix or guard band is used to handle multipath delay more effectively. The cyclic prefix is a temporal period symbol that is added to each OFDM symbol to avoid symbol overlapping from multipath delay. For simplicity, we exclude the impact of resource loss from using the cyclic prefix. This assumption will not affect the performance comparison between different transmission methods that use the same cyclic prefix length.

2.7 Summary

In this chapter, we provided background details on the models and assumptions that will be applied throughout this thesis. Specifically, we discussed background information for the channel model, MIMO system, massive MIMO, multicell scenario and OFDM.

Chapter 3

Time-shifted pilots in massive MIMO

3.1 Introduction

As discussed in Chapter 1 and 2, pilot contamination is a major limiting factor in massive MIMO because the interference does not diminish even when the number of antennas becomes very large. In the case of a transmission that uses a linear channel estimator, pilot contamination is an issue that must be dealt with [27]. In [1, 4], the worst case pilot contamination occurs when neighboring cells send the same pilot sequence simultaneously. To avoid the problem of synchronized pilots between adjacent cells, [3] proposed a time-shifted pilot method, where some cells send downlink data while others transmit pilots. Although the method can significantly improve the transmission rate, the analysis in [3] is based on the assumption of an infinite number of antennas at the BS. Using the time-shifted method along with a certain cell group arrangement, all the interference from the nearest cells can be eliminated in the asymptotic case. As the number of antennas goes to infinity, [3] shows that the interference from different cell groups during data transmission becomes negligible and the rate performance can be overestimated.

In the case of a limited numbers of antennas at the BS, we cannot ignore the impact of the aforementioned interference on the transmission performance. Therefore, in this chapter, we study the effect of massive, but finite MIMO systems on the time-shifted performance.

We derive novel closed form expressions for downlink and uplink transmission rates using MF for two different time-shifted methods. The closed form achievable rate for a finite number of antennas using estimated channels has been analyzed in massive MIMO systems [5–8]. These achievable rate derivations are based on [38], where the lower bound rate is obtained by assuming the interference behaves like Gaussian noise. We use a similar concept to obtain a closed form achievable rate for the time-shifted pilots in [3] with a finite number of antennas. In [49] the time-shifted method from [3] is studied with a zero-forcing scheme.

This chapter also investigates another variant of the time-shifted method, where the uplink pilots overlap with uplink data from other cells instead of downlink data from other cells. The impact of uplink data interference on pilots was briefly analyzed in [5], and said to have a similar effect as pilot to pilot interference (time-synchronized pilot). Subsequent works on massive MIMO [52–54] also use similar assumptions as in [5] regarding the impact of uplink data interference on the pilots. The assumptions arising from these works are built on a relatively brief analysis and provide no detailed sum rate derivations and performance comparisons of the time-shifted methods. Hence, in this chapter, we provide a more complete analysis of the time-shifted methods by deriving novel closed form rate expressions for time-shifted methods with finite number of antennas. The rate expressions allow us to obtain significantly faster results compared to Monte Carlo simulation and provide insights into performance. Specifically, the results show that time-shifted pilots with downlink data overlap perform the best when there are a small number of UEs while time-shifted pilots with uplink data overlap has the advantage when there are high numbers of UEs. We also show that time-synchronized pilots are not necessarily the worst case scenario in terms of sum rate performance.

The rest of this chapter is arranged as follows. Section 3.2 provides the general system model. Section 3.3 describes the time-shifted pilot method with downlink data overlap and Section 3.4 describes the time-shifted pilot method with uplink data overlap. Both sections 3.3 and 3.4 investigate channel estimation, downlink transmission rate and uplink transmission rate. The time-synchronized method is discussed in Section 3.5. This is followed by numerical results in Section 3.6 and a summary in Section 3.7.

3.2 System model

We assume an OFDM system, where each subcarrier has a narrow bandwidth such that we can assume flat fading transmission. Each BS has M antennas and each UE has one antenna. The number of coherent subcarriers is N. If we limit the discrete length of the pilot in the time domain as τ, then the total length of pilot sequences in the time-frequency domain is $K = \tau N$ symbols. K is also the maximum number of orthogonal pilot sequences, which means there can be K UEs performing parallel transmission using the same time-frequency resource through spatial multiplexing [4]. Since K is proportional to N, this implies that coherence in the frequency domain can be used to achieve a higher spatial multiplexing gain.

We consider a block fading channel model with a frame duration of T symbols. Each frame includes uplink pilot, downlink data and uplink data transmission within a time-coherent channel block. We use a similar Rayleigh fading channel model as that discussed in Chapter 2, but extended to include cell and UE numbers. We define the channel vector between the k-th UE in cell i and the BS in cell l as $\boldsymbol{g}_{ilk} = \sqrt{\beta_{ilk}}\boldsymbol{h}_{ilk}$. $\boldsymbol{h}_{ilk}$ is a $1 \times M$ small-scale fading vector where $\boldsymbol{h}_{ilk} \sim \mathcal{CN}(0, \boldsymbol{I}_M)$ and β_{ilk} is a time-invariant large-scale fading coefficient that depends on the path-loss model [3]. We assume all UEs have the same average transmit power and all BSs have the same average transmit power. All UEs and BSs have the same average noise power.

Due to the large-scale of massive MIMO channel information, it is difficult for a single antenna UE to obtain instantaneous channel information when the coherence time is limited [41]. However, the BS can estimate the channel by using the received uplink pilot from the UE. The BS can use the estimated channel not only to equalize data received from the UE during uplink data transmission, but also to beamform data from BS to UE during downlink data transmission. This assumes channel reciprocity, where both uplink and downlink channels are the same. Field tests in [33] show this is achievable in massive MIMO using TDD. We use MF for precoding and equalization. We consider any signal that originates from a different cell is treated as interference.

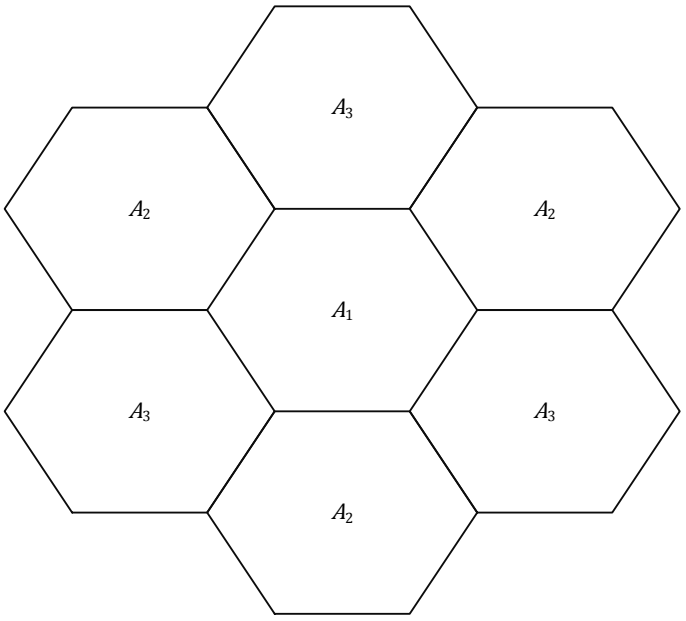

FIGURE 3.1: Cell group arrangement for 7 cell system.

3.3 Time-shifted pilot with downlink data overlap

In this section, we consider the time-shifted pilot arrangement of [3], where the cells are arranged according to groups and pilots overlap with downlink data. We refer to this as TShdown (and superscript TD in equations). Since [3] assumed an infinite number of antennas, the novelty of this section is the derivation of the closed form ergodic sum rate for a finite number of antennas. We consider 3 cell groups, denoted A_1, A_2 and A_3. Cells from different groups are often adjacent or close to each other, which means they can interfere with each other's signals. As shown in Figure 3.2, each group is set to have an uplink pilot interval that does not interfere with the pilot from other groups, but interferes with the downlink data from other groups. In this arrangement, pilot contamination from different cell groups can be avoided, but pilot contamination from the same cell group can still occur. An example group arrangement for a 7 cell system is shown in Figure 3.1. For group A_2 and A_3, the precoding vector is obtained using the channel estimate from the previous frame. Due to the block fading assumption, we assume the coherence block for group A_2 and A_3 is from the beginning of the pilot in the previous frame to the beginning of the pilot in the current frame. Note that the downlink data duration is twice as long as the pilot duration. However, it is still possible to have longer data duration, as long as there is no pilot overlap between different cell

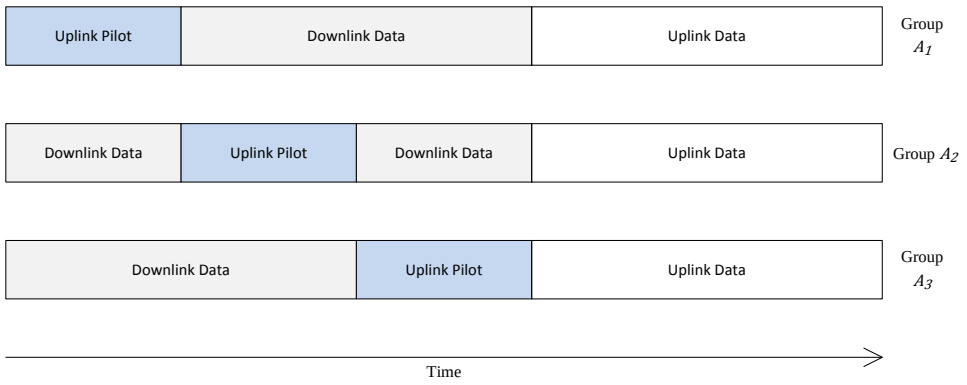

FIGURE 3.2: The arrangement of time-shifted pilots for TShdown.

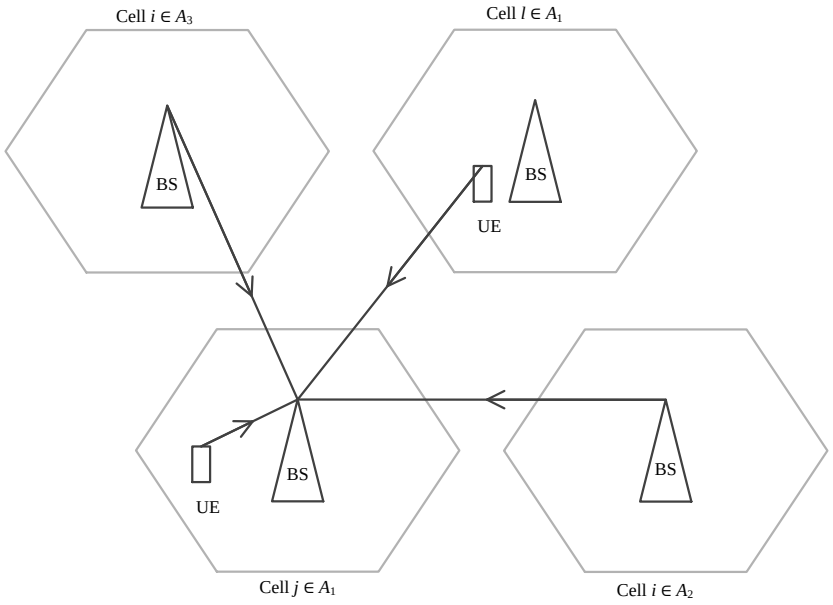

FIGURE 3.3: Received signal at cell j BS during group A_1 cells uplink pilot transmission phase for TShdown.

groups. This requires some modification on the sum rate equation, which can potentially be extended as future research.

3.3.1 Channel estimation

Channel estimation is required to define the precoding matrix during downlink transmission and the equalizer matrix for uplink transmission. To find a lower bound on transmission rate, we use LMMSE channel estimation [38]. The uplink pilot signal from the UE is used to estimate the channel. As shown in Figure 3.2, when uplink pilots are transmitted from group A_1 cells, the group A_2 and A_3 cells transmit downlink data. This transmission can be illustrated in Figure 3.3. If cell j belongs to group A_1, then

the received signal at BS cell j during this phase can be represented as

$$\boldsymbol{Y}_j = \underbrace{\sum_{i\in A_1}\sum_{k=1}^{K}\sqrt{p_{\mathrm{r}}\beta_{ijk}}\boldsymbol{\psi}_k\boldsymbol{h}_{ijk}}_{\text{Uplink pilot from UEs in group } A_1} + \underbrace{\sum_{i\notin A_1}\sum_{k=1}^{K}\sqrt{p_{\mathrm{f}}\delta_{ji}}\boldsymbol{q}_{ik}\boldsymbol{a}_{ik}^{T}\boldsymbol{G}_{ji}}_{\text{Downlink data from BSs in groups } A_2 \text{ and } A_3} + \boldsymbol{Z}_j, \quad (3.1)$$

where

- $\boldsymbol{Y}_j$ is the $K \times M$ received signal matrix at BS cell j.
- p_{r} is the average UE transmit pilot power.
- $\boldsymbol{\psi}_k$ is the $K \times 1$ pilot sequence vector of UE k with $\|\boldsymbol{\psi}_k\|^2 = K$.
- p_{f} is the average BS downlink data power.
- δ_{ji} is the large-scale fading from BS cell i to BS cell j.
- $\boldsymbol{q}_{ik}$ is the $K \times 1$ downlink data vector from BS cell i to UE k with $\boldsymbol{q}_{ik} \sim \mathcal{CN}(0, \boldsymbol{I}_K)$.
- $\boldsymbol{a}_{ik}$ is the $M \times 1$ precoding vector from BS in cell i for UE k in cell i.
- $\boldsymbol{G}_{ji}$ is the $M \times M$ small-scale fading matrix from BS cell i to BS cell j, where $\mathrm{vec}(\boldsymbol{G}_{ji}) \sim \mathcal{CN}(0, \boldsymbol{I}_{MM})$.
- $\boldsymbol{Z}_j$ is the $K \times M$ noise matrix at cell j BS, where $\mathrm{vec}\,(\boldsymbol{Z}_j) \sim \mathcal{CN}(0, \boldsymbol{I}_{KM})$.

Next, we derive the LMMSE channel estimate. First, we correlate the received signal with the pilot. k' represents UEs that have the same pilot sequence. The pilot sequence of the k'-th UE is correlated with the received pilot signal (3.1) as

$$\check{\boldsymbol{h}}_{jk'} = \boldsymbol{\psi}_{k'}^{H}\boldsymbol{Y}_j. \quad (3.2)$$

Any two distinct pilot sequences are orthogonal to each other, meaning $\boldsymbol{\psi}_{k'}^{H}\boldsymbol{\psi}_k = 0$ for $k \neq k'$, while identical sequences are completely correlated, $\boldsymbol{\psi}_{k'}^{H}\boldsymbol{\psi}_{k'} = K$. As a result, (3.2) becomes

$$\check{\boldsymbol{h}}_{jk'} = \sum_{i\in A_1}\sqrt{p_{\mathrm{r}}\beta_{ijk'}}K\boldsymbol{h}_{ijk'} + \sum_{i\notin A_1}\sum_{k=1}^{K}\sqrt{p_{\mathrm{f}}\delta_{ji}}\boldsymbol{\psi}_{k'}^{H}\boldsymbol{q}_{ik}\boldsymbol{a}_{ik}^{T}\boldsymbol{G}_{ji} + \boldsymbol{\psi}_{k'}^{H}\boldsymbol{Z}_j. \quad (3.3)$$

To analyze the impact of pilot contamination, we obtain channel estimates between UEs and BSs in all same group cells. Similar to [6], we use the correlation between the received signal and pilot in (3.3) to obtain LMMSE channel estimation.

Theorem 3.1. If l and j are in the same cell group, then LMMSE estimation of the channel between UE k' in cell l and BS in cell j is given by

$$\hat{\boldsymbol{h}}_{ljk'} = \frac{\sqrt{p_{\mathrm{r}}\beta_{ljk'}}}{\alpha_{jk'}^{\mathrm{TD}}}\check{\boldsymbol{h}}_{jk'}, \tag{3.4}$$

where

$$\alpha_{jk'}^{\mathrm{TD}} = \sum_{i\in A_1} p_{\mathrm{r}}\beta_{ijk'}K + \sum_{i\notin A_1}\sum_{k=1}^{K}\frac{p_{\mathrm{f}}\delta_{ji}}{K}\left(1+\frac{p_{\mathrm{f}}\delta_{ji}(1-\frac{1}{M})}{\phi_{ik}}\right)+1,$$
$$\phi_{ik} = \sum_{p\in A_\gamma} p_{\mathrm{r}}\beta_{pik}K + \sum_{p\notin A_\gamma} p_{\mathrm{f}}\delta_{ip} + 1.$$

Proof. The LMMSE estimate is given by [46]

$$\hat{\boldsymbol{h}}_{ljk'}^{T} = \boldsymbol{R}_{\boldsymbol{h}_{ljk'}\check{\boldsymbol{h}}_{jk'}}\left(\boldsymbol{R}_{\check{\boldsymbol{h}}_{jk'}\check{\boldsymbol{h}}_{jk'}}\right)^{-1}\check{\boldsymbol{h}}_{jk'}^{T}, \tag{3.5}$$

where $\boldsymbol{R}_{\check{\boldsymbol{h}}_{jk'}\check{\boldsymbol{h}}_{jk'}}$ is the covariance of $\check{\boldsymbol{h}}_{jk'}$ while $\boldsymbol{R}_{\boldsymbol{h}_{ljk'}\check{\boldsymbol{h}}_{jk'}}$ is the cross-covariance of $\boldsymbol{h}_{ljk'}$ and $\check{\boldsymbol{h}}_{jk'}$.

The covariance of (3.3) can be written as

$$\boldsymbol{R}_{\check{\boldsymbol{h}}_{jk'}\check{\boldsymbol{h}}_{jk'}} = \sum_{i\in A_1} p_{\mathrm{r}}\beta_{ijk'}K^2\mathbb{E}\left[\boldsymbol{h}_{ijk'}^{H}\boldsymbol{h}_{ijk'}\right] + \sum_{i\notin A_1}\sum_{k=1}^{K} p_{\mathrm{f}}\delta_{ji}\mathbb{E}\left[\boldsymbol{G}_{ji}^{H}\boldsymbol{a}_{ik}^{*}\boldsymbol{q}_{ik}^{H}\boldsymbol{\psi}_{k'}\boldsymbol{\psi}_{k'}^{H}\boldsymbol{q}_{ik}\boldsymbol{a}_{ik}^{T}\boldsymbol{G}_{ji}\right]$$
$$+\mathbb{E}\left[\boldsymbol{Z}_{j}^{H}\boldsymbol{\psi}_{k'}\boldsymbol{\psi}_{k'}^{H}\boldsymbol{Z}_{j}\right]. \tag{3.6}$$

If we assume that every random variable in (3.6) is uncorrelated with each other, then we obtain following solution for (3.6)

$$\boldsymbol{R}_{\check{\boldsymbol{h}}_{jk'}\check{\boldsymbol{h}}_{jk'}} = \left(\sum_{i\in A_1} p_{\mathrm{r}}\beta_{ijk'}K^2 + \sum_{i\notin A_1} p_{\mathrm{f}}\delta_{ji}K + K\right)\boldsymbol{I}_M$$
$$= \phi_{jk'}K\boldsymbol{I}_M, \tag{3.7}$$

where $\phi_{jk'} = \sum_{i \in A_1} p_{\mathrm{r}} \beta_{ijk'} K + \sum_{i \notin A_1} p_{\mathrm{f}} \delta_{ji} + 1$. This expression is considered as an approximation because in reality, there will be correlation between $\boldsymbol{G}_{ji}$ and $\boldsymbol{a}_{ik}$ because $\boldsymbol{a}_{ik}$ is a function of $\boldsymbol{G}_{ji}$. To see this more clear, we use (3.3) and (3.19) to obtain the precoding vector at different cell groups ($i \notin A_1$) as

$$\boldsymbol{a}_{ik} = \frac{\hat{\boldsymbol{h}}_{iik}^H}{\sqrt{K}\left\|\hat{\boldsymbol{h}}_{iik}^H\right\|} = \frac{\check{\boldsymbol{h}}_{ik}^H}{\sqrt{K}\left\|\check{\boldsymbol{h}}_{ik}^H\right\|}, \tag{3.8}$$

where $\check{\boldsymbol{h}}_{ik'}^H$ is the correlation between the pilot and received pilot signal at BS in cell $i \notin A_1$. Since $\check{\boldsymbol{h}}_{ik'}^H$ contains the variable $\boldsymbol{G}_{ij}$ and $\boldsymbol{G}_{ij}^T = \boldsymbol{G}_{ji}$ (channel reciprocity), this means that $\boldsymbol{a}_{ik}$ is partially correlated with $\boldsymbol{G}_{ji}$.

Lemma 3.2. Let

$$\boldsymbol{w} = \sqrt{a}\boldsymbol{x} + \sqrt{b}\frac{\boldsymbol{y}}{\|\boldsymbol{y}\|}\boldsymbol{Z}, \tag{3.9}$$

where $\boldsymbol{w}$, $\boldsymbol{x}$ and $\boldsymbol{y}$ are $1 \times M$ vectors and $\boldsymbol{Z}$ is an $M \times M$ matrix. $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{Z}$ are independent from each other and have elements which are i.i.d. $\mathcal{CN}(0,1)$. We can show that

$$\mathbb{E}\left[\boldsymbol{Z}\frac{\boldsymbol{w}^H}{\|\boldsymbol{w}\|}\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}\boldsymbol{Z}^H\right] = \left(1 + \frac{b(1-\frac{1}{M})}{a+b}\right)\boldsymbol{I}_M. \tag{3.10}$$

Proof. See Appendix A. □

To solve (3.6) using Lemma 3.2, first we simplify (3.6) as

$$\boldsymbol{R}_{\check{\boldsymbol{h}}_{jk'}\check{\boldsymbol{h}}_{jk'}} = \sum_{i \in A_1} p_{\mathrm{r}} \beta_{ijk'} K^2 \boldsymbol{I}_M + \sum_{i \notin A_1} \sum_{k=1}^{K} p_{\mathrm{f}} \delta_{ji} \mathbb{E}\left[\boldsymbol{G}_{ji}^H \frac{\check{\boldsymbol{h}}_{ik}^T}{\left\|\check{\boldsymbol{h}}_{ik}\right\|} \frac{\check{\boldsymbol{h}}_{ik}^*}{\left\|\check{\boldsymbol{h}}_{ik}\right\|} \boldsymbol{G}_{ji}\right] + K\boldsymbol{I}_M. \tag{3.11}$$

Note that $\check{\boldsymbol{h}}_{ik}$ has a similar expression to that in (3.3), but for cell $i \notin A_1$. Since $\check{\boldsymbol{h}}_{ik}$ and $\boldsymbol{G}_{ji}$ are partially correlated, the expectation term in (3.11) has a similar property to that in (3.10), where $\boldsymbol{w}$ can be replaced with $\check{\boldsymbol{h}}_{ik}$ and $\boldsymbol{Z}$ can be replaced with $\boldsymbol{G}_{ji}$. As a result, b in (3.10) is replaced with $p_{\mathrm{f}}\delta_{ji}K$ while $a+b$ is replaced with the norm of the covariance of $\check{\boldsymbol{h}}_{ik}$, or $\left\|\boldsymbol{R}_{\check{\boldsymbol{h}}_{ik}\check{\boldsymbol{h}}_{ik}}\right\|$ for $i \notin A_1$. Hence, (3.11) becomes

$$\boldsymbol{R}_{\check{\boldsymbol{h}}_{jk'}\check{\boldsymbol{h}}_{jk'}} = \sum_{i \in A_1} p_{\mathrm{r}} \beta_{ijk'} K^2 \boldsymbol{I}_M + \sum_{i \notin A_1} \sum_{k=1}^{K} p_{\mathrm{f}} \delta_{ji} \left(1 + \frac{p_{\mathrm{f}}\delta_{ji}K(1-\frac{1}{M})}{\left\|\boldsymbol{R}_{\check{\boldsymbol{h}}_{ik}\check{\boldsymbol{h}}_{ik}}\right\|}\right)\boldsymbol{I}_M + K\boldsymbol{I}_M \tag{3.12}$$

The term $\left\|\boldsymbol{R}_{\check{\boldsymbol{h}}_{ik}\check{\boldsymbol{h}}_{ik}}\right\|$ for $i \notin A_1$ in (3.12) is similar to $\left\|\boldsymbol{R}_{\check{\boldsymbol{h}}_{jk}\check{\boldsymbol{h}}_{jk}}\right\|$ for $j \in A_1$, which is the norm of (3.12). For example, if $i \in A_2$, then

$$\left\|\boldsymbol{R}_{\check{\boldsymbol{h}}_{ik}\check{\boldsymbol{h}}_{ik}}\right\| = \sum_{q\in A_2} p_{\mathrm{r}}\beta_{qik'}K^2 + \sum_{q\notin A_2}\sum_{v=1}^{K} p_{\mathrm{f}}\delta_{iq}\left(1+\frac{p_{\mathrm{f}}\delta_{iq}K(1-\frac{1}{M})}{\left\|\boldsymbol{R}_{\check{\boldsymbol{h}}_{qv}\check{\boldsymbol{h}}_{qv}}\right\|}\right) + K. \tag{3.13}$$

However, this also means that we also need to solve $\left\|\boldsymbol{R}_{\check{\boldsymbol{h}}_{qv}\check{\boldsymbol{h}}_{qv}}\right\|$ for $q \notin A_2$ in (3.13). If we continue to solve this term, there will be another similar term within this term. As a result, there will be an endless sequence of functions within functions. However, deriving the sequence of functions within functions has diminishing impact on the calculation accuracy of (3.12) because the term $\frac{p_{\mathrm{f}}\delta_{ji}K(1-\frac{1}{M})}{\left\|\boldsymbol{R}_{\check{\boldsymbol{h}}_{ik}\check{\boldsymbol{h}}_{ik}}\right\|}$ is less than 1. For example, using transmission setting from [6] with 100 BS antennas and 3 UEs, we get $\frac{p_{\mathrm{f}}\delta_{ji}K(1-\frac{1}{M})}{\left\|\boldsymbol{R}_{\check{\boldsymbol{h}}_{ik}\check{\boldsymbol{h}}_{ik}}\right\|} \approx 0.04$. Hence, to simplify the calculation, we approximate (3.13) similar to (3.7). As a result, (3.12) becomes

$$\boldsymbol{R}_{\check{\boldsymbol{h}}_{jk'}\check{\boldsymbol{h}}_{jk'}} \approx \sum_{i\in A_1} p_{\mathrm{r}}\beta_{ijk'}K^2\boldsymbol{I}_M + \sum_{i\notin A_1}\sum_{k=1}^{K} p_{\mathrm{f}}\delta_{ji}\left(1+\frac{p_{\mathrm{f}}\delta_{ji}(1-\frac{1}{M})}{\phi_{ik}}\right)\boldsymbol{I}_M + K\boldsymbol{I}_M. \tag{3.14}$$

Next we find the cross-covariance of the small-scale channel and (3.3) given by

$$\boldsymbol{R}_{\boldsymbol{h}_{ljk'}\check{\boldsymbol{h}}_{jk'}} = \mathbb{E}\left[\boldsymbol{h}_{ljk'}^{H}\check{\boldsymbol{h}}_{jk'}\right].$$

The term $\check{\boldsymbol{h}}_{jk'}$ contains a summation of other interference terms, see (3.3), in addition to $\boldsymbol{h}_{ljk'}$. However, the expectation of the uncorrelated terms is zero. As a result, $\boldsymbol{R}_{\boldsymbol{h}_{ljk'}\check{\boldsymbol{h}}_{jk'}}$ becomes

$$\boldsymbol{R}_{\boldsymbol{h}_{ljk'}\check{\boldsymbol{h}}_{jk'}} = \sqrt{p_{\mathrm{r}}\beta_{ljk'}}K\boldsymbol{I}_M. \tag{3.15}$$

Substituting (3.14) and (3.15) into (3.5), the LMMSE estimate becomes (3.4).

□

Using (3.4) and (3.14), we can obtain the covariance of the LMMSE channel estimate as

$$\begin{aligned}\boldsymbol{R}_{\hat{\boldsymbol{h}}_{ljk'}\hat{\boldsymbol{h}}_{ljk'}} =& \mathbb{E}\left[\hat{\boldsymbol{h}}_{ljk'}^{H}\hat{\boldsymbol{h}}_{ljk'}\right] \\ =& \left(\frac{\sqrt{p_{\mathrm{r}}\beta_{ljk'}}}{\alpha_{jk'}^{\mathrm{TD}}}\right)^{2}\boldsymbol{R}_{\check{\boldsymbol{h}}_{jk'}\check{\boldsymbol{h}}_{jk'}} \\ =& \frac{p_{\mathrm{r}}\beta_{ljk'}K}{\alpha_{jk'}^{\mathrm{TD}}}\boldsymbol{I}_{M}. \end{aligned} \tag{3.16}$$

The LMMSE channel estimation error is $\tilde{\boldsymbol{h}}_{ljk'} = \boldsymbol{h}_{ljk'} - \hat{\boldsymbol{h}}_{ljk'}$, where $\tilde{\boldsymbol{h}}_{ljk'}$ is uncorrelated with $\hat{\boldsymbol{h}}_{ljk'}$ (see Appendix B for proof). Hence, the covariance matrix of the channel estimation error is given by

$$\begin{aligned}\boldsymbol{R}_{\tilde{\boldsymbol{h}}_{ljk'}\tilde{\boldsymbol{h}}_{ljk'}} =& \boldsymbol{R}_{\boldsymbol{h}_{ljk'}\boldsymbol{h}_{ljk'}} - \boldsymbol{R}_{\hat{\boldsymbol{h}}_{ljk'}\hat{\boldsymbol{h}}_{ljk'}} \\ =& \left(1 - \frac{p_{\mathrm{r}}\beta_{ljk'}K}{\alpha_{jk'}^{\mathrm{TD}}}\right)\boldsymbol{I}_{M}. \end{aligned} \tag{3.17}$$

We approximate the noise and interference during the channel estimation to be Gaussian random variables. This means that the estimated channel will also have a Gaussian distribution. Hence, based on results in (3.16) and (3.17), $\hat{\boldsymbol{h}}_{ljk'}$ and $\tilde{\boldsymbol{h}}_{ljk'}$ have the distributions $\mathcal{CN}\left(0, \frac{p_{\mathrm{r}}\beta_{ljk'}K}{\alpha_{jk'}^{\mathrm{TD}}}\boldsymbol{I}_{M}\right)$ and $\mathcal{CN}\left(0, \left(1 - \frac{p_{\mathrm{r}}\beta_{ljk'}K}{\alpha_{jk'}^{\mathrm{TD}}}\right)\boldsymbol{I}_{M}\right)$, respectively. We will use this property throughout the chapter.

3.3.2 Downlink transmission rate

The derivation of downlink transmission rate is based on the first half of the downlink data transmission phase of group A_1 (refer to Figure 3.2). In this phase, UEs in the group A_1 cells receive downlink data from the BSs in group A_1 and A_3 cells and uplink pilots from UEs in group A_2 cells. This transmission can be illustrated in Figure 3.4. If UE k' is located in cell j of group A_1, then the received signal for this UE during this

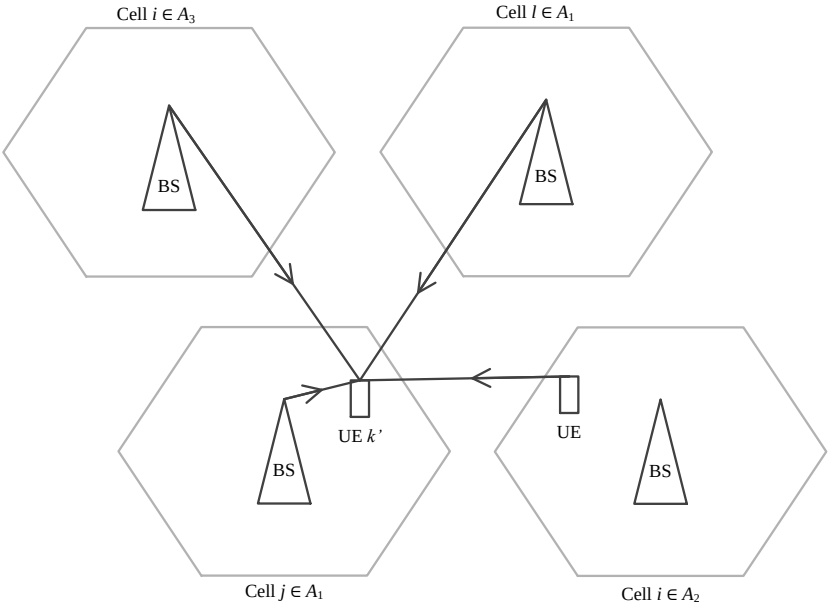

FIGURE 3.4: Received signal at UE k' in cell j during the downlink data transmission phase for group A_1 cells with TShdown.

phase is given by

$$x_{jk'}^{\mathrm{TD}} = \underbrace{\sum_{i \in A_1} \sum_{k=1}^{K} \sqrt{p_{\mathrm{f}} \beta_{jik'}} \boldsymbol{h}_{jik'} \boldsymbol{a}_{ik} q_{ik}}_{\text{Downlink data from BSs in group } A_1} + \underbrace{\sum_{i \in A_2} \sum_{k=1}^{K} \sqrt{p_{\mathrm{r}} \sigma_{jik'k}} v_{jik'k} \psi_k}_{\text{Uplink pilot from UEs in group } A_2}$$
$$+ \underbrace{\sum_{i \in A_3} \sum_{k=1}^{K} \sqrt{p_{\mathrm{f}} \beta_{jik'}} \boldsymbol{h}_{jik'} \boldsymbol{a}_{ik} q_{ik}}_{\text{Downlink data from BSs in group } A_3} + z_{jk'}, \tag{3.18}$$

where

- q_{ik} is the downlink data from BS cell i to UE k with $q_{ik} \sim \mathcal{CN}(0,1)$.
- $\sigma_{jik'k}$ is the large-scale fading from UE k in cell i to UE k' in cell j.
- $v_{jik'k}$ is the small-scale fading from UE k in cell i to UE k' in cell j with $v_{jik'k} \sim \mathcal{CN}(0,1)$.
- $z_{jk'}$ is the noise at UE k' in cell j with $z_{jk'} \sim \mathcal{CN}(0,1)$.

We define the MF precoding vector for UE k' in cell j as

$$\boldsymbol{a}_{jk'} \triangleq \frac{\hat{\boldsymbol{h}}_{jjk'}^{H}}{\sqrt{K} \left\| \hat{\boldsymbol{h}}_{jjk'}^{H} \right\|}. \tag{3.19}$$

The denominator of (3.19) is a normalization factor that ensures the expectation of the total precoding power is equal to 1.

Theorem 3.3. The closed form lower bound ergodic downlink rate of the TShdown method for UE k' in cell j is given by

$$R_{jk'}^{\mathrm{TD,DL}} = \frac{T_{\mathrm{d}}}{T} \times \log_2\left(1 + \frac{S_{\mathrm{a}}}{N_{\mathrm{a}}}\right), \tag{3.20}$$

where

$$S_{\mathrm{a}} = \frac{p_{\mathrm{f}} p_{\mathrm{r}} \beta_{jjk'}^2 \xi\left(M\right)^2}{\alpha_{jk'}^{\mathrm{TD}}} \tag{3.21}$$

$$\begin{aligned} N_{\mathrm{a}} = &\sum_{l \in A_1} \frac{p_{\mathrm{f}} \beta_{jlk'}}{K} \left(1 + \frac{p_{\mathrm{r}} \beta_{jlk'} K(M-1)}{\alpha_{lk'}^{\mathrm{TD}}}\right) - \frac{p_{\mathrm{f}} p_{\mathrm{r}} \beta_{jjk'}^2 \xi\left(M\right)^2}{\alpha_{jk'}^{\mathrm{TD}}} + \sum_{i \in A_1} \sum_{k \neq k'}^{K} \frac{p_{\mathrm{f}} \beta_{jik'}}{K} \\ &+ \sum_{i \in A_2} \sum_{k=1}^{K} p_{\mathrm{r}} \sigma_{jik'k} + \sum_{i \in A_3} \sum_{k=1}^{K} \frac{p_{\mathrm{f}} \beta_{jik'}}{K} + 1. \end{aligned} \tag{3.22}$$

T_{d} is the downlink data length, DL represents downlink, T is the frame length, S_{a} is effective downlink signal power for TShdown, N_{a} is the total downlink interference and noise power for TShdown, $\xi\left(M\right) = \frac{\Gamma(M+1/2)}{\Gamma(M)}$ and Γ is the gamma function.

Proof. Since the UE receiver does not have a channel estimate, we use a similar approach to [8] to obtain the achievable ergodic downlink rate for UE k' in cell j. This is given by

$$R_{jk'}^{\mathrm{TD,DL}} = \frac{T_{\mathrm{d}}}{T} \times \log_2\left(1 + \frac{S_{\mathrm{a}}}{N_{\mathrm{a}}}\right), \tag{3.23}$$

Using the method from [8], the average effective channel for the k'-th UE of cell j can be assumed to be $\sqrt{p_{\mathrm{f}} \beta_{jjk'}} \mathbb{E}\left[\boldsymbol{h}_{jjk'} \boldsymbol{a}_{jk'}\right]$. The variance of the effective channel gives the effective signal (desired) power of (3.23), given by

$$S_{\mathrm{a}} = p_{\mathrm{f}} \beta_{jjk'} \left|\mathbb{E}\left[\boldsymbol{h}_{jjk'} \boldsymbol{a}_{jk'}\right]\right|^2. \tag{3.24}$$

N_{a} in (3.23) is the total interference and noise power. It can be obtained from the total received power (variance of (3.18)) minus the effective signal power in (3.24),

$N_{\mathrm{a}} = \mathbb{E}\left[\left|x_{jk'}^{\mathrm{TD}}\right|^2\right] - S_{\mathrm{a}}$. To simplify the analysis, we expand N_{a} as

$$N_{\mathrm{a}} = \underbrace{\textstyle\sum_{l\in A_1} p_{\mathrm{f}}\beta_{jlk'}\mathbb{E}\left[\left|\boldsymbol{h}_{jlk'}\boldsymbol{a}_{lk'}\right|^2\right]}_{\text{Downlink data from BSs in group } A_1 \text{cell to UE } k'} + \underbrace{\textstyle\sum_{i\in A_1}\sum_{k\neq k'}^{K} p_{\mathrm{f}}\beta_{jik'}\mathbb{E}\left[\left|\boldsymbol{h}_{jik'}\boldsymbol{a}_{ik}\right|^2\right]}_{\text{Downlink data from BSs in cell } l\in A_1 \text{to UE } k\neq k'}$$
$$+ \underbrace{\textstyle\sum_{i\in A_2}\sum_{k=1}^{K} p_{\mathrm{r}}\sigma_{jik'k}\mathbb{E}\left[\left|v_{jik'k}\right|^2\right]}_{\text{Uplink pilot from UEs in cell group } A_2} + \underbrace{\textstyle\sum_{i\in A_3}\sum_{k=1}^{K} p_{\mathrm{f}}\beta_{jik'}\mathbb{E}\left[\left|\boldsymbol{h}_{jik'}\boldsymbol{a}_{ik}\right|^2\right]}_{\text{Downlink data from BSs in cell group } A_3}$$
$$+ \underbrace{1}_{\text{AWGN noise}} - \underbrace{p_{\mathrm{f}}\beta_{jjk'}\left|\mathbb{E}\left[\boldsymbol{h}_{jjk'}\boldsymbol{a}_{jk'}\right]\right|^2}_{S_{\mathrm{a}}}. \tag{3.25}$$

To solve the expectation term in the effective signal power, (3.24), we separate the correlated and uncorrelated terms as [8]

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{h}_{jjk'}\boldsymbol{a}_{jk'}\right] &= \mathbb{E}\left[\hat{\boldsymbol{h}}_{jjk'}\frac{\hat{\boldsymbol{h}}_{jjk'}^H}{\sqrt{K}\left\|\hat{\boldsymbol{h}}_{jjk'}^H\right\|} + \tilde{\boldsymbol{h}}_{jjk'}\frac{\hat{\boldsymbol{h}}_{jjk'}^H}{\sqrt{K}\left\|\hat{\boldsymbol{h}}_{jjk'}^H\right\|}\right] \\ &= \mathbb{E}\left[\frac{\left\|\hat{\boldsymbol{h}}_{jjk'}\right\|}{\sqrt{K}}\right] \\ &= \sqrt{\frac{\left\|\boldsymbol{R}_{\hat{\boldsymbol{h}}_{jjk'}\hat{\boldsymbol{h}}_{jjk'}}\right\|}{K}}\mathbb{E}\left[\theta\right].\end{aligned} \tag{3.26}$$

Note that θ has a Chi distribution that is scaled by a factor of $1/\sqrt{2}$ and has $2M$ degrees of freedom [8]. Using (3.16) and the expectation of a Chi distribution [50], (3.26) becomes

$$\mathbb{E}\left[\boldsymbol{h}_{jjk'}\boldsymbol{a}_{jk'}\right] = \sqrt{\frac{p_{\mathrm{r}}\beta_{jjk'}}{\alpha_{jk'}^{\mathrm{TD}}}}\xi\left(M\right), \tag{3.27}$$

where $\xi\left(M\right) = \frac{\Gamma(M+1/2)}{\Gamma(M)}$ and Γ is the gamma function.

The first expectation term in (3.25) involves the signal power from the BS to UEs of the same group and same pilot sequence, $k = k'$. In this term, the precoding vector is correlated with the channel which the signal propagates through. Therefore, we separate the channel $\boldsymbol{h}_{jlk'}$ into correlated and uncorrelated terms [8]. This gives

$$\mathbb{E}\left[\left|\boldsymbol{h}_{jlk'}\boldsymbol{a}_{lk'}\right|^2\right] = \mathbb{E}\left[\left|\hat{\boldsymbol{h}}_{jlk'}\frac{\hat{\boldsymbol{h}}_{llk'}^H}{\sqrt{K}\left\|\hat{\boldsymbol{h}}_{llk'}^H\right\|}\right|^2\right] + \mathbb{E}\left[\left|\tilde{\boldsymbol{h}}_{jlk'}\frac{\hat{\boldsymbol{h}}_{llk'}^H}{\sqrt{K}\left\|\hat{\boldsymbol{h}}_{llk'}^H\right\|}\right|^2\right]. \tag{3.28}$$

Using (3.3) and (3.4), it can be shown that $\hat{\boldsymbol{h}}_{llk'} = \sqrt{\frac{\beta_{llk'}}{\beta_{jlk'}}}\hat{\boldsymbol{h}}_{jlk'}$. This also means that $\frac{\hat{\boldsymbol{h}}_{llk'}^H}{\left\|\hat{\boldsymbol{h}}_{llk'}^H\right\|} = \frac{\hat{\boldsymbol{h}}_{jlk'}^H}{\left\|\hat{\boldsymbol{h}}_{jlk'}^H\right\|}$. Using this, along with Lemma C.1 in Appendix C, (3.28) becomes

$$\begin{aligned}\mathbb{E}\left[\left|\boldsymbol{h}_{jlk'}\boldsymbol{a}_{lk'}\right|^2\right] =& \mathbb{E}\left[\left|\hat{\boldsymbol{h}}_{jlk'}\frac{\hat{\boldsymbol{h}}_{jlk'}^H}{\sqrt{K}\left\|\hat{\boldsymbol{h}}_{jlk'}^H\right\|}\right|^2\right] + \mathbb{E}\left[\left|\tilde{\boldsymbol{h}}_{jlk'}\frac{\hat{\boldsymbol{h}}_{jlk'}^H}{\sqrt{K}\left\|\hat{\boldsymbol{h}}_{jlk'}^H\right\|}\right|^2\right] \\ =& \frac{1}{K}\left(\mathrm{tr}\left(\boldsymbol{R}_{\hat{\boldsymbol{h}}_{jlk'}\hat{\boldsymbol{h}}_{jlk'}}\right) + \frac{1}{M}\mathrm{tr}\left(\boldsymbol{R}_{\tilde{\boldsymbol{h}}_{jlk'}\tilde{\boldsymbol{h}}_{jlk'}}\right)\right). \end{aligned} \tag{3.29}$$

Using (3.16) and (3.17), (3.29) becomes

$$\mathbb{E}\left[\left|\boldsymbol{h}_{jlk'}\boldsymbol{a}_{lk'}\right|^2\right] = \left(1 + \frac{p_{\mathrm{r}}\beta_{jlk'}K(M-1)}{\alpha_{lk'}^{\mathrm{TD}}}\right)\frac{1}{K}. \tag{3.30}$$

The second and fourth terms in (3.25) are the transmitted signals to UEs of the same group, but with different pilot sequences ($k \neq k'$) and to UEs of different cell groups, respectively. In this case, the precoding vector is uncorrelated from the channel vector. As a result, the expectation value in these terms can be simplified to

$$\begin{aligned}\mathbb{E}\left[\left|\boldsymbol{h}_{jik'}\boldsymbol{a}_{ik}\right|^2\right] =& \mathbb{E}\left[\left|\boldsymbol{h}_{jik'}\frac{\hat{\boldsymbol{h}}_{iik}^H}{\sqrt{K}\left\|\hat{\boldsymbol{h}}_{iik}^H\right\|}\right|^2\right] \\ =& \frac{1}{K}\left(\frac{1}{M}\mathrm{tr}\left(\boldsymbol{R}_{\boldsymbol{h}_{jik'}\boldsymbol{h}_{jik'}}\right)\right) \\ =& \frac{1}{K}. \end{aligned} \tag{3.31}$$

The third expectation term in (3.25) involves the signal between UEs of different cell groups. This is given by

$$\mathbb{E}\left[\left|v_{jik'k}\right|^2\right] = 1. \tag{3.32}$$

Substituting (3.27), (3.30), (3.31) and (3.32) into (3.24) and (3.25), the lower bound ergodic downlink rate for UE k' in cell j in (3.23) becomes (3.20).

□

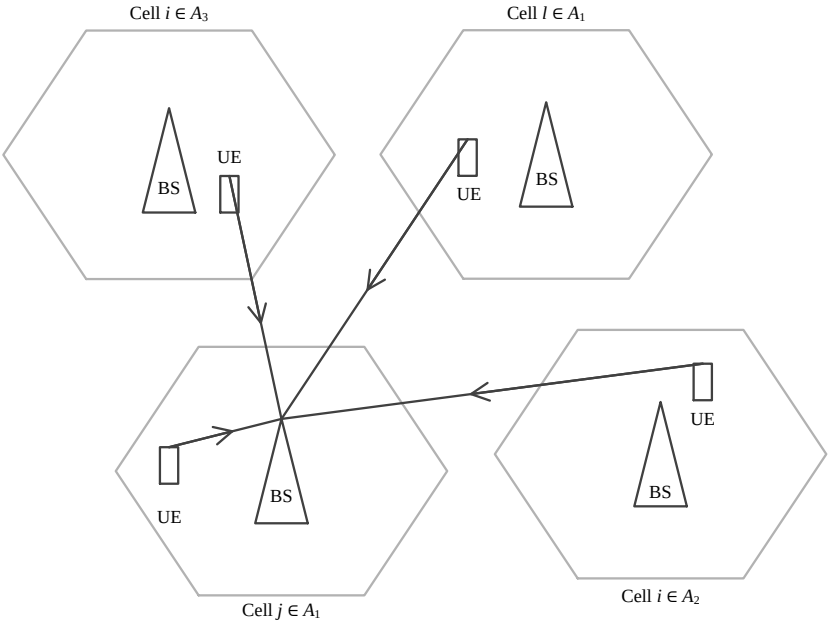

FIGURE 3.5: Received signal at BS in cell j during group A_1 cells uplink data transmission phase for TShdown.

The closed form achievable ergodic downlink rate for TShdown, (3.20), shows that the interference from signals directed to UEs with the same pilot sequence in the same cell group increases linearly with the number of antennas. However, interference from different cell groups is not affected by the number of antennas.

3.3.3 Uplink transmission rate

As seen in Figure 3.2, uplink data transmission is performed simultaneously in all cells. This transmission can be illustrated in Figure 3.5. If the received signal at the cell j BS of group A_1 is $\boldsymbol{y}_j$ during uplink data, then the estimated uplink data from UE k' in cell j is $\hat{d}_{jk}^{\mathrm{TD}} = \boldsymbol{y}_j \boldsymbol{b}_{jk'}$. This can be expanded as

$$\hat{d}_{jk'}^{\mathrm{TD}} = \underbrace{\sum_{i \in A_1} \sum_{k=1}^{K} \sqrt{p_{\mathrm{u}} \beta_{ijk}} d_{ik} \boldsymbol{h}_{ijk} \boldsymbol{b}_{jk'}}_{\text{Uplink data from UEs in group } A_1} + \underbrace{\sum_{i \notin A_1} \sum_{k=1}^{K} \sqrt{p_{\mathrm{u}} \beta_{ijk}} d_{ik} \boldsymbol{h}_{ijk} \boldsymbol{b}_{jk'}}_{\text{Uplink data from UEs in groups } A_2 \text{ and } A_3} + \boldsymbol{z}_j \boldsymbol{b}_{jk'}, \tag{3.33}$$

where

– p_{u} is the average UE transmit data power.

– d_{ik} is the uplink data from UE k, cell i with $d_{ik} \sim \mathcal{CN}(0,1)$.

– $\boldsymbol{z}_j$ is the $1 \times M$ noise vector at cell j BS with $\boldsymbol{z}_j \sim \mathcal{CN}(0, \boldsymbol{I}_M)$.

– $\boldsymbol{b}_{jk'}$ is the $M \times 1$ equalizer vector for UE k' in cell j.

We define the equalizer vector for UE k' in cell j as

$$\boldsymbol{b}_{jk'} \triangleq \frac{\hat{\boldsymbol{h}}_{jjk'}^{H}}{\left\| \hat{\boldsymbol{h}}_{jjk'}^{H} \right\|}. \tag{3.34}$$

Unlike the precoding vector in (3.19), we exclude the term $1/\sqrt{K}$ in the equalizer because the term will be cancelled out during the division process in the uplink rate equation.

Theorem 3.4. The closed form lower bound ergodic uplink rate of the TShdown method for UE k' in cell j is given by

$$\tilde{R}_{jk'}^{\mathrm{TD,UL}} = \frac{T_{\mathrm{u}}}{T} \times \log_2 \left(1 + \frac{\tilde{S}_{\mathrm{b}}}{\tilde{N}_{\mathrm{b}}} \right), \tag{3.35}$$

where

$$\tilde{S}_{\mathrm{b}} = \frac{p_{\mathrm{u}} p_{\mathrm{r}} \beta_{jjk'}^2 K(M-1)}{\alpha_{jk'}^{\mathrm{TD}}}, \tag{3.36}$$

$$\begin{aligned} \tilde{N}_{\mathrm{b}} =& p_{\mathrm{u}} \beta_{jjk'} \left(1 - \frac{p_{\mathrm{r}} \beta_{jjk'} K}{\alpha_{jk'}^{\mathrm{TD}}} \right) + \sum_{l \in A_1 \cap l \neq j} p_{\mathrm{u}} \beta_{ljk'} \left(1 + \frac{p_{\mathrm{r}} \beta_{ljk'} K(M-2)}{\alpha_{jk'}^{\mathrm{TD}}} \right) \\ &+ \sum_{i \in A_1} \sum_{k \neq k'}^{K} p_{\mathrm{u}} \beta_{ijk} + \sum_{i \notin A_1} \sum_{k=1}^{K} p_{\mathrm{u}} \beta_{ijk} + 1. \end{aligned} \tag{3.37}$$

T_{u} is the uplink data length and UL represents uplink, $\tilde{S}_{\mathrm{b}}$ is effective uplink signal power for TShdown, $\tilde{N}_{\mathrm{b}}$ is the total uplink interference and noise power for TShdown.

Proof. The information-theoretic method to derive downlink rate in Theorem 3.3 is based on [8]. Such a method is specifically designed for downlink transmission where the receiver (UE) does not have channel information and the transmission depends on BS beamforming to overcome channel distortion. For the uplink, the channel is estimated at the receiver (BS). In this case, we use a similar approach to that in [5] to obtain the

lower bound ergodic uplink rate for UE k' in cell j as

$$R_{jk'}^{\mathrm{TD,UL}} = \frac{T_{\mathrm{u}}}{T} \times \mathbb{E}\left[\log_2\left(1 + \frac{S_{\mathrm{b}}}{N_{\mathrm{b}}}\right)\right], \tag{3.38}$$

Due to the complexity of the expectation in (3.38), we employ Jensen's inequality as in [5], to obtain

$$\begin{aligned} R_{jk'}^{\mathrm{TD,UL}} \geq & \tilde{R}_{jk'}^{\mathrm{TD,UL}} \\ \tilde{R}_{jk'}^{\mathrm{TD,UL}} = & \frac{T_{\mathrm{u}}}{T} \times \log_2\left(1 + \left(\mathbb{E}\left[\frac{N_{\mathrm{b}}}{S_{\mathrm{b}}}\right]\right)^{-1}\right). \end{aligned} \tag{3.39}$$

S_{b} in (3.39) is the effective signal (desired) power of the estimated data. Using the method in [5], the uplink effective signal is $\sqrt{p_{\mathrm{u}}\beta_{jjk'}}d_{jk'}\hat{\boldsymbol{h}}_{jjk'}\boldsymbol{b}_{jk'}$. Hence, S_{b} can be written as

$$S_{\mathrm{b}} = p_{\mathrm{u}}\beta_{jjk'}\left|\hat{\boldsymbol{h}}_{jjk'}\boldsymbol{b}_{jk'}\right|^2. \tag{3.40}$$

N_{b} in (3.39) is the total interference and noise power. It equals the total power of estimated data (3.33) minus the effective signal power (3.40), $N_{\mathrm{b}} = \left|\hat{d}_{jk'}^{\mathrm{TD}}\right|^2 - S_{\mathrm{b}}$. We can expand N_{b} into uncorrelated interference terms and $\mathbb{E}\left[\frac{N_{\mathrm{b}}}{S_{\mathrm{b}}}\right]$ becomes

$$\begin{aligned} \mathbb{E}\left[\frac{N_{\mathrm{b}}}{S_{\mathrm{b}}}\right] = & \underbrace{p_{\mathrm{u}}\beta_{jjk'}\mathbb{E}\left[\frac{\left|\tilde{\boldsymbol{h}}_{jjk'}\boldsymbol{b}_{jk'}\right|^2}{S_{\mathrm{b}}}\right]}_{\text{Channel estimation error for UE } k' \text{ in cell } j} + \underbrace{\sum_{l\in A_1 \cap l\neq j} p_{\mathrm{u}}\beta_{ljk'}\mathbb{E}\left[\frac{\left|\boldsymbol{h}_{ljk'}\boldsymbol{b}_{jk'}\right|^2}{S_{\mathrm{b}}}\right]}_{\text{Uplink data from UE } k' \text{ in cell } l\in A_1\cap l\neq j} \\ & + \underbrace{\sum_{i\in A_1}\sum_{k\neq k'}^{K} p_{\mathrm{u}}\beta_{ijk}\mathbb{E}\left[\frac{\left|\boldsymbol{h}_{ijk}\boldsymbol{b}_{jk'}\right|^2}{S_{\mathrm{b}}}\right]}_{\text{Uplink data from UE } k\neq k' \text{ in cell } i\in A_1} + \underbrace{\sum_{i\notin A_1}\sum_{k=1}^{K} p_{\mathrm{u}}\beta_{ijk}\mathbb{E}\left[\frac{\left|\boldsymbol{h}_{ijk}\boldsymbol{b}_{jk'}\right|^2}{S_{\mathrm{b}}}\right]}_{\text{Uplink data from UEs in cell group } A_2 \text{ and } A_3} \\ & + \underbrace{\mathbb{E}\left[\frac{\left|\boldsymbol{z}_j\boldsymbol{b}_{jk'}\right|^2}{S_{\mathrm{b}}}\right]}_{\text{AWGN noise}}. \end{aligned} \tag{3.41}$$

Note that the channel estimation error term in (3.41) is obtained from $\mathbb{E}\left[\frac{\left|\tilde{\boldsymbol{h}}_{jjk'}\boldsymbol{b}_{jk'}\right|^2}{S_{\mathrm{b}}}\right] = \mathbb{E}\left[\frac{\left|\boldsymbol{h}_{jjk'}\boldsymbol{b}_{jk'}\right|^2}{S_{\mathrm{b}}}\right] - \mathbb{E}\left[\frac{S_{\mathrm{b}}}{S_{\mathrm{b}}}\right]$. This is due to the property where $\tilde{\boldsymbol{h}}_{jjk'} = \boldsymbol{h}_{jjk'} - \hat{\boldsymbol{h}}_{jjk'}$, and $\tilde{\boldsymbol{h}}_{jjk'}$ is uncorrelated with $\hat{\boldsymbol{h}}_{jjk'}$.

To find the closed form of the expectation terms in (3.41), we first solve $\mathbb{E}\left[\frac{1}{S_{\mathrm{b}}}\right]$. Using (3.34) and (3.40), we can write

$$\begin{aligned}\mathbb{E}\left[\frac{1}{S_{\mathrm{b}}}\right] =& \mathbb{E}\left[\frac{1}{p_{\mathrm{u}}\beta_{jjk'}\left|\hat{\boldsymbol{h}}_{jjk'}\boldsymbol{b}_{jk'}\right|^2}\right] \\ =& \frac{1}{p_{\mathrm{u}}\beta_{jjk'}\left\|\boldsymbol{R}_{\hat{\boldsymbol{h}}_{jjk'}\hat{\boldsymbol{h}}_{jjk'}}\right\|}\mathbb{E}\left[\frac{1}{\theta^2}\right].\end{aligned} \tag{3.42}$$

Note that θ^2 has a Chi-squared distribution that is scaled by a factor of $1/2$ and has $2M$ degrees of freedom. It can be shown that $\mathbb{E}\left[\frac{1}{\theta^2}\right] = 1/(M-1)$ [51]. Using this and (3.16), then (3.42) becomes

$$\begin{aligned}\mathbb{E}\left[\frac{1}{S_{\mathrm{b}}}\right] =& \frac{1}{p_{\mathrm{u}}p_{\mathrm{r}}\beta_{jjk'}^2 K(M-1)/\alpha_{jk'}^{\mathrm{TD}}} \\ =& \frac{1}{\lambda^{\mathrm{TD}}},\end{aligned} \tag{3.43}$$

where $\lambda^{\mathrm{TD}} = p_{\mathrm{u}}p_{\mathrm{r}}\beta_{jjk'}^2 K(M-1)/\alpha_{jk'}^{\mathrm{TD}}$.

The first expectation term in (3.41) involves interference due to channel estimation error. Since we approximate the channel estimate as Gaussian, and since $\tilde{\boldsymbol{h}}_{jjk}$ and $\hat{\boldsymbol{h}}_{jjk}$ is uncorrelated, we can apply similar derivation in [5, eq. (76)] to this expectation term and get

$$\begin{aligned}\mathbb{E}\left[\frac{\left|\tilde{\boldsymbol{h}}_{jjk}\boldsymbol{b}_{jk'}\right|^2}{S_{\mathrm{b}}}\right] =& \mathbb{E}\left[\frac{\left|\tilde{\boldsymbol{h}}_{jjk}\boldsymbol{b}_{jk'}\right|^2}{p_{\mathrm{u}}\beta_{jjk'}\left|\hat{\boldsymbol{h}}_{jjk'}\boldsymbol{b}_{jk'}\right|^2}\right] \\ =& \mathbb{E}\left[\left|\tilde{\boldsymbol{h}}_{jjk}\boldsymbol{b}_{jk'}\right|^2\right]\mathbb{E}\left[\frac{1}{p_{\mathrm{u}}\beta_{jjk'}\left|\hat{\boldsymbol{h}}_{jjk'}\boldsymbol{b}_{jk'}\right|^2}\right]\end{aligned} \tag{3.44}$$

Using (3.34), (3.43) and Lemma C.1, (3.44) becomes

$$\begin{aligned}\mathbb{E}\left[\frac{\left|\tilde{\boldsymbol{h}}_{jjk}\boldsymbol{b}_{jk'}\right|^2}{S_{\mathrm{b}}}\right] =& \mathbb{E}\left[\left|\tilde{\boldsymbol{h}}_{jjk}\frac{\hat{\boldsymbol{h}}_{jjk'}^H}{\left\|\hat{\boldsymbol{h}}_{jjk'}^H\right\|}\right|^2\right]\mathbb{E}\left[\frac{1}{S_{\mathrm{b}}}\right] \\ =& \frac{1}{M}\mathrm{tr}\left(\boldsymbol{R}_{\tilde{\boldsymbol{h}}_{jjk}\tilde{\boldsymbol{h}}_{jjk}}\right)\mathbb{E}\left[\frac{1}{S_{\mathrm{b}}}\right] \\ =& \left(1-\frac{p_{\mathrm{r}}\beta_{jjk'}K}{\alpha_{jk'}^{\mathrm{TD}}}\right)\frac{1}{\lambda^{\mathrm{TD}}}.\end{aligned} \tag{3.45}$$

The second expectation term in (3.41) involves the signal power received from UEs of the same group and the same pilot sequence. Due to pilot contamination, the channel is partially correlated with the equalizer. Since we approximate the channel estimate as complex Gaussian, we can separate the channel into correlated and uncorrelated parts as follows

$$\begin{aligned}\mathbb{E}\left[\frac{\left|\boldsymbol{h}_{ljk'}\boldsymbol{b}_{jk'}\right|^2}{S_\mathrm{b}}\right] =&\mathbb{E}\left[\frac{\left|\hat{\boldsymbol{h}}_{ljk'}\boldsymbol{b}_{jk'}\right|^2}{S_\mathrm{b}}\right]+\mathbb{E}\left[\frac{\left|\tilde{\boldsymbol{h}}_{ljk'}\boldsymbol{b}_{jk'}\right|^2}{S_\mathrm{b}}\right]\\ =&\mathbb{E}\left[\frac{\left|\hat{\boldsymbol{h}}_{ljk'}\boldsymbol{b}_{jk'}\right|^2}{p_\mathrm{u}\beta_{jjk'}\left|\hat{\boldsymbol{h}}_{jjk'}\boldsymbol{b}_{jk'}\right|^2}\right]+\mathbb{E}\left[\frac{\left|\tilde{\boldsymbol{h}}_{ljk'}\boldsymbol{b}_{jk'}\right|^2}{p_\mathrm{u}\beta_{jjk'}\left|\hat{\boldsymbol{h}}_{jjk'}\boldsymbol{b}_{jk'}\right|^2}\right].\end{aligned} \tag{3.46}$$

Using (3.3) and (3.4), it can be shown that $\hat{\boldsymbol{h}}_{jjk'} = \sqrt{\frac{\beta_{jjk'}}{\beta_{ljk'}}}\hat{\boldsymbol{h}}_{ljk'}$. Using this, we can simplify the first expectation in (3.46) with simple division, as follow

$$\begin{aligned}\mathbb{E}\left[\frac{\left|\boldsymbol{h}_{ljk'}\boldsymbol{b}_{jk'}\right|^2}{S_\mathrm{b}}\right] =&\mathbb{E}\left[\frac{\beta_{ljk'}\left|\hat{\boldsymbol{h}}_{ljk'}\boldsymbol{b}_{jk'}\right|^2}{p_\mathrm{u}\beta_{jjk'}^2\left|\hat{\boldsymbol{h}}_{ljk'}\boldsymbol{b}_{jk'}\right|^2}\right]+\mathbb{E}\left[\frac{\left|\tilde{\boldsymbol{h}}_{ljk'}\boldsymbol{b}_{jk'}\right|^2}{p_\mathrm{u}\beta_{jjk'}\left|\hat{\boldsymbol{h}}_{jjk'}\boldsymbol{b}_{jk'}\right|^2}\right]\\ =&\frac{\beta_{ljk'}}{p_\mathrm{u}\beta_{jjk'}^2}+\mathbb{E}\left[\frac{\left|\tilde{\boldsymbol{h}}_{ljk'}\boldsymbol{b}_{jk'}\right|^2}{p_\mathrm{u}\beta_{jjk'}\left|\hat{\boldsymbol{h}}_{jjk'}\boldsymbol{b}_{jk'}\right|^2}\right].\end{aligned} \tag{3.47}$$

The second term in (3.47) is the channel estimation error term which can be solved using a similar method to (3.45). Therefore, (3.47) becomes

$$\begin{aligned}\mathbb{E}\left[\frac{\left|\boldsymbol{h}_{ljk'}\boldsymbol{b}_{jk'}\right|^2}{S_\mathrm{b}}\right] =&\frac{\beta_{ljk'}}{p_\mathrm{u}\beta_{jjk'}^2}+\mathbb{E}\left[\left|\tilde{\boldsymbol{h}}_{ljk}\frac{\hat{\boldsymbol{h}}_{jjk'}^H}{\left\|\hat{\boldsymbol{h}}_{jjk'}^H\right\|}\right|^2\right]\mathbb{E}\left[\frac{1}{S_\mathrm{b}}\right]\\ =&\frac{\beta_{ljk'}}{p_\mathrm{u}\beta_{jjk'}^2}+\frac{1}{M}\mathrm{tr}\left(\boldsymbol{R}_{\tilde{\boldsymbol{h}}_{ljk}\tilde{\boldsymbol{h}}_{ljk}}\right)\mathbb{E}\left[\frac{1}{S_\mathrm{b}}\right]\\ =&\frac{\beta_{ljk'}}{p_\mathrm{u}\beta_{jjk'}^2}+\left(1-\frac{p_\mathrm{r}\beta_{ljk'}K}{\alpha_{jk'}^{\mathrm{TD}}}\right)\frac{1}{\lambda^{\mathrm{TD}}}\\ =&\left(1+\frac{p_\mathrm{r}\beta_{ljk'}K(M-2)}{\alpha_{jk'}^{\mathrm{TD}}}\right)\frac{1}{\lambda^{\mathrm{TD}}}.\end{aligned} \tag{3.48}$$

The third and fourth terms in (3.41) involve signal power from UEs of the same group, but with different pilot sequences and from UEs of different cell groups, respectively. Since the equalizer vector, $\boldsymbol{b}_{jk'}$, is uncorrelated with the channel vector, $\boldsymbol{h}_{ijk}$, for both

terms, then their expectation terms can be written as

$$\begin{aligned}\mathbb{E}\left[\frac{\left|\boldsymbol{h}_{ijk}\boldsymbol{b}_{jk'}\right|^2}{S_{\mathrm{b}}}\right] =& \mathbb{E}\left[\left|\boldsymbol{h}_{ijk}\frac{\hat{\boldsymbol{h}}_{jjk'}^H}{\left\|\hat{\boldsymbol{h}}_{jjk'}^H\right\|}\right|^2\right]\mathbb{E}\left[\frac{1}{S_{\mathrm{b}}}\right] \\ =& \frac{1}{M}\mathrm{tr}\left(\boldsymbol{R}_{\boldsymbol{h}_{ijk}\boldsymbol{h}_{ijk}}\right)\mathbb{E}\left[\frac{1}{S_{\mathrm{b}}}\right] \\ =& \frac{1}{\lambda^{\mathrm{TD}}}. \end{aligned} \tag{3.49}$$

Equation (3.49) also applies to the AWGN noise term. Substituting (3.43), (3.45), (3.48) and (3.49) into (3.41), then the lower bound ergodic rate for UE k' in (3.39) becomes (3.35).

□

Similar to the downlink rate in (3.20), we can see the relationship between interference and the number of antennas in the uplink rate of (3.35). It shows that the interference from UEs with the same pilot sequence in the same cell group increases linearly with the number of antennas while interference from a different cell group is not affected by the number of antennas.

3.3.4 Power optimization

In the time-shifted methods, if the data power is too high, then there will be high interference during channel estimation resulting in a low transmission rate. This is because if we refer to Figure 3.2 for TShdown, the downlink data is transmitted at the same time as other cells transmit their pilot. Hence, the high data power will cause high channel estimation error. However, if the data power is too low, then the received power will be too low. Therefore, we need to find a balance for the transmit power level. This type of power optimization is a unique advantage of time-shifted methods over a time-synchronized pilot. The same strategy cannot be applied to a time-synchronized pilot due to the fact that the pilot and the data transmission time with a time-synchronized pilot do not intersect. For TShdown, since the pilot is overlapped with the downlink data, we find the optimum downlink transmit power, p_{f}, that maximizes the downlink sum rate. The optimum power can be found using an exhaustive search based on the available large-scale fading information. We can also derive the optimum power based on

the sum rate equation. An expression for optimum transmit power can be approximated by differentiating (3.20) with respect to p_{f}, and solving for p_{f} in $\frac{d}{dp_{\mathrm{f}}}\left(R_{jk'}^{\mathrm{TD,DL}}\right) = 0$ with $p_{\mathrm{f}} > 0$ (see Appendix D.1 for more detail). The result is given as

$$p_{\mathrm{f}} = \sqrt{\frac{\left(1 + p_{\mathrm{r}} K \sum_{i \in A_1} \beta_{ijk'}\right)\left(1 + p_{\mathrm{r}} \sum_{i \in A_2} \sum_{k=1}^{K} \sigma_{jik'k}\right)}{\left(\sum_{i \notin A_1} \delta_{ji}\right)\left(\sum_{i \in A_1} \sum_{k=1}^{K} \frac{\beta_{jik'}}{K} + \sum_{i \in A_3} \sum_{k=1}^{K} \frac{\beta_{jik'}}{K}\right)}}. \tag{3.50}$$

Note that the p_{f} obtained in (3.50) is only optimized for UE k' in cell j. In order to obtain a common value for p_{f} across all UEs in all cells, we average the result in (3.50). Simulations show that (3.50) generates values with a distribution resembling log-normal (as shown in Figure 3.14 in the results section). Therefore, we use the geometric mean to find the average.

Note that (3.50) requires the use of large scale fading information. The acquisition of the large scale fading values should be similar to the power measurement when a UE starts to establish a connection with a BS. This power measurement will require an overhead signal, which is already being used by current wireless standards for cell search and power control [11]. Therefore, the measured values can be collected and we can apply it to (3.50). The large scale value must also be shared among the BSs. This can be achieved through networked BSs.

Based on (3.50), as K increases, the optimum power also increases. This is expected because as there are more UEs, there will be more interference from uplink pilots from other cells' UEs during the downlink data transmission. Thus the downlink data power needs to increase to overcome the increase in the interference. To get some mathematical insights on the relationship between optimum power and K, we assume that the values of $\beta_{jik'}$ and $\sigma_{jik'k}$ are the same for all UEs. Applying this assumption on (3.50), as K grows very large, the optimum power will increase approximately linearly to K.

3.4 Time-shifted pilot with uplink data overlap

In Section 3.3, the pilot is set to overlap with downlink data. In this section, we analyze another variation where the uplink pilot overlaps with uplink data instead of downlink data. From now on, the term for the time-shifted pilot method with uplink data overlap

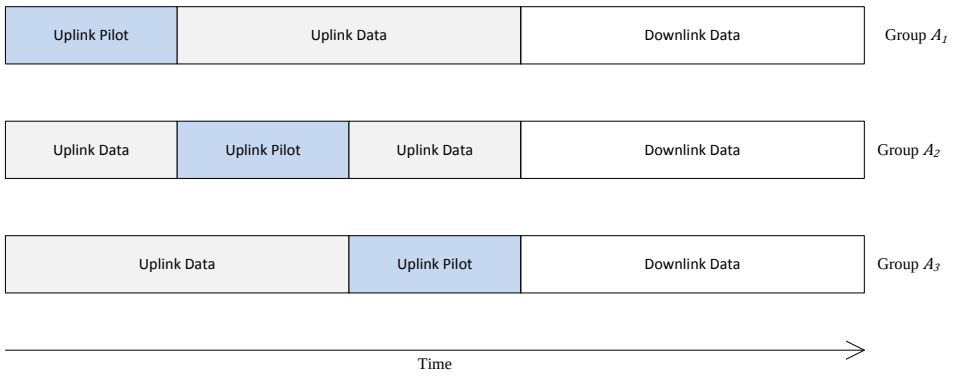

FIGURE 3.6: The arrangement of time-shifted pilots for TShup.

is simplified to TShup (and superscript TU in equations). As discussed in the introduction, TShup has been briefly analyzed in [52–54] without detailed sum rate derivations and performance analysis. Hence, in this section we derive novel closed form ergodic achievable rate expressions for TShup in order to provide a more complete analysis of the method. As shown in Figure 3.6, for TShup the uplink pilot from one group will only overlap with uplink data from other groups so that pilot to pilot interference between different groups does not occur, but uplink data to pilot interference can still occur. The main motivation to investigate TShup is the practical advantage over TShdown in terms of uplink and downlink time coordination. This is because the TShdown method requires the uplink and downlink to be changed more than once within one transmission frame (see Figure 3.2), but this only occurs once in the TShup method.

The important finding in this section is the inter-cell interference behaves differently in TShup compared to TShdown. This is because in TShup, both interference from the same and different cell groups increase linearly with the number of antennas. For interference from different cell groups, this effect can occur for all UEs rather than only UEs with the same pilot sequences. The proof of this finding will be discussed in the sum rate derivation sections.

3.4.1 Channel estimation

Channel estimation is done at the BS using the uplink pilot. Referring to Figure 3.6, during the uplink pilot transmission phase in group A_1, UEs in this group transmit pilot signals while UEs in other groups transmit data signals. This transmission can be illustrated in Figure 3.7. If cell j belongs to group A_1, then the received signal at BS

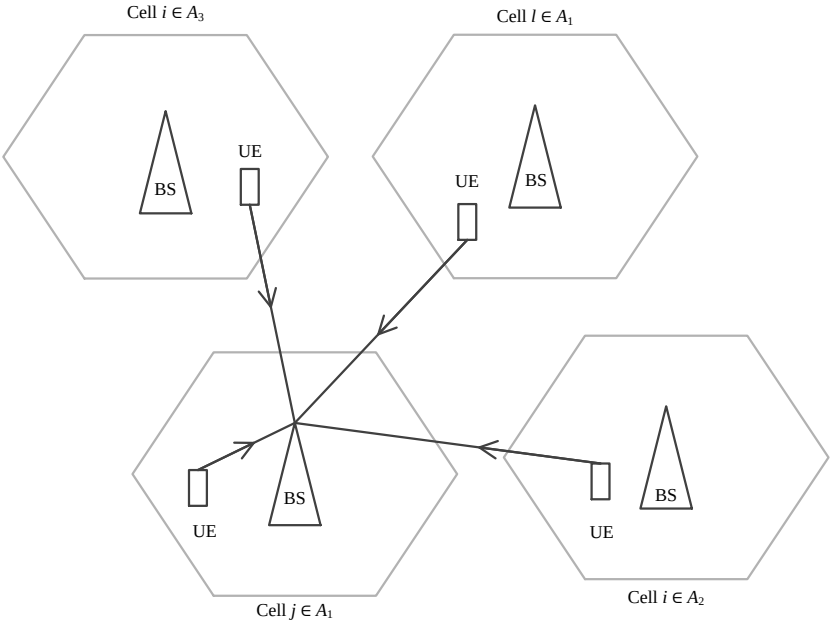

FIGURE 3.7: Received signal at BS in cell j during the uplink pilot transmission phase for group A_1 cells with TShup.

cell j during this phase can be expressed as

$$\boldsymbol{Y}_j = \underbrace{\sum_{i \in A_1} \sum_{k=1}^{K} \sqrt{p_{\mathrm{r}} \beta_{ijk}} \boldsymbol{\psi}_k \boldsymbol{h}_{ijk}}_{\text{Uplink pilot from UEs in group } A_1} + \underbrace{\sum_{i \notin A_1} \sum_{k=1}^{K} \sqrt{p_{\mathrm{u}} \beta_{ijk}} \boldsymbol{d}_{ik} \boldsymbol{h}_{ijk}}_{\text{Uplink data from UEs in groups } A_2 \text{ and } A_3} + \boldsymbol{Z}_j, \quad (3.51)$$

where $\boldsymbol{d}_{ik}$ is the $K \times 1$ uplink data vector from the BS in cell i to UE k with $\boldsymbol{d}_{ik} \sim \mathcal{CN}(0, \boldsymbol{I}_K)$.

We can use (3.51) to acquire the correlation between the pilot and received signal using the same method used to obtain (3.3). This gives

$$\check{\boldsymbol{h}}_{jk'} = \sum_{i \in A_1} \sqrt{p_{\mathrm{r}} \beta_{ijk'}} K \boldsymbol{h}_{ijk'} + \sum_{i \notin A_1} \sum_{k=1}^{K} \sqrt{p_{\mathrm{u}} \beta_{ijk}} \boldsymbol{\psi}_{k'}^H \boldsymbol{d}_{ik} \boldsymbol{h}_{ijk} + \boldsymbol{\psi}_{k'}^H \boldsymbol{Z}_j. \quad (3.52)$$

Theorem 3.5. Using (3.52), if cells l and j are in the same cell group, then LMMSE estimation of the channel between user k' in cell l to the BS in cell j is

$$\hat{\boldsymbol{h}}_{ljk'} = \frac{\sqrt{p_{\mathrm{r}} \beta_{ljk'}}}{\alpha_{jk'}^{\mathrm{TU}}} \check{\boldsymbol{h}}_{ljk'}, \quad (3.53)$$

where $\alpha_{jk'}^{\mathrm{TU}} = \sum_{i \in A_1} p_{\mathrm{r}} \beta_{ijk'} K + \sum_{i \notin A_1} \sum_{k=1}^{K} p_{\mathrm{u}} \beta_{ijk} + 1$.

Proof. We use the general expression (3.5) to obtain the LMMSE channel estimate. Therefore, we need to find $\boldsymbol{R}_{\check{\boldsymbol{h}}_{jk'}\check{\boldsymbol{h}}_{jk'}}$, which is the covariance of $\check{\boldsymbol{h}}_{jk'}$, and $\boldsymbol{R}_{\boldsymbol{h}_{ljk'}\check{\boldsymbol{h}}_{jk'}}$, which is the cross-covariance of $\boldsymbol{h}_{ljk'}$ and $\check{\boldsymbol{h}}_{jk'}$. The covariance of $\check{\boldsymbol{h}}_{jk'}$ in (3.52) can be written as

$$\begin{aligned}\boldsymbol{R}_{\check{\boldsymbol{h}}_{jk'}\check{\boldsymbol{h}}_{jk'}} = &\sum_{i\in A_1} p_{\mathrm{r}}\beta_{ijk'}K^2\mathbb{E}\left[\boldsymbol{h}_{ijk'}^H\boldsymbol{h}_{ijk'}\right] + \sum_{i\notin A_1}\sum_{k=1}^{K} p_{\mathrm{u}}\beta_{ijk}\mathbb{E}\left[\boldsymbol{h}_{ijk}^H\boldsymbol{d}_{ik}^H\boldsymbol{\psi}_{k'}\boldsymbol{\psi}_{k'}^H\boldsymbol{d}_{ik}\boldsymbol{h}_{ijk}\right] \\ &+ \mathbb{E}\left[\boldsymbol{Z}_j^H\boldsymbol{\psi}_{k'}\boldsymbol{\psi}_{k'}^H\boldsymbol{Z}_j\right]. \end{aligned} \tag{3.54}$$

Since every variable in (3.54) is uncorrelated with each other, then every expectation term in (3.54) will become a scaled identity matrix. As a result, (3.54) simplifies to

$$\begin{aligned}\boldsymbol{R}_{\check{\boldsymbol{h}}_{jk'}\check{\boldsymbol{h}}_{jk'}} &= \left(\sum_{i\in A_1} p_{\mathrm{r}}\beta_{ijk'}K^2 + \sum_{i\notin A_1}\sum_{k=1}^{K} p_{\mathrm{u}}\beta_{ijk}K + K\right)\boldsymbol{I}_M \\ &= \alpha_{jk'}^{\mathrm{TU}}K\boldsymbol{I}_M, \end{aligned} \tag{3.55}$$

where $\alpha_{jk'}^{\mathrm{TU}} = \sum_{i\in A_1} p_{\mathrm{r}}\beta_{ijk'}K + \sum_{i\notin A_1}\sum_{k=1}^{K} p_{\mathrm{u}}\beta_{ijk} + 1$.

The cross-covariance of $\boldsymbol{h}_{ljk'}$ and $\check{\boldsymbol{h}}_{jk'}$ has the same derivation as in (3.15), which gives

$$\begin{aligned}\boldsymbol{R}_{\boldsymbol{h}_{ljk'}\check{\boldsymbol{h}}_{jk'}} =&\mathbb{E}\left[\boldsymbol{h}_{ljk'}^H\check{\boldsymbol{h}}_{jk'}\right] \\ =&\sqrt{p_{\mathrm{r}}\beta_{ljk'}}K\boldsymbol{I}_M. \end{aligned} \tag{3.56}$$

Using (3.5), (3.55) and (3.56), we obtain the LMMSE channel estimation in (3.53).

□

The covariance properties for (3.53) have the same derivation as the TShdown method in (3.16) and (3.17), but with the term $\alpha_{jk'}^{\mathrm{TD}}$ replaced by $\alpha_{jk'}^{\mathrm{TU}}$. Similar to TShdown, we approximate the noise and interference to have Gaussian distributions. Consequently, the distributions for the channel estimate and channel estimation error will be $\hat{\boldsymbol{h}}_{ljk'} \sim \mathcal{CN}\left(0, \frac{p_{\mathrm{r}}\beta_{ljk'}K}{\alpha_{jk'}^{\mathrm{TU}}}\boldsymbol{I}_M\right)$ and $\tilde{\boldsymbol{h}}_{ljk'} \sim \mathcal{CN}\left(0, \left(1 - \frac{p_{\mathrm{r}}\beta_{ljk'}K}{\alpha_{jk'}^{\mathrm{TU}}}\right)\boldsymbol{I}_M\right)$.

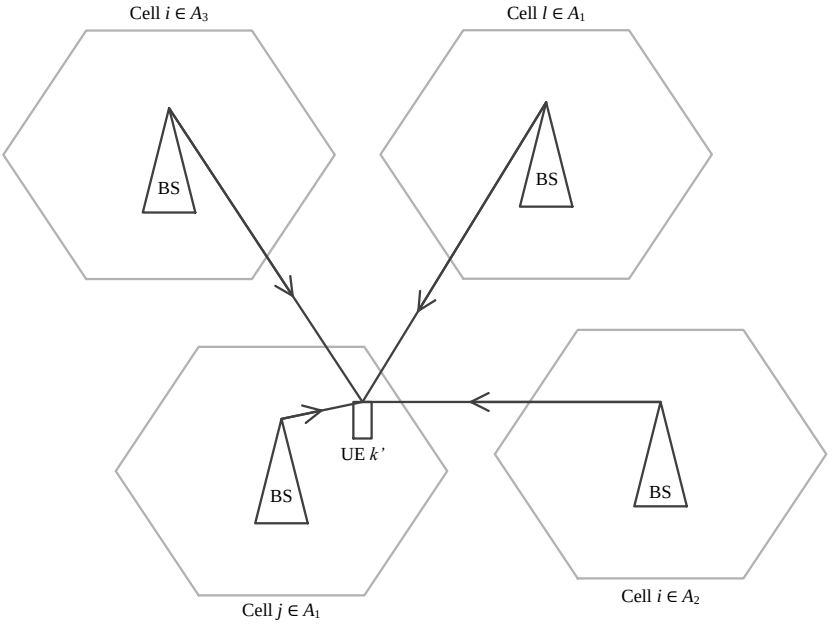

FIGURE 3.8: Received signal at UE k' in cell j during the downlink data transmission phase for group A_1 cells with TShup.

3.4.2 Downlink transmission rate

As shown in Figure 3.6 and Figure 3.8, during the downlink transmission in group A_1, BSs from cells in all groups simultaneously transmit downlink data. Therefore, the received signal at UE k' in cell j of group A_1 will be

$$x_{jk'}^{\mathrm{TU}} = \underbrace{\sum_{i \in A_1} \sum_{k=1}^{K} \sqrt{p_{\mathrm{f}} \beta_{jik'}} \boldsymbol{h}_{jik'} \boldsymbol{a}_{ik} q_{ik}}_{\text{Downlink data from BSs in group } A_1} + \underbrace{\sum_{i \notin A_1} \sum_{k=1}^{K} \sqrt{p_{\mathrm{f}} \beta_{jik'}} \boldsymbol{h}_{jik'} \boldsymbol{a}_{ik} q_{ik}}_{\text{Downlink data from BSs in groups } A_2 \text{ and } A_3} + z_{jk'}. \tag{3.57}$$

The precoding vector $\boldsymbol{a}_{ik}$ is defined in (3.19), but using channel estimation from (3.53).

Theorem 3.6. The closed form lower bound ergodic downlink rate of the TShup method for UE k' in cell j is given by

$$R_{jk'}^{\mathrm{TU,DL}} = \frac{T_{\mathrm{d}}}{T} \times \log_2 \left(1 + \frac{S_{\mathrm{c}}}{N_{\mathrm{c}}} \right), \tag{3.58}$$

where

$$S_{\mathrm{c}} = \frac{p_{\mathrm{f}} p_{\mathrm{r}} \beta_{jjk'}^2 \xi \left(M\right)^2}{\alpha_{jk'}^{\mathrm{TU}}} \tag{3.59}$$

$$N_{\mathrm{c}} = \sum_{l \in A_1} \frac{p_{\mathrm{f}} \beta_{jlk'}}{K} \left(1 + \frac{p_{\mathrm{r}} \beta_{jlk'} K(M-1)}{\alpha_{lk'}^{\mathrm{TU}}}\right) - \frac{p_{\mathrm{f}} p_{\mathrm{r}} \beta_{jjk'}^2 \xi \left(M\right)^2}{\alpha_{jk'}^{\mathrm{TU}}} + \sum_{i \in A_1} \sum_{k \neq k'}^{K} \frac{p_{\mathrm{f}} \beta_{jik'}}{K}$$
$$+ \sum_{f \notin A_1} \sum_{k=1}^{K} \frac{p_{\mathrm{f}} \beta_{jfk'}}{K} \left(1 + \frac{p_{\mathrm{u}} \beta_{jfk'}(M-1)}{\alpha_{fk}^{\mathrm{TU}}}\right) + 1. \tag{3.60}$$

S_{c} is effective downlink signal power for TShup, N_{c} is the total downlink interference and noise power for TShup.

Proof. The lower bound ergodic downlink rate for TShup has the same general form as the TShdown downlink rate in (3.23) which is

$$R_{jk'}^{\mathrm{TU,DL}} = \frac{T_{\mathrm{d}}}{T} \times \log_2 \left(1 + \frac{S_{\mathrm{c}}}{N_{\mathrm{c}}}\right), \tag{3.61}$$

Using the same assumption for downlink rate as in the previous section, we can write

$$S_{\mathrm{c}} = p_{\mathrm{f}} \beta_{jjk'} \left| \mathbb{E}\left[\boldsymbol{h}_{jjk'} \boldsymbol{a}_{jk'}\right] \right|^2. \tag{3.62}$$

The total interference and noise power, N_{c}, can be obtained from the total received power (variance of (3.57)) minus the effective signal power in (3.62), $N_{\mathrm{c}} = \mathbb{E}\left[\left|x_{jk'}^{\mathrm{TU}}\right|^2\right] - S_{\mathrm{c}}$. Similar to the previous section, we expand the interference power terms in order to analyze each of them separately. Therefore,

$$N_{\mathrm{c}} = \underbrace{\sum_{l \in A_1} p_{\mathrm{f}} \beta_{jlk'} \mathbb{E}\left[\left|\boldsymbol{h}_{jlk'} \boldsymbol{a}_{lk'}\right|^2\right]}_{\text{Downlink data from BSs in group } A_1 \text{cell to UE } k'} + \underbrace{\sum_{i \in A_1} \sum_{k \neq k'}^{K} p_{\mathrm{f}} \beta_{jik'} \mathbb{E}\left[\left|\boldsymbol{h}_{jik'} \boldsymbol{a}_{ik}\right|^2\right]}_{\text{Downlink data from BSs in cell } l \in A_1 \text{to UE } k \neq k'}$$
$$+ \underbrace{\sum_{f \notin A_1} \sum_{k=1}^{K} p_{\mathrm{f}} \beta_{jfk'} \mathbb{E}\left[\left|\boldsymbol{h}_{jfk'} \boldsymbol{a}_{fk}\right|^2\right]}_{\text{Downlink data from UEs in cell group } A_2 \text{ and } A_3} + \underbrace{1}_{\text{AWGN noise}} - \underbrace{p_{\mathrm{f}} \beta_{jjk'} \left| \mathbb{E}\left[\boldsymbol{h}_{jjk'} \boldsymbol{a}_{jk'}\right] \right|^2}_{S_{\mathrm{c}}}. \tag{3.63}$$

The expectation term of the effective power in (3.62) has the same derivation procedure as (3.27), but uses the channel estimation result in (3.53). As a result,

$$\mathbb{E}\left[\boldsymbol{h}_{jjk'}\boldsymbol{a}_{jk'}\right] = \sqrt{\frac{p_{\mathrm{r}}\beta_{jjk'}}{\alpha_{jk'}^{\mathrm{TU}}}}\xi\left(M\right). \tag{3.64}$$

where $\xi(M)$ defined in (3.27).

The first expectation term in (3.63) involves the signal power from the BS to UEs of the same group and same pilot sequence, which has the similar derivation as (3.30), but uses the channel estimation result in (3.53). Therefore,

$$\mathbb{E}\left[\left|\boldsymbol{h}_{jlk'}\boldsymbol{a}_{lk'}\right|^2\right] = \left(1 + \frac{p_{\mathrm{r}}\beta_{jlk'}K(M-1)}{\alpha_{lk'}^{\mathrm{TU}}}\right)\frac{1}{K}. \tag{3.65}$$

The second expectation term in (3.63) involves the signal power from the BS to UEs of the same group with different pilot sequences. Since the precoding and the channel is uncorrelated, this means

$$\mathbb{E}\left[\left|\boldsymbol{h}_{jik'}\boldsymbol{a}_{ik}\right|^2\right] = \frac{1}{K}. \tag{3.66}$$

For the third expectation term in (3.63), which involves signals from different cell groups, the precoding vector, $\boldsymbol{a}_{fk}$, is correlated with the channel from different cell groups, $\boldsymbol{h}_{jfk'}$. This means that uplink data to pilot interference will also contribute to the contamination effect. Consequently, the interference has a similar derivation as the expression for pilot contamination in (3.65). However, since the uplink data, $\boldsymbol{d}_{ik}$, and pilot, $\boldsymbol{\psi}_{k'}$, are uncorrelated, this means that the expectation term will not have exactly the same expression as the pilot to pilot interference in (3.65). This is because the expected power of the correlation between pilot and the received uplink data signal is $\mathbb{E}\left[\left|\sqrt{p_{\mathrm{u}}\beta_{jfk}}\boldsymbol{\psi}_{k'}^H\boldsymbol{d}_{ik}\right|^2\right] = p_{\mathrm{u}}\beta_{jfk}K$ while the expected power of the correlation between pilot and the received pilot of the same sequence is $\mathbb{E}\left[\left|\sqrt{p_{\mathrm{r}}\beta_{jlk'}}\boldsymbol{\psi}_{k'}^H\boldsymbol{\psi}_{k'}\right|^2\right] = p_{\mathrm{r}}\beta_{jlk'}K^2$. Based on this power difference, we replace the variables related to received pilot power, namely $p_{\mathrm{r}}\beta_{jlk'}K$ in expectation term (3.65) with variables related to received uplink data power, namely $p_{\mathrm{u}}\beta_{jfk'}$, in order to get the expectation for different cell groups

interference as

$$\mathbb{E}\left[\left|\boldsymbol{h}_{jfk'}\boldsymbol{a}_{fk}\right|^2\right] = \left(1 + \frac{p_{\mathrm{u}}\beta_{jfk'}(M-1)}{\alpha_{fk}^{\mathrm{TU}}}\right)\frac{1}{K}. \tag{3.67}$$

Substituting (3.64)-(3.67) into (3.62) and (3.63), the ergodic achievable downlink rate for UE k' in cell j in (3.61) becomes (3.58).

□

The inter-cell interference behaves differently in TShup (3.58) compared to TShdown (3.20). This is because in TShup, both interference from the same and different cell groups increase linearly with the number of antennas. For interference from different cell groups, this effect can occur for all UEs rather than only UEs with the same pilot sequences.

3.4.3 Uplink transmission rate

Referring to Figure 3.6 and Figure 3.9, during uplink data transmission in cells from group A_1, the BS in group A_1 also receives uplink pilots from UEs in group A_2 and uplink data from group A_3. If $\boldsymbol{y}_j$ is the signal received by the cell j BS in group A_1, then the estimated data is $\hat{d}_{jk'}^{\mathrm{TU}} = \boldsymbol{y}_j\boldsymbol{b}_{jk'}$. This can be expanded as

$$\begin{aligned}\hat{d}_{jk'}^{\mathrm{TU}} = &\underbrace{\sum_{i\in A_1}\sum_{k=1}^{K}\sqrt{p_{\mathrm{u}}\beta_{ijk}}d_{ik}\boldsymbol{h}_{ijk}}_{\text{Uplink data from UEs in group } A_1}\boldsymbol{b}_{jk'} + \underbrace{\sum_{i\in A_2}\sum_{k=1}^{K}\sqrt{p_{\mathrm{r}}\beta_{ijk}}\psi_k\boldsymbol{h}_{ijk}}_{\text{Uplink pilot from UEs in group } A_2}\boldsymbol{b}_{jk'} \\ &+ \underbrace{\sum_{i\in A_3}\sum_{k=1}^{K}\sqrt{p_{\mathrm{u}}\beta_{ijk}}d_{ik}\boldsymbol{h}_{ijk}}_{\text{Uplink data from UEs in group } A_3}\boldsymbol{b}_{jk'} + \boldsymbol{z}_j\boldsymbol{b}_{jk'},\end{aligned} \tag{3.68}$$

where $\boldsymbol{b}_{jk'}$ is the equalizer defined in (3.34), but using channel estimation from (3.53).

Theorem 3.7. The closed form lower bound ergodic uplink rate of the TShup method for UE k' in cell j is given by

$$\tilde{R}_{jk'}^{\mathrm{TU,UL}} = \frac{T_{\mathrm{u}}}{T} \times \log_2\left(1 + \frac{\tilde{S}_{\mathrm{d}}}{\tilde{N}_{\mathrm{d}}}\right), \tag{3.69}$$

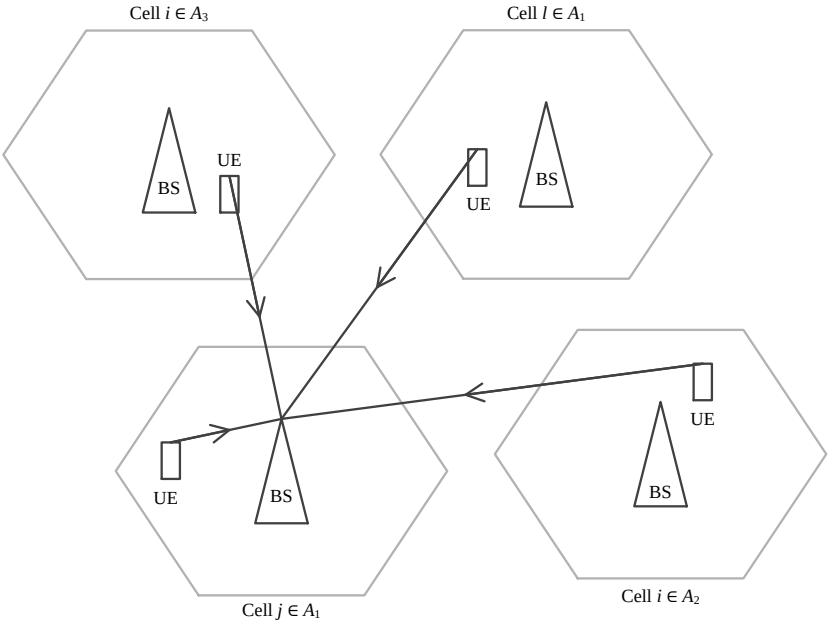

FIGURE 3.9: Received signal at BS in cell j during the uplink data transmission phase for group A_1 cells with TShup.

where

$$\tilde{S}_{\mathrm{d}} = \frac{p_{\mathrm{u}} p_{\mathrm{r}} \beta_{jjk'}^2 K(M-1)}{\alpha_{jk'}^{\mathrm{TU}}} \tag{3.70}$$

$$\begin{aligned}\tilde{N}_{\mathrm{d}} =& p_{\mathrm{u}} \beta_{jjk'} \left(1 - \frac{p_{\mathrm{r}} \beta_{jjk'} K}{\alpha_{jk'}^{\mathrm{TU}}}\right) + \sum_{l \in A_1 \cap l \neq j} p_{\mathrm{u}} \beta_{ljk'} \left(1 + \frac{p_{\mathrm{r}} \beta_{ljk'} K(M-2)}{\alpha_{jk'}^{\mathrm{TU}}}\right) \\ &+ \sum_{i \in A_1} \sum_{k \neq k'}^{K} p_{\mathrm{u}} \beta_{ijk} + \sum_{f \in A_2} \sum_{k=1}^{K} p_{\mathrm{r}} \beta_{fjk} \left(1 + \frac{p_{\mathrm{u}} \beta_{fjk}(M-2)}{\alpha_{jk'}^{\mathrm{TU}}}\right) \\ &+ \sum_{f \in A_3} \sum_{k=1}^{K} p_{\mathrm{u}} \beta_{fjk} \left(1 + \frac{p_{\mathrm{u}} \beta_{fjk}(M-2)}{\alpha_{jk'}^{\mathrm{TU}}}\right) + 1. \end{aligned} \tag{3.71}$$

$\tilde{S}_{\mathrm{d}}$ is effective uplink signal power for TShup, $\tilde{N}_{\mathrm{b}}$ is the total uplink interference and noise power for TShup.

Proof. The expression for the lower bound ergodic uplink rate for the TShup method has the same general form as for the TShdown method in (3.39), which is

$$\tilde{R}_{jk'}^{\mathrm{TU,UL}} = \frac{T_{\mathrm{u}}}{T} \times \log_2 \left(1 + \left(\mathbb{E}\left[\frac{N_{\mathrm{d}}}{S_{\mathrm{d}}}\right]\right)^{-1}\right), \tag{3.72}$$

The effective signal power, S_{d}, is given by

$$S_{\mathrm{d}} = p_{\mathrm{u}}\beta_{jjk'}\left|\hat{\boldsymbol{h}}_{jjk'}\boldsymbol{b}_{jk'}\right|^2. \tag{3.73}$$

The total interference and noise power, N_{d}, can be obtained from the total power of estimated data (3.68) minus the effective signal power (3.73), $N_{\mathrm{d}} = \left|\hat{d}_{jk}^{\mathrm{TU}}\right|^2 - S_{\mathrm{d}}$. Similar to the previous section, we expand N_{d} into uncorrelated interference terms to get

$$\begin{aligned}\mathbb{E}\left[\frac{N_{\mathrm{d}}}{S_{\mathrm{d}}}\right] = & \underbrace{p_{\mathrm{u}}\beta_{jjk'}\mathbb{E}\left[\frac{\left|\tilde{\boldsymbol{h}}_{jjk'}\boldsymbol{b}_{jk'}\right|^2}{S_{\mathrm{d}}}\right]}_{\text{Channel estimation error for UE } k' \text{in cell } j} + \underbrace{\sum_{l\in A_1\cap l\neq j} p_{\mathrm{u}}\beta_{ljk'}\mathbb{E}\left[\frac{\left|\boldsymbol{h}_{ljk'}\boldsymbol{b}_{jk'}\right|^2}{S_{\mathrm{d}}}\right]}_{\text{Uplink data from UE } k' \text{in cell } l\in A_1\cap l\neq j} \\ & + \underbrace{\sum_{i\in A_1}\sum_{k\neq k'}^{K} p_{\mathrm{u}}\beta_{ijk}\mathbb{E}\left[\frac{\left|\boldsymbol{h}_{ijk}\boldsymbol{b}_{jk'}\right|^2}{S_{\mathrm{d}}}\right]}_{\text{Uplink data from UE } k\neq k' \text{in cell } i\in A_1} + \underbrace{\sum_{f\in A_2}\sum_{k=1}^{K} p_{\mathrm{r}}\beta_{fjk}\mathbb{E}\left[\frac{\left|\boldsymbol{h}_{fjk}\boldsymbol{b}_{jk'}\right|^2}{S_{\mathrm{d}}}\right]}_{\text{Uplink pilot from UEs in cell group } A_2} \\ & + \underbrace{\sum_{f\in A_3}\sum_{k=1}^{K} p_{\mathrm{u}}\beta_{fjk}\mathbb{E}\left[\frac{\left|\boldsymbol{h}_{fjk}\boldsymbol{b}_{jk'}\right|^2}{S_{\mathrm{d}}}\right]}_{\text{Uplink data from UEs in cell group } A_3} + \underbrace{\mathbb{E}\left[\frac{\left|\boldsymbol{z}_j\boldsymbol{b}_{jk'}\right|^2}{S_{\mathrm{d}}}\right]}_{\text{AWGN noise}}.\end{aligned} \tag{3.74}$$

Similar to (3.41), the channel estimation error term in (3.74) is obtained from $\mathbb{E}\left[\frac{\left|\tilde{\boldsymbol{h}}_{jjk'}\boldsymbol{b}_{jk'}\right|^2}{S_{\mathrm{d}}}\right] = \mathbb{E}\left[\frac{\left|\boldsymbol{h}_{jjk'}\boldsymbol{b}_{jk'}\right|^2}{S_{\mathrm{d}}}\right] - \mathbb{E}\left[\frac{S_{\mathrm{d}}}{S_{\mathrm{d}}}\right]$.

The first expectation term in (3.74) involves channel estimation error, which has the same derivation as in (3.45), but uses the channel estimate of (3.53). As a result,

$$\mathbb{E}\left[\frac{\left|\tilde{\boldsymbol{h}}_{jjk}\boldsymbol{b}_{jk'}\right|^2}{S_{\mathrm{d}}}\right] = \left(1 - \frac{p_{\mathrm{r}}\beta_{jjk'}K}{\alpha_{jk'}^{\mathrm{TU}}}\right)\frac{1}{\lambda^{\mathrm{TU}}}, \tag{3.75}$$

where λ^{TU} in (3.75) has a similar expression to λ^{TD} in (3.43), but $\alpha_{jk'}^{\mathrm{TD}}$ is replaced by $\alpha_{jk'}^{\mathrm{TU}}$. Therefore, $\lambda^{\mathrm{TU}} = \frac{p_{\mathrm{u}}p_{\mathrm{r}}\beta_{jjk'}^2K(M-1)}{\alpha_{jk'}^{\mathrm{TU}}}$.

The second expectation term in (3.74) involves the signal power received from UEs of the same group and the same pilot sequence. Since this term is related to pilot contamination, it has a similar derivation as in (3.48), but uses the channel estimate of

(3.53). Therefore,

$$\mathbb{E}\left[\frac{\left|\boldsymbol{h}_{ljk'}\boldsymbol{b}_{jk'}\right|^2}{S_{\mathrm{d}}}\right] = \left(1+\frac{p_{\mathrm{r}}\beta_{ljk'}K(M-2)}{\alpha_{jk'}^{\mathrm{TU}}}\right)\frac{1}{\lambda^{\mathrm{TU}}}. \tag{3.76}$$

The third expectation term in (3.74) involves signal power from UEs of the same group but with different pilot sequences. Since the vector and equalizer are uncorrelated, this means

$$\mathbb{E}\left[\frac{\left|\boldsymbol{h}_{ijk}\boldsymbol{b}_{jk'}\right|^2}{S_{\mathrm{d}}}\right] = \frac{1}{\lambda^{\mathrm{TU}}}. \tag{3.77}$$

Note that (3.77) also applies to the AWGN noise term.

The expectation for different cell group interference (fourth and fifth terms in (3.74)) has a contamination effect, which means it has a similar expression to (3.76). However, since this contamination is caused by data instead of pilots, we need to replace the variables related to received pilot power, namely $p_{\mathrm{r}}\beta_{ljk'}K$ in expectation term (3.76), with variables related to received uplink data power, namely $p_{\mathrm{u}}\beta_{fjk}$ (similar strategy used to obtain (3.67)). As a result we get

$$\mathbb{E}\left[\frac{\left|\boldsymbol{h}_{fjk}\boldsymbol{b}_{jk'}\right|^2}{S_{\mathrm{d}}}\right] = \left(1+\frac{p_{\mathrm{u}}\beta_{fjk}(M-2)}{\alpha_{jk'}^{\mathrm{TU}}}\right)\frac{1}{\lambda^{\mathrm{TU}}}. \tag{3.78}$$

Substituting (3.75)-(3.78) into (3.74), the ergodic achievable uplink rate for UE k' in cell j in (3.72) becomes (3.69).

□

The uplink rate for TShup (3.69) and TShdown (3.35) are quite different because in TShup interference from the same and different cell groups both increase linearly with the number of antennas. For different cell group interference, this affects all UEs rather than for UEs with the same pilot sequence.

3.4.4 Power optimization

Similar to the TShdown method in Subsection 3.3.4, we can derive an optimum power allocation for TShup as well. However, for TShup, since the pilot is overlapped with the

uplink data (refer to Figure 3.6), optimization is done by finding the uplink transmit power, p_{u}, that maximizes the uplink sum rate. An expression for optimum transmit power is approximated by differentiating (3.58) with respect to p_u, and solving for p_{u} in $\frac{d}{dp_{\mathrm{u}}}\left(R_{jk'}^{\mathrm{TU,UL}}\right) = 0$ with $p_{\mathrm{u}} > 0$ (see Appendix D.2 for more detail). The result is given as

$$p_{\mathrm{u}} = \sqrt{\frac{p_{\mathrm{r}}^2 K \left(\sum_{f\in A_2}\sum_{k=1}^{K}\beta_{fjk}\right)\left(\sum_{i\in A_1}\beta_{ijk'}\right) + p_{\mathrm{r}}\left(\sum_{f\in A_2}\sum_{k=1}^{K}\beta_{fjk} + K\sum_{i\in A_1}\beta_{ijk'}\right)}{\left(\sum_{i\in A_1\cup A_3}\sum_{k=1}^{K}\beta_{ijk}\right)\left(\sum_{i\notin A_1}\sum_{k=1}^{K}\beta_{ijk}\right) + \sum_{i\in A_3}\sum_{k=1}^{K}\beta_{ijk}^2\left(M-1\right)}}. \tag{3.79}$$

Similar to the TShdown method, the optimum power that we obtained in (3.79) is only optimized for UE k' in cell j. In order to obtain a common value for p_u across all UEs in all cells, we average the value of (3.79) across multiple UEs and cells. Since the value generated from (3.79) resembles a log-normal distribution (as shown in Figure 3.15 in the results section), geometric mean is used to find the average.

3.5 Time-synchronized pilot

To analyze the performance of the time-synchronized pilot (TSync) method, we can use the same rate equation as either of the time-shifted methods, but we have to modify the system so that all cells only belong to one cell group. This is because if all cells in the time-shifted method belong to the same group, such as in Figure 3.10, then the pilot transmission in all cells will be time-synchronized. In other words, the TSync method is a special case of the time-shifted methods. For example, we can set all cells in TShup to be in group A_1. As a result, the pilot and data transmission phase is arranged as in Figure 3.11. In this case, we need to replace all interference terms from cell groups A_2 and A_3 in the downlink rate equation (3.58) and uplink rate equation (3.69) with interference terms in group A_1 in order to obtain the downlink and uplink rates for TSync. We also need to replace $\alpha_{jk'}^{\mathrm{TU}}$ with $\alpha_{jk'}^{\mathrm{TC}} = \sum_{i\in A_1} p_{\mathrm{r}}\beta_{ilk'}K + 1$ for the TSync method. We use TC as the abbreviation of the variables related to the time-synchronized method. Therefore, the ergodic downlink rate for TSync can be expressed as

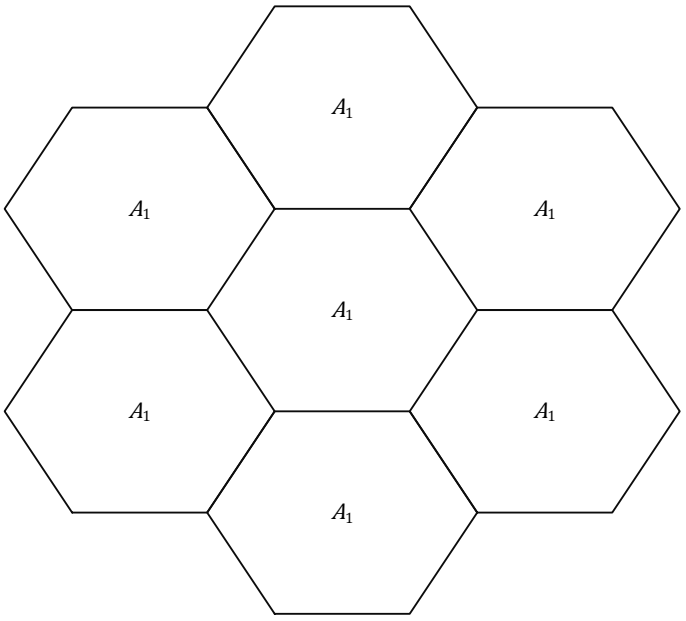

FIGURE 3.10: 7 cells group arrangement for TSync.

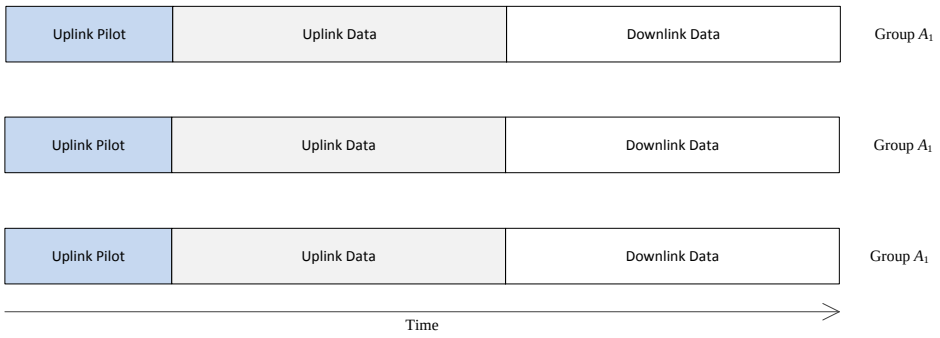

FIGURE 3.11: The arrangement of time-shifted pilots for TSync.

$$R_{jk'}^{\mathrm{TC,DL}} = \frac{T_{\mathrm{d}}}{T} \times \log_2\left(1 + \frac{S_{\mathrm{e}}}{N_{\mathrm{e}}}\right), \tag{3.80}$$

where

$$S_{\mathrm{e}} = \frac{p_{\mathrm{f}} p_{\mathrm{r}} \beta_{jjk'}^2 \xi \left(M\right)^2}{\alpha_{jk'}^{\mathrm{TC}}}, \tag{3.81}$$

$$N_{\mathrm{e}} = \sum_{l \in A_1} \frac{p_{\mathrm{f}} \beta_{jlk'}}{K} \left(1 + \frac{p_{\mathrm{r}} \beta_{jlk'} K(M-1)}{\alpha_{lk'}^{\mathrm{TC}}}\right) - \frac{p_{\mathrm{f}} p_{\mathrm{r}} \beta_{jjk'}^2 \xi \left(M\right)^2}{\alpha_{jk'}^{\mathrm{TC}}} + \sum_{i \in A_1} \sum_{k \neq k'}^{K} \frac{p_{\mathrm{f}} \beta_{jik'}}{K} + 1. \tag{3.82}$$

S_{e} is effective downlink signal power for TSync, N_{e} is the total downlink interference and noise power for TSync.

The ergodic uplink rate for TSync can be expressed as

$$\tilde{R}_{jk'}^{\mathrm{TC,UL}} = \frac{T_{\mathrm{u}}}{T} \times \log_2\left(1 + \frac{\tilde{S}_{\mathrm{f}}}{\tilde{N}_{\mathrm{f}}}\right), \tag{3.83}$$

where

$$\tilde{S}_{\mathrm{f}} = \frac{p_{\mathrm{u}} p_{\mathrm{r}} \beta_{jjk'}^2 K(M-1)}{\alpha_{jk'}^{\mathrm{TC}}}, \tag{3.84}$$

$$\tilde{N}_{\mathrm{f}} = p_{\mathrm{u}}\beta_{jjk'}\left(1 - \frac{p_{\mathrm{r}}\beta_{jjk'}K}{\alpha_{jk'}^{\mathrm{TC}}}\right) + \sum_{l \in A_1 \cap l \neq j} p_{\mathrm{u}}\beta_{ljk'}\left(1 + \frac{p_{\mathrm{r}}\beta_{ljk'}K(M-2)}{\alpha_{jk'}^{\mathrm{TC}}}\right) + \sum_{i \in A_1}\sum_{k \neq k'}^{K} p_{\mathrm{u}}\beta_{ijk} + 1. \tag{3.85}$$

$\tilde{S}_{\mathrm{f}}$ is effective uplink signal power for TSync, $\tilde{N}_{\mathrm{f}}$ is the total uplink interference and noise power for TSync.

3.6 Numerical results

We use the cell arrangement in Figure 3.1 for performance analysis of both time-shifted methods. In this arrangement, there will be no same group interference from the nearest cells. This is similar to the optimum arrangement for transmission with a frequency reuse of 3. We also assume that the interference outside the 7 cells has a negligible impact on the middle cell. For TSync, all cells belong to the same group, as shown in Figure 3.10. With these settings, each transmission method will experience different types of interference from the nearest cells during pilot transmission. First, we compare the SINR performance of Monte Carlo simulation with the derived closed form for both time-shifted methods. SINR is defined as the effective signal power divided with effective interference and noise power. For example, for TShdown downlink, the SINR is equal to S_{a} in (3.21) divided by N_{a} in (3.22). For Monte Carlo simulation, we use Zadoff-Chu sequences for the pilot [10], QPSK modulation for data and model small-scale fading channel using a random complex Gaussian variable. For every small scale channel realization, we obtain the channel estimate, precoding and equalizer values. We insert these values in the expectation term in (3.24), (3.25), (3.40) and (3.41) and find the average values over all channel realizations for each expectation term in order to obtain downlink and uplink TShdown SINR. A similar method can be applied to the expectation

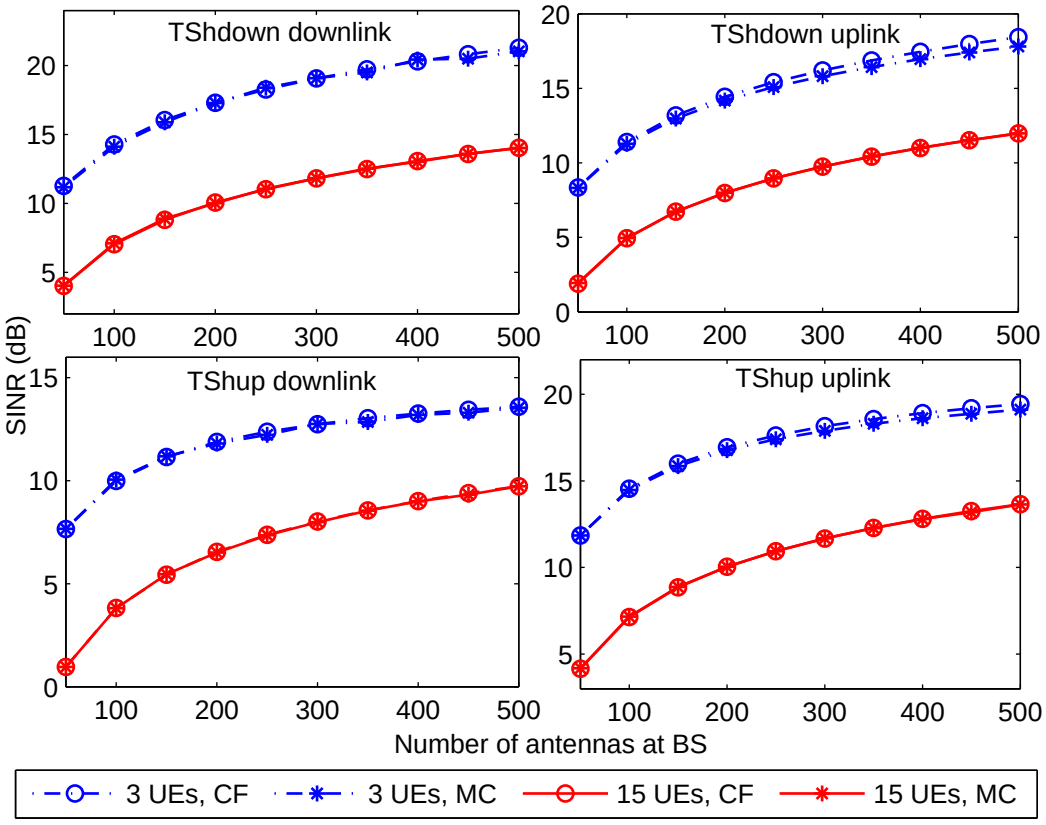

FIGURE 3.12: Downlink and uplink SINR for the TShdown and TShup method using Monte Carlo (MC) and closed form (CF) simulations.

terms in (3.62), (3.63), (3.73) and (3.74) to obtain downlink and uplink TShup SINR. Then we compare the Monte Carlo result with the derived closed form expressions. The transmit powers for all transmission methods are $p_{\mathrm{f}} = 20$ dB, $p_{\mathrm{u}} = 10$ dB and $p_{\mathrm{r}} = 10$ dB. The transmissions are tested with coherent subcarrier values of $N = 1$ and $N = 5$. Since $K = \tau N$, if we set $\tau = 3$, this means the number of spatial multiplexed UEs will be 3 and 15 for each test. Having more coherent subcarriers enables the pilot length to be longer, which means more parallel data streams or more UEs can be included in spatial multiplexing transmission. We set all the large-scale fading values using the path-loss equation $\beta = zd^{-\gamma}$, where z is log-normal shadow fading, γ is the path-loss exponent and d is the distance between transmitter and receiver, normalized to the inner radius of the cell. The path-loss exponent is set to 3.8. We assume the UEs are the same distance from the BS, namely 2/3 of the cell radius and without shadowing (or $z = 1$) following the scenario in [6]. The results in Figure 3.12 show that derived closed form expressions are a very close approximation to the Monte Carlo simulations. The results also show that SINR for 3 UEs is higher than that of 15 UEs. However, 15 UEs will have a higher spatial multiplexing gain resulting in a higher sum rate, which we will see next.

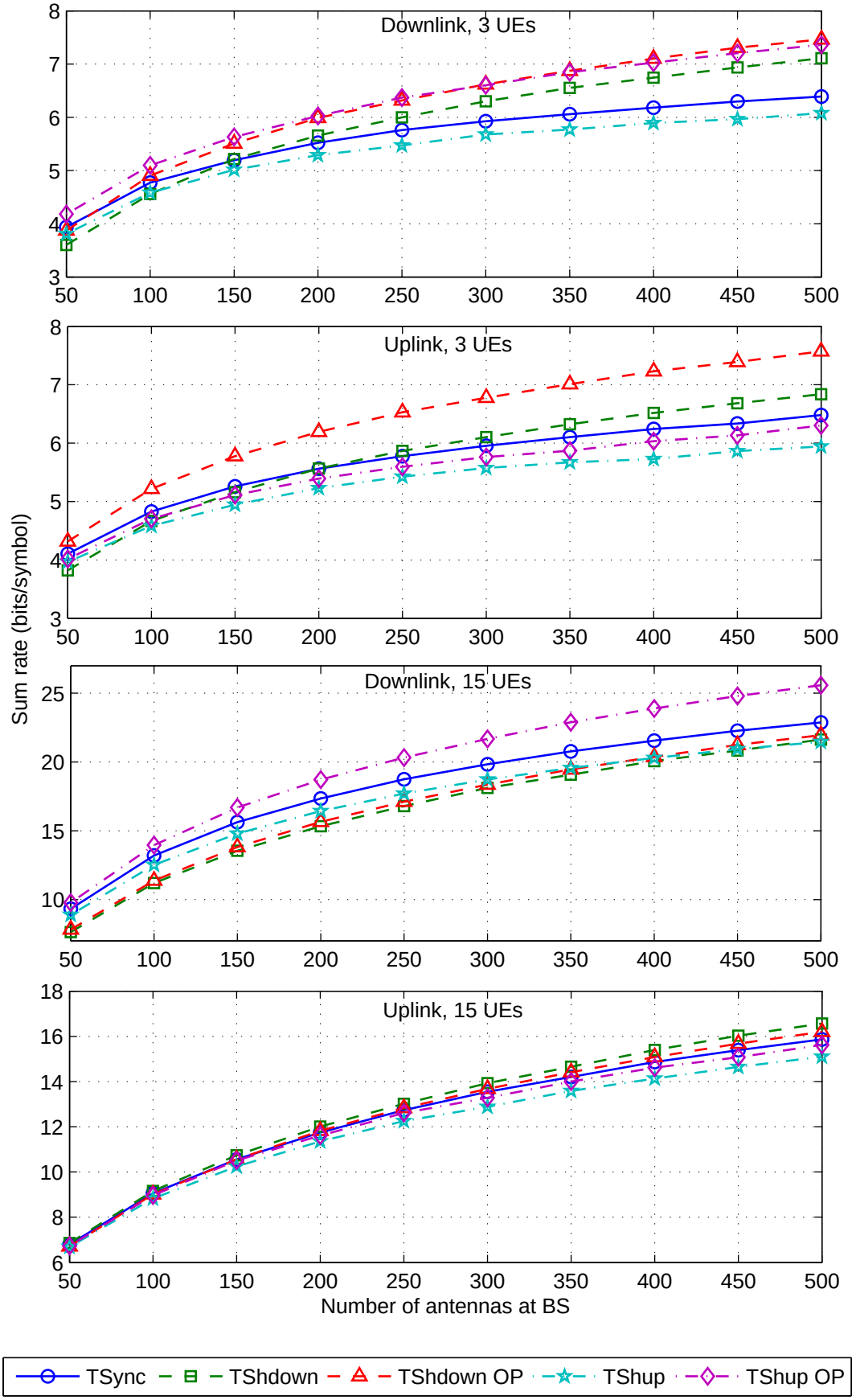

FIGURE 3.13: Average downlink and uplink sum rate for TSync, TShdown and TShup and without power optimization (OP).

Now we proceed using the closed form expressions to obtain sum rate performance. In this simulation, the UEs are located at uniform random locations within each cell excluding the area inside a 0.1 normalized distance from the BSs. We also include the shadowing effect with standard deviation of 8dB [5]. Each UE will be located in a cell which has the largest large-scale fading between the UE and the BS in that cell. This is because UE cell allocation will be based on power level between the UE and the BS [11]. For TShdown, we find the downlink and uplink rates for each UE using (3.20) and (3.35). For TShup, we find the downlink and uplink rates for each UE using (3.58) and (3.69). For TSync, we find the downlink and uplink rate for each UE using (3.80) and (3.83). After obtaining the transmission rate for each UE, the sum rate is obtained by summing the transmission rate for all UEs in one cell. Then the average sum rate is obtained by averaging the sum rate across different random locations of the UEs. All the transmission methods have $\tau = 3$, $T_{\mathrm{d}} = 6$, $T_{\mathrm{u}} = 6$, and $T = 15$. As well as using the same transmit power, we also use optimized power to maximize the sum rate. These optimum powers are found using an exhaustive search. We use 3 UEs and 15 UEs for each test.

The average sum rate per cell is shown in Figure 3.13. The TShdown method can improve both uplink and downlink rates compared to the TSync method when we use 3 UEs and can be further improved by using optimized power. However, for 15 UEs, TShdown performs poorly in the downlink. This is because, for high number of UEs, there will be more pilot interference signals from other cells received by the UEs during the downlink transmission. For TShup, the downlink rate is significantly improved compared to TSync for 3 UEs and 15 UEs when the power is optimized. The uplink rate is slightly lower than TSync. Note that if we apply power optimization, then the transmit data power from UEs in TShup will be between 3 dB and 7 dB while the transmit data power from BS in TShdown will be between 14 dB and 22 dB. Therefore, it is possible that the power optimization can also reduce the power consumption, especially for TShup.

Next, to get a more complete understanding of the behavior of time-shifted methods, we consider two more types of path-loss model in addition to the path-loss model used for Figure 3.13. The additional two path-loss model are random UE locations without shadowing and equal UE distance from the BS without shadowing [6]. The results in Table 3.1 show the percentage changes in sum rate for the time-shifted methods with respect to TSync using 200 antennas for the 3 different path-loss models. Note that the

TABLE 3.1: Percentage change in the average sum rate with respect to TSync for 200 antennas

Pathloss model	Number of UEs	TShdown		TShup	
		DL	UL	DL	UL
Random UE locations with shadowing (Figure 3.13)	3	9%	11%	9%	-3%
	15	-10%	1%	8%	-1%
Random UE locations without shadowing	3	17%	15%	12%	1%
	15	0%	4%	9%	2%
Equal UE distances from BS [6]	3	6%	9%	6%	2%
	15	2%	4%	5%	- 1%

results in Table 3.1 are in comparison to TSync in each specific condition and both of the time-shifted methods use power optimization. A positive percentage means a sum rate increase while a negative percentage means a sum rate decrease with respect to TSync. The results show that for all 3 path-loss models, TShdown generally performs the best for 3 UEs. However, TShdown has less advantage in the downlink rate for 15 UEs. This is because, as the number of UEs increases, the interference from uplink pilots from other cells will also increase. For TShup, increasing the number of UEs does not significantly impact its downlink performance with respect to TSync. This because, unlike TShdown, TShup does not have UE to UE interference since all the UEs perform either uplink or downlink simultaneously. This is an important advantage of TShup because practically the demand for the downlink is usually higher than on the uplink. In other words, TShdown is the preferable method when there are small numbers of UEs, but TShup has an advantage when there are higher numbers of UEs. This means that in addition to being more practical than TShdown in term of downlink and uplink phase change (see the explanation in Section 3.4 introduction), TShup has a more consistent downlink rate improvement with respect to TSync when the number of UEs changes. If we include shadowing effects, TShdown has a significantly worse performance than TSync in the downlink for 15 UEs. This is because UE to UE interference in the TShdown downlink can be very large due to the log-normal distribution of the large-scale fading from the shadowing effect. In other words, under certain conditions, it is not accurate to assume that TSync is the worst case scenario for transmission. Compared to the random UE location model, the equal UE distance model for TShdown has less significant performance changes between 3 UEs and 15 UEs with respect to TSync. This is because the random UE location model gives more cell edge interference. This significantly increases UE to UE interference in TShdown when the number of UEs

increases. Meanwhile, the equal UE distance scenario has less cell edge interference due to the position of the UEs. Hence, the changes in the number of UEs has less impact on TShdown performance for this model.

The results in Figure 3.13 are based on power optimization using an exhaustive search. As we discussed in Section 3.3.4 and Section 3.4.4, we can also obtain the power using the optimal power equations. Due to the variation of the UEs large scale fading, the optimum power equation in (3.50) and (3.79) will follow certain distribution, as shown in Figure 3.14 and 3.15. Since the optimum power distribution in these figures are based on a logarithmic scale, we use geometric mean to find the average optimum power for each method. Using the acquired optimum power, we can obtain the sum rate results as given in Figure 3.16. The results show that the sum rate of an optimized power exhaustive search agrees well with the equation-based optimum power.

3.7 Summary

We had derived novel closed form downlink and uplink transmission rate expressions for TShdown and TShup with a finite number of BS antennas. Results showed that if there were a small number of UEs, the TShdown method with power optimization provided the best sum rate performance. However, for high numbers of UEs, the TShup method with power optimization had an advantage compared to other methods. The time shifted methods do not always outperform TSync when there is no power optimization. As discussed in Section 3.3.4, data power optimization is a unique advantage that time shifted methods have over TSync due to the fact that there is no pilot and data overlap in TSync. In addition to being more practical than TShdown in terms of downlink and uplink phase change, the TShup performance improvement with respect to TSync was also more consistent than TShdown when the number of UEs increased. The results also showed that the TSync method is not necessarily the worst case scenario in terms of sum rate performance, especially when path-loss models with shadowing are considered.

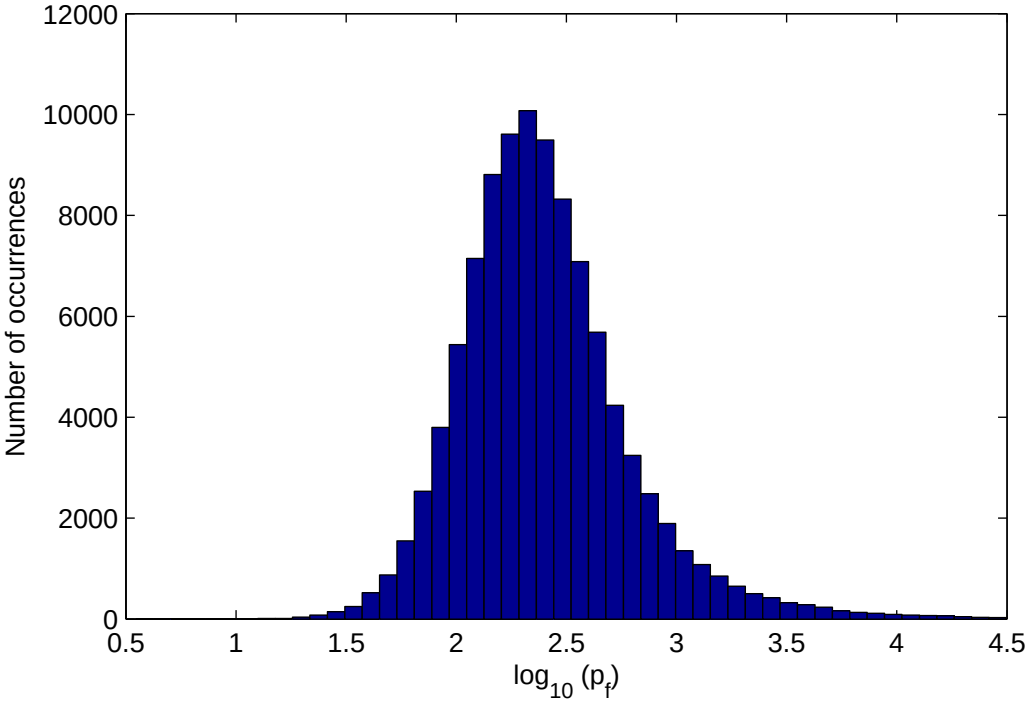

FIGURE 3.14: Distribution for optimum downlink data power for TShdown generated using (3.50) with 100 BS antennas, 15 UEs and 10^4 random UEs location drops.

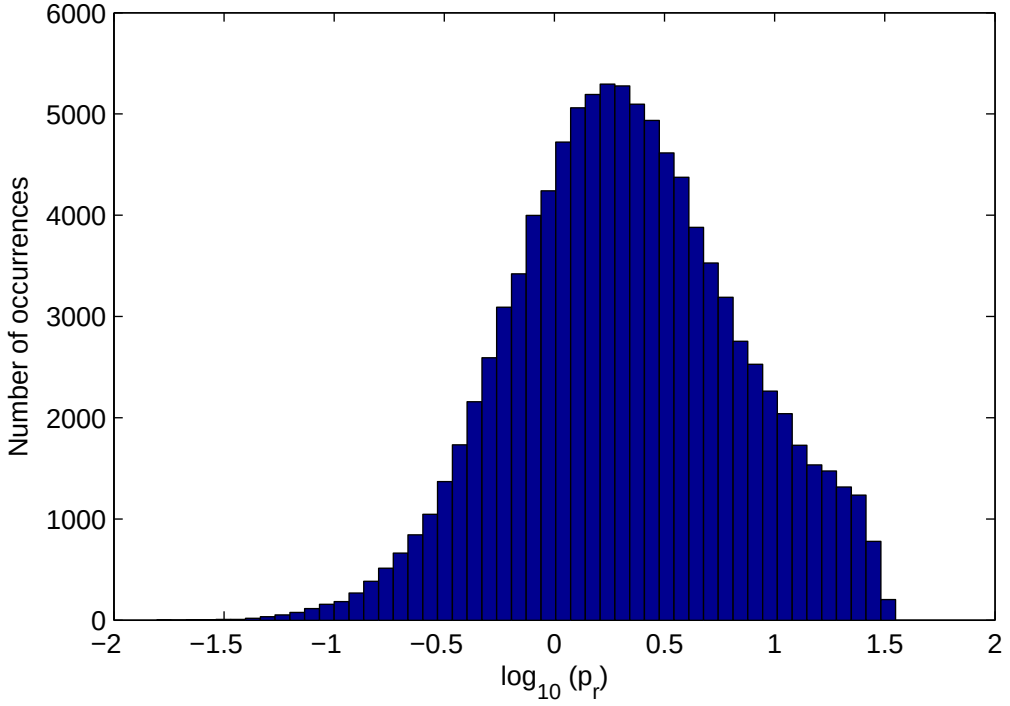

FIGURE 3.15: Distribution for optimum uplink data power for TShup generated using (3.79) with 100 BS antennas, 15 UEs and 10^4 random UEs location drops.

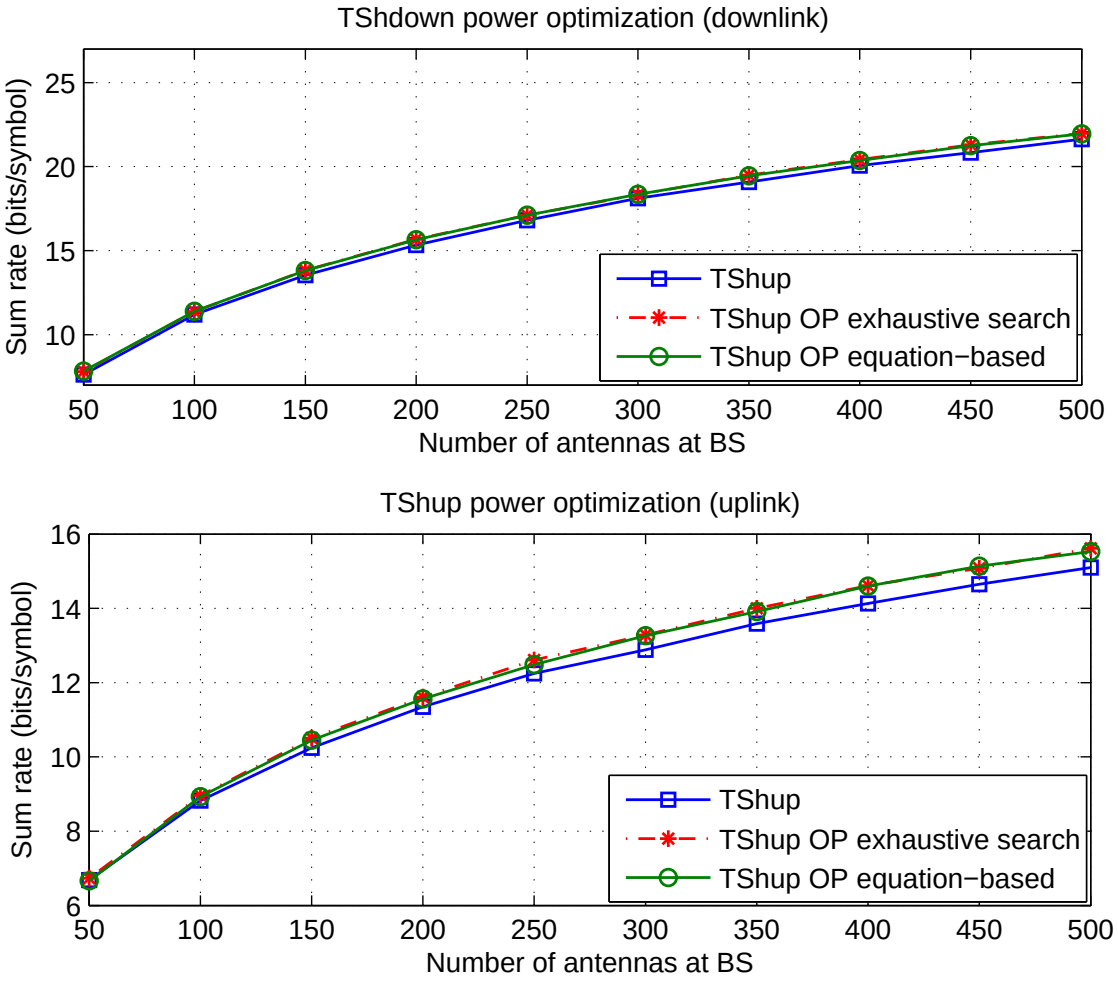

FIGURE 3.16: Power optimization (OP) using exhaustive search and equation-based methods for TShdown and TShup for 15 UEs.

Chapter 4

Massive MIMO systems in time-selective channels

4.1 Introduction

As we discussed in Chapter 2, the channel may vary over time due the UE mobility [29]. The block-fading channel model is a slow fading approximation, which assumes the channel to be constant within a coherence block and to vary randomly between blocks. Research in massive MIMO systems commonly uses the block fading model due to its simplicity [8]. Unlike the block-fading model, selective fading channel model allows the channel to vary within the coherence block and thus provides a more realistic representation of channel variation. We have shown in Figure 2.3 that there can be significant variation over time between the block fading model and time-selective model, especially for faster fading (higher speed UE mobility). Hence, this chapter investigates the performance of massive MIMO systems in time-selective channels, where the channel conditions can vary within the frame. The key challenge to analyze selective fading performance is the calculation complexity. This is because there is channel variation within the frame in selective fading while there is no channel variation within the frame of block fading model. Therefore, more complicated mathematical expression is required to obtain the sum rate for selective fading model compared to block fading model.

Various works have studied the optimization of training and data length in time-selective channels [55–58, 60]. Such optimization depends on the type of performance criteria

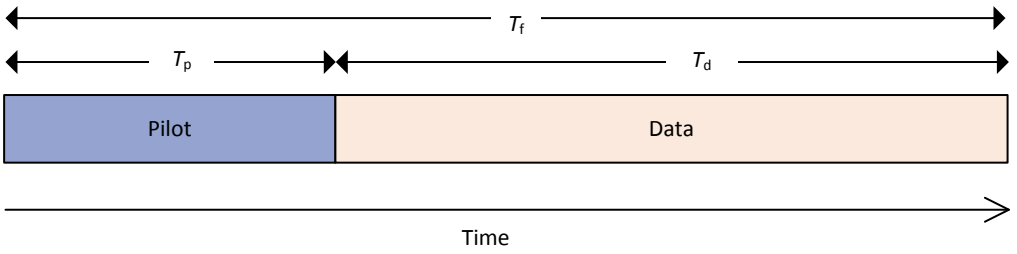

FIGURE 4.1: Pilot and data arrangement in one frame.

used. For example, bit error rate (BER) is used in [55, 56] as a performance metric. In this chapter, we use achievable sum rate for training optimization because we aim to measure the improvement in sum rate provided by the spatial multiplexing gains. Optimization based on sum rate and throughput is considered in [56–58, 60], where [57, 58] study a single antenna system and [56] uses a constant spatial multiplexing gain. In [60] the channel is assumed constant during training. In contrast, we consider the effect of time-variation during both training and data phases. Our contribution, is in deriving a novel analytical proof to show that using constant amplitude pilots, the interference effect from time-selective channels does not diminish in the asymptotic case (infinite BS antennas). We also provide a sum rate performance comparison between block and time-selective fading channels in order to look at optimal training. So far, we have not found any research that provides a comparison between optimal training of block and time-selective fading models.

The rest of this chapter is arranged as follows. Section 4.2 discusses the general system model. Section 4.3 describes the derivation of the channel estimation process. Section 4.4 provides the derivation of the sum rate. This is followed by numerical results in Section 4.5 and a summary in Section 4.6.

4.2 System model

We assume a narrow band channel in a single cell system. The uplink pilot is used to estimate the channel. Then, the estimated channel is used to beamform the downlink data for the remainder of the frame, as shown in Figure 4.1. Pilot and data transmission for all UEs are completely synchronized. We assume channel reciprocity (equal uplink and downlink channels). Each transmission frame has a length of T_{f} symbols, consisting of a pilot of length T_{p} and data of length $T_{\mathrm{d}} = T_{\mathrm{f}} - T_{\mathrm{p}}$. We set the number of spatial

multiplexed UEs equal to the pilot length, $K = T_{\mathrm{p}}$. The number of BS antennas is M. The channel vector between the k-th UE and the BS at time t is $\sqrt{\beta_k}\boldsymbol{h}_{kt}$, where $\boldsymbol{h}_{kt}$ is $1 \times M$ small scale fading with $\boldsymbol{h}_{kt} \sim \mathcal{CN}(0, \boldsymbol{I}_M)$ and $\sqrt{\beta_k}$ is large scale fading that is time-invariant. We assume no channel correlation between different antennas. We use a first order Gauss-Markov time-selective channel model for the small scale fading. We allow the channel to vary within the transmission frame, but stay constant within each symbol. The relationship between the channel at time t and s for user k is given by [58]

$$\boldsymbol{h}_{kt} = \delta^{|t-s|}\boldsymbol{h}_{ks} + \sqrt{1-\delta^{2|t-s|}}\boldsymbol{e}_{kt} \tag{4.1}$$

where $\delta \in [0, 1]$ is the correlation between two consecutive symbols and $\boldsymbol{e}_{kt}$ has a $\mathcal{CN}(0, \boldsymbol{I}_M)$ distribution, which is mutually independent for different values of k and t. The value of δ is determined by bandwidth, B, and coherence time, T_{C} (which depends on maximum Doppler shift [29]). δ can be expressed as $\delta^{T_{\mathrm{C}}B} = w$ [59], where w is the correlation value between two ends of the channel in one coherence time that is typically set to $w = 0.5$ [29]. The typical value of δ varies between 0.9 and 0.99 [59]. For example, for $w = 0.5$, 200 Hz of maximum Doppler shift and 10 kHz bandwidth will result to $\delta = 0.968$. High value δ means slow fading due to a slow moving UE while small value δ means fast fading as a result of a high speed UE. From (4.1), the cross-covariance between channels at time t and s is

$$\mathbb{E}[\boldsymbol{h}_{kt}^H \boldsymbol{h}_{ks}] = \boldsymbol{I}_M \delta^{|t-s|}. \tag{4.2}$$

We assume the BS knows the channel statistical properties, namely the large scale fading and time-correlation coefficient.

4.3 Channel estimation

To estimate the channel, we use the pilot signal received at the BS. We define the user k pilot as $\boldsymbol{\psi}_k = [\psi_{k1}\ \psi_{k2} \cdots \psi_{k\tau} \cdots \psi_{kK}]^T$, where $|\psi_{k\tau}|^2 = 1$ for any discrete time value τ (i.e. constant amplitude) and $\psi_{k\tau}$ is the pilot sequence for UE k at training time τ. If the target UE is k', its pilot sequence is set to be orthogonal with pilot sequences from different UEs, or $\boldsymbol{\psi}_{k'}^H \boldsymbol{\psi}_k = 0$ for $k \neq k'$. To achieve this, we consider Zadoff-Chu

sequences, which are defined as

$$\psi_{k\tau} = \exp\left(-\left(\tau - k\right)\left(\tau - k + (K \bmod 2)\right) u\pi i/K\right),$$

where u is a constant integer with $0 < u < K$ and $\gcd(K, u) = 1$. The received pilot signal matrix at the BS from all users is given by the $K \times M$ matrix $\boldsymbol{Y} = [\boldsymbol{y}_1\ \boldsymbol{y}_2 \cdots \boldsymbol{y}_\tau \cdots \boldsymbol{y}_K]^T$, where

$$\boldsymbol{y}_\tau = \sum_{k=1}^{K} \sqrt{P_\mathrm{p}\beta_k}\psi_{k\tau}\boldsymbol{h}_{k\tau} + \boldsymbol{n}_\tau, \tag{4.3}$$

P_p is the UE average transmit pilot power and $\boldsymbol{n}_\tau$ is a $1 \times M$ noise vector with a $\mathcal{CN}(0, \boldsymbol{I}_M)$ distribution. The received signal is correlated with pilot sequence k' as follows

$$\begin{aligned}
\check{\boldsymbol{h}}_{k'} =& \boldsymbol{\psi}_{k'}^H \boldsymbol{Y} \\
=& \sum_{\tau=1}^{K} \psi_{k'\tau}^* \boldsymbol{y}_\tau \\
=& \underbrace{\sum_{\tau=1}^{K} \sqrt{P_\mathrm{p}\beta_{k'}}\psi_{k'\tau}^*\psi_{k'\tau}\boldsymbol{h}_{k'\tau}}_{A} + \underbrace{\sum_{k\neq k'}^{K}\sum_{\tau=1}^{K} \sqrt{P_\mathrm{p}\beta_k}\psi_{k'\tau}^*\psi_{k\tau}\boldsymbol{h}_{k\tau}}_{B} + \sum_{\tau=1}^{K}\psi_{k'\tau}^*\boldsymbol{n}_\tau,
\end{aligned} \tag{4.4}$$

where A is the received signal from the target UE k' and B is interference from pilot signals with different sequences. If $\boldsymbol{h}_{k\tau}$ in B is constant across τ, such as in the case of block fading, then the term B is eliminated due to pilot orthogonality, or $\sum_{\tau=1}^{K} \psi_{k'\tau}^*\psi_{k\tau} = 0$. However, since $\boldsymbol{h}_{k\tau}$ is varying across τ in selective fading, B becomes non-zero.

To obtain a lower bound on sum rate [8], we use the linear minimum mean square error (LMMSE) channel estimate [38].

Theorem 4.1. The LMMSE estimate of the channel for UE k' at time t can be expressed as

$$\hat{\boldsymbol{h}}_{k't} = \theta_{k't}\check{\boldsymbol{h}}_{k'}, \tag{4.5}$$

where

$$\theta_{k't} = \frac{\sqrt{P_{\mathrm{p}}\beta_{k'}}\sum_{\tau=1}^{K}\delta^{|t-\tau|}}{\sum_{k=1}^{K}\sum_{\tau=1}^{K}\sum_{s=1}^{K}P_{\mathrm{p}}\beta_k\psi_{k'\tau}\psi_{k\tau}^{*}\psi_{k's}^{*}\psi_{ks}\delta^{|\tau-s|}+K}.$$

Proof. We aim to obtain LMMSE estimation of channel $\boldsymbol{h}_{k't}$ using information from $\check{\boldsymbol{h}}_{k'}$ in (4.4) and channel statistical properties. From [46], the LMMSE channel estimate is given by

$$\hat{\boldsymbol{h}}_{k't}^{T} = \boldsymbol{R}_{\boldsymbol{h}_{k't}\check{\boldsymbol{h}}_{k'}}\left(\boldsymbol{R}_{\check{\boldsymbol{h}}_{k'}\check{\boldsymbol{h}}_{k'}}\right)^{-1}\check{\boldsymbol{h}}_{k'}^{T} \tag{4.6}$$

where $\boldsymbol{R}_{\boldsymbol{h}_{k't}\check{\boldsymbol{h}}_{k'}}$ is the cross-covariance between $\boldsymbol{h}_{k't}$ and $\check{\boldsymbol{h}}_{k'}$, and $\boldsymbol{R}_{\check{\boldsymbol{h}}_{k'}\check{\boldsymbol{h}}_{k'}}$ is the covariance of $\check{\boldsymbol{h}}_{k'}$. Using (4.2) and (4.4), and by eliminating uncorrelated terms, we get

$$\begin{aligned}\boldsymbol{R}_{\boldsymbol{h}_{k't}\check{\boldsymbol{h}}_{k'}} =& \mathbb{E}\left[\boldsymbol{h}_{k't}^{H}\check{\boldsymbol{h}}_{k'}\right]\\ =&\sqrt{P_{\mathrm{p}}\beta_{k'}}\sum_{\tau=1}^{K}\delta^{|t-\tau|}\boldsymbol{I}_M,\end{aligned} \tag{4.7}$$

and

$$\begin{aligned}\boldsymbol{R}_{\check{\boldsymbol{h}}_{k'}\check{\boldsymbol{h}}_{k'}} =& \mathbb{E}\left[\check{\boldsymbol{h}}_{k'}^{H}\check{\boldsymbol{h}}_{k'}\right]\\ =&\mathbb{E}\left[\sum_{k=1}^{K}\sum_{\tau=1}^{K}\sum_{l=1}^{K}\sum_{s=1}^{K}\sqrt{P_{\mathrm{p}}\beta_k P_{\mathrm{p}}\beta_l}\psi_{k'\tau}\psi_{k\tau}^{*}\psi_{k's}^{*}\psi_{ls}\boldsymbol{h}_{k\tau}^{H}\boldsymbol{h}_{ls}\right]+\sum_{\tau=1}^{K}\mathbb{E}\left[\boldsymbol{n}_{\tau}^{H}\boldsymbol{n}_{\tau}\right].\end{aligned} \tag{4.8}$$

Using (4.2) and since $\boldsymbol{h}_{k\tau}$ is independent for different k, we simplifiy (4.8) as

$$\boldsymbol{R}_{\check{\boldsymbol{h}}_{k'}\check{\boldsymbol{h}}_{k'}} = \sum_{k=1}^{K}\sum_{\tau=1}^{K}\sum_{s=1}^{K}P_{\mathrm{p}}\beta_k\psi_{k'\tau}\psi_{k\tau}^{*}\psi_{k's}^{*}\psi_{ks}\delta^{|\tau-s|}\boldsymbol{I}_M + K\boldsymbol{I}_M. \tag{4.9}$$

Substituting (4.7) and (4.9) into (4.6), gives (4.5). □

Using (4.5) and (4.9), the covariance of $\hat{\boldsymbol{h}}_{k't}$ is given by

$$\begin{aligned}\boldsymbol{R}_{\hat{\boldsymbol{h}}_{k't}\hat{\boldsymbol{h}}_{k't}} =& \mathbb{E}\left[\hat{\boldsymbol{h}}_{k't}^{H}\hat{\boldsymbol{h}}_{k't}\right]\\ =&\theta_{k't}^{2}\boldsymbol{R}_{\check{\boldsymbol{h}}_{k'}\check{\boldsymbol{h}}_{k'}}\\ =&\alpha_{k't}\boldsymbol{I}_M,\end{aligned} \tag{4.10}$$

where

$$\alpha_{k't} = \frac{P_{\mathrm{p}}\beta_{k'}\left(\sum_{\tau=1}^{K}\delta^{|t-\tau|}\right)^2}{\sum_{k=1}^{K}\sum_{\tau=1}^{K}\sum_{s=1}^{K}P_{\mathrm{p}}\beta_k\psi_{k'\tau}\psi_{k\tau}^*\psi_{k's}^*\psi_{ks}\delta^{|\tau-s|}+K}. \tag{4.11}$$

The channel estimation error at time t is $\tilde{\boldsymbol{h}}_{k't} = \boldsymbol{h}_{k't} - \hat{\boldsymbol{h}}_{k't}$ and $\tilde{\boldsymbol{h}}_{k't}$ is uncorrelated with $\hat{\boldsymbol{h}}_{k't}$. The covariance of $\tilde{\boldsymbol{h}}_{k't}$ is

$$\begin{aligned}\boldsymbol{R}_{\tilde{\boldsymbol{h}}_{k't}\tilde{\boldsymbol{h}}_{k't}} =& \boldsymbol{R}_{\boldsymbol{h}_{k't}\boldsymbol{h}_{k't}} - \boldsymbol{R}_{\hat{\boldsymbol{h}}_{k't}\hat{\boldsymbol{h}}_{k't}}\\ =& \boldsymbol{I}_M - \boldsymbol{R}_{\hat{\boldsymbol{h}}_{k't}\hat{\boldsymbol{h}}_{k't}}\\ =& \left(1-\alpha_{k't}\right)\boldsymbol{I}_M.\end{aligned} \tag{4.12}$$

The covariance in (4.10) and (4.12) will be used to derive the closed form transmission rate.

4.4 Achievable sum rate

The received downlink data signal at UE k' at time t can be expressed as

$$x_{k't} = \sum_{k=1}^{K}\sqrt{P_{\mathrm{d}}\beta_{k'}}\boldsymbol{h}_{k't}\boldsymbol{a}_{kt}q_{kt} + v_{k't}, \tag{4.13}$$

where P_{d} is the average downlink data power, $\boldsymbol{a}_{kt}$ is an $M \times 1$ precoding vector for UE k at time t, $v_{k't}$ is noise at UE k' at time t with $v_{k't} \sim \mathcal{CN}(0,1)$ and q_{kt} is downlink data to UE k at time t with $q_{kt} \sim \mathcal{CN}(0,1)$ [8]. For the precoding vector, we use conjugate beamforming defined as

$$\boldsymbol{a}_{kt} = \frac{\hat{\boldsymbol{h}}_{kt}^H}{\sqrt{K}\left\|\hat{\boldsymbol{h}}_{kt}\right\|}. \tag{4.14}$$

Theorem 4.2. A lower bound on the sum rate of the time-selective channel for K users is given by (4.15),

$$R_{\text{sum}} = \frac{1}{T_{\text{f}}} \sum_{k'=1}^{K} \sum_{t=K+1}^{T_{\text{f}}} \log_2 \left(1 + \frac{\frac{P_{\text{d}} \beta_{k'} \xi(M)^2 \alpha_{k't}}{K}}{\left(\begin{array}{c} \frac{P_{\text{d}} \beta_{k'}}{K} \left(1 + (M - 1 - \xi(M)^2) \alpha_{k't} \right) \\ + \sum_{k \neq k'}^{K} \frac{P_{\text{d}} \beta_{k'}}{K} \left(1 + (M-1) \phi_{k'kt}^2 \right) + 1 \end{array} \right)} \right), \quad (4.15)$$

where $\alpha_{k't}$ is defined in (4.11) and $\phi_{k'kt}^2$ is defined as

$$\phi_{k'kt}^2 = \frac{P_{\text{p}} \beta_{k'} \left| \sum_{\tau=1}^{K} \psi_{k\tau}^* \psi_{k'\tau} \delta^{|t-\tau|} \right|^2}{\sum_{q=1}^{K} \sum_{\tau=1}^{K} \sum_{s=1}^{K} P_{\text{p}} \beta_q \psi_{k\tau} \psi_{q\tau}^* \psi_{ks}^* \psi_{qs} \delta^{|\tau - s|} + K}. \quad (4.16)$$

Proof. We use a similar lower bound to that in [8]. However, since the signal to interference and noise ratio (SINR) for the time-selective channel model will be time dependent, this also means that there will be different achievable rates at different times. Therefore, we sum the rate equation from $t = K + 1$ to $t = T_{\text{f}}$ (the data time interval) and divide by the frame length T_{f} [56] to obtain the average data rate per symbol,

$$r_{k'} = \frac{1}{T_{\text{f}}} \sum_{t=K+1}^{T_{\text{f}}} \log_2 \left(1 + \frac{S_{k't}}{N_{k't}} \right), \quad (4.17)$$

where $S_{k't}$ in (4.17) is the effective (desired) power for UE k' at time t. Using a similar method as in [8], we assume that the effective channel of the transmission is $\sqrt{P_{\text{d}} \beta_{k'}} \mathbb{E}\left[\boldsymbol{h}_{k't} \boldsymbol{a}_{k't} \right]$. As a result,

$$S_{k't} = P_{\text{d}} \beta_{k'} \mathbb{E}\left[\boldsymbol{h}_{k't} \boldsymbol{a}_{k't} \right]^2. \quad (4.18)$$

Using a similar method as in [8, eq. (13)] and (3.27), but with the precoding in (4.14), we can solve the expectation in (4.18) to give

$$\begin{aligned} S_{k't} &= P_{\text{d}} \beta_{k'} \mathbb{E}\left[\boldsymbol{h}_{k't} \frac{\hat{\boldsymbol{h}}_{k't}^H}{\sqrt{K} \left\| \hat{\boldsymbol{h}}_{k't} \right\|} \right]^2 \\ &= \frac{P_{\text{d}} \beta_{k'}}{K} \mathbb{E}\left[\left\| \hat{\boldsymbol{h}}_{k't} \right\| \right]^2 \\ &= \frac{P_{\text{d}} \beta_{k'}}{K} \xi(M)^2 \left\| \boldsymbol{R}_{\hat{\boldsymbol{h}}_{k't} \hat{\boldsymbol{h}}_{k't}} \right\| \\ &= \frac{P_{\text{d}} \beta_{k'}}{K} \xi(M)^2 \alpha_{k't}, \end{aligned} \quad (4.19)$$

where $\xi(M)$ is defined in (3.27) in Chapter 3. $N_{k't}$ in (4.17) is the interference and noise for UE k' at time t which equals the total received power minus the effective power, $N_{k't} = \mathbb{E}\left[|x_{k't}|^2\right] - S_{k't}$. This can be expanded as

$$N_{k't} = P_{\mathrm{d}}\beta_{k'}\mathbb{E}\left[|\boldsymbol{h}_{k't}\boldsymbol{a}_{k't}|^2\right] + \sum_{k\neq k'}^{K} P_{\mathrm{d}}\beta_{k'}\mathbb{E}\left[|\boldsymbol{h}_{k't}\boldsymbol{a}_{kt}|^2\right] + 1 - S_{k't}. \tag{4.20}$$

The first expectation term in (4.20) can be solved using a similar method to [8, eq. (14)], but using precoding defined in (4.14). Similar derivation can also be seen in (3.29) in Chapter 3. Using results from (4.10), (4.12) and Lemma C.1, the expectation term becomes

$$\begin{aligned}
\mathbb{E}\left[|\boldsymbol{h}_{k't}\boldsymbol{a}_{k't}|^2\right] =& \mathbb{E}\left[\left|\boldsymbol{h}_{k't}\frac{\hat{\boldsymbol{h}}_{k't}^H}{\sqrt{K}\left\|\hat{\boldsymbol{h}}_{k't}\right\|}\right|^2\right] \\
=& \mathbb{E}\left[\left|\hat{\boldsymbol{h}}_{k't}\frac{\hat{\boldsymbol{h}}_{k't}^H}{\sqrt{K}\left\|\hat{\boldsymbol{h}}_{k't}\right\|}\right|^2\right] + \mathbb{E}\left[\left|\tilde{\boldsymbol{h}}_{k't}\frac{\hat{\boldsymbol{h}}_{k't}^H}{\sqrt{K}\left\|\hat{\boldsymbol{h}}_{k't}\right\|}\right|^2\right] \\
=& \frac{1}{K}\left(\mathrm{tr}\left(\boldsymbol{R}_{\hat{\boldsymbol{h}}_{k't}\hat{\boldsymbol{h}}_{k't}}\right) + \frac{1}{M}\mathrm{tr}\left(\boldsymbol{R}_{\tilde{\boldsymbol{h}}_{k't}\tilde{\boldsymbol{h}}_{k't}}\right)\right) \\
=& \frac{1}{K}(\alpha_{k't}M + 1 - \alpha_{k't}).
\end{aligned} \tag{4.21}$$

The second expectation term in (4.20) is interference from downlink data to other UEs. From (4.5) and (4.14), it can be shown that $\boldsymbol{a}_{kt} = \frac{\check{\boldsymbol{h}}_k^H}{\sqrt{K}\|\check{\boldsymbol{h}}_k\|}$. $\check{\boldsymbol{h}}_k$ has a similar expression as in (4.4), but for different UEs ($k \neq k'$). Since the interference from other UEs during training (B in (4.4)) is non-zero in time varying fading, this means that $\check{\boldsymbol{h}}_k$ will become a function of $\boldsymbol{h}_{k't}$. Therefore, $\boldsymbol{h}_{k't}$ and $\check{\boldsymbol{h}}_k$ are correlated. To solve this expectation term, we expand $\boldsymbol{h}_{k't}$ in terms of $\check{\boldsymbol{h}}_k$ as

$$\boldsymbol{h}_{k't} = \phi_{k'kt}\frac{\check{\boldsymbol{h}}_k}{\sqrt{\left\|\boldsymbol{R}_{\check{\boldsymbol{h}}_k\check{\boldsymbol{h}}_k}\right\|}} + \sqrt{1-\phi_{k'kt}^2}\boldsymbol{\epsilon}_{k't}, \tag{4.22}$$

where $\phi_{k'kt}$ is the correlation coefficient between $\boldsymbol{h}_{k't}$ and $\check{\boldsymbol{h}}_k$ given in (4.25). Both $\frac{\check{\boldsymbol{h}}_k}{\sqrt{\left\|\boldsymbol{R}_{\check{\boldsymbol{h}}_k\check{\boldsymbol{h}}_k}\right\|}}$ and $\boldsymbol{\epsilon}_{k't}$ have $\mathcal{CN}(0, \boldsymbol{I}_M)$ distributions and are uncorrelated with each other.

Using (4.22), we expand the expectation term for multiuser interference in (4.20) as

$$\begin{aligned}\mathbb{E}\left[|\boldsymbol{h}_{k't}\boldsymbol{a}_{kt}|^2\right] &= \mathbb{E}\left[\left|\boldsymbol{h}_{k't}\frac{\check{\boldsymbol{h}}_k^H}{\sqrt{K}\left\|\check{\boldsymbol{h}}_k\right\|}\right|^2\right] \\ &= \frac{\phi_{k'kt}^2}{K\left\|\boldsymbol{R}_{\check{\boldsymbol{h}}_k\check{\boldsymbol{h}}_k}\right\|}\mathbb{E}\left[\left|\check{\boldsymbol{h}}_k\frac{\check{\boldsymbol{h}}_k^H}{\left\|\check{\boldsymbol{h}}_k\right\|}\right|^2\right] + \frac{1-\phi_{k'kt}^2}{K}\mathbb{E}\left[\left|\boldsymbol{\epsilon}_{k't}\frac{\check{\boldsymbol{h}}_k^H}{\left\|\check{\boldsymbol{h}}_k\right\|}\right|^2\right] \\ &= \frac{\phi_{k'kt}^2}{K\left\|\boldsymbol{R}_{\check{\boldsymbol{h}}_k\check{\boldsymbol{h}}_k}\right\|}\mathrm{tr}\left(\boldsymbol{R}_{\check{\boldsymbol{h}}_k\check{\boldsymbol{h}}_k}\right) + \frac{1-\phi_{k'kt}^2}{K}\frac{1}{M}\mathrm{tr}\left(\mathbb{E}\left[\boldsymbol{\epsilon}_{k't}^H\boldsymbol{\epsilon}_{k't}\right]\right) \\ &= \frac{1}{K}(\phi_{k'kt}^2 M + 1 - \phi_{k'kt}^2). \end{aligned} \tag{4.23}$$

To obtain the correlation coefficient between $\boldsymbol{h}_{k't}$ and $\check{\boldsymbol{h}}_k$, we first find the covariance between these two variables. Note that $\check{\boldsymbol{h}}_k$ has a similar definition to that in (4.4). Using (4.2), and eliminating uncorrelated terms, we can simplify the covariance as

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{h}_{k't}^H\check{\boldsymbol{h}}_k\right] =& \mathbb{E}\left[\sqrt{P_{\mathrm{p}}\beta_{k'}}\sum_{\tau=1}^{K}\psi_{k\tau}^*\psi_{k'\tau}\boldsymbol{h}_{k't}^H\boldsymbol{h}_{k'\tau}\right] \\ =& \sqrt{P_{\mathrm{p}}\beta_{k'}}\sum_{\tau=1}^{K}\psi_{k\tau}^*\psi_{k'\tau}\delta^{|t-\tau|}\boldsymbol{I}_M. \end{aligned} \tag{4.24}$$

The correlation coefficient is obtained as follows

$$\begin{aligned}\phi_{k'kt} =& \frac{\left\|\mathbb{E}\left[\boldsymbol{h}_{k't}^H\check{\boldsymbol{h}}_k\right]\right\|}{\sqrt{\left\|\mathbb{E}\left[\boldsymbol{h}_{k't}^H\boldsymbol{h}_{k't}\right]\right\|\left\|\mathbb{E}\left[\check{\boldsymbol{h}}_k^H\check{\boldsymbol{h}}_k\right]\right\|}} \\ =& \frac{\left\|\mathbb{E}\left[\boldsymbol{h}_{k't}^H\check{\boldsymbol{h}}_k\right]\right\|}{\sqrt{\|\boldsymbol{I}_M\|\left\|\boldsymbol{R}_{\check{\boldsymbol{h}}_k\check{\boldsymbol{h}}_k}\right\|}}. \end{aligned} \tag{4.25}$$

Using (4.9), (4.24) and (4.25), then $\phi_{k'kt}^2$ becomes (4.16).

The closed form equivalent of the first and second expectation terms in (4.20) can be obtained from (4.21) and (4.23), respectively. Therefore, $N_{k't}$ can be written as

$$N_{k't} = \frac{P_{\mathrm{d}}\beta_{k'}}{K}\left(1+(M-1)\alpha_{k't}\right) + \sum_{k\neq k'}^{K}\frac{P_{\mathrm{d}}\beta_{k'}}{K}\left(1+(M-1)\phi_{k'kt}^2\right) + 1 - S_{k't}. \tag{4.26}$$

The total sum rate for K UEs is

$$R_{\text{sum}} = \sum_{k'=1}^{K} r_{k'}. \tag{4.27}$$

Substituting (4.17), (4.19) and (4.26) into (4.27), the final sum rate becomes (4.15). □

It can be shown that $\frac{S_{k't}}{N_{k't}} \rightarrow \frac{\beta_{k'}\alpha_{k't}}{\sum_{k\neq k'}^{K}\beta_k\phi^2_{k'kt}}$ as $M \rightarrow \infty$. This means that even if we use orthogonal pilot sequences with constant amplitude in a single cell system (no pilot contamination effect [1]) and we assume no correlation between antennas, the asymptotic SINR is still limited by the interference effect due to the time-selective channel.

4.5 Results

We let $P_{\text{p}} = P_{\text{d}} = 10$ dB and $\beta_k = 1$ for all k. We consider correlation values of $\delta = 0.95$ and $\delta = 0.99$ [55]. For the time-selective channel, we use the sum rate derived in (4.15) to obtain the optimum frame length, data and pilot length using an exhaustive search. The results are shown in Figure 4.2. The frame length is varied in order to obtain the best sum rate performance. The results show that the frame length increases along with data length and pilot length when there are more antennas at the BS. This is because as the number of BS antennas increases, the effective signal power in (4.19) also increases. The improvement in SINR enables the transmission to have longer pilot and data length, thus achieving higher spatial multiplexing gain. The results also show that the optimal length for pilot is between 20% to 35% of the frame length.

To obtain the block-fading sum rate, we must define the coherence time. Using the Gauss-Markov model, we can obtain the coherence time in terms of the number of symbols as

$$T_{\text{coh}} = \text{Round}\left(\frac{\log\left(\frac{v}{100}\right)}{\log(\delta)}\right), \tag{4.28}$$

where Round$(\cdot)$ is rounding to the closest integer. If we let the frame length equal the coherence time ($T_{\text{f}} = T_{\text{coh}}$), then v will be the percentage correlation between the first symbol of two consecutive frames. The rule of thumb value for v% is defined as 50% [29]. We can obtain the sum rate for the block-fading channel using (4.15), by setting $\delta = 1$ within the coherence time and $T_{\text{f}} = T_{\text{coh}}$. Using an exhaustive search, we obtain results

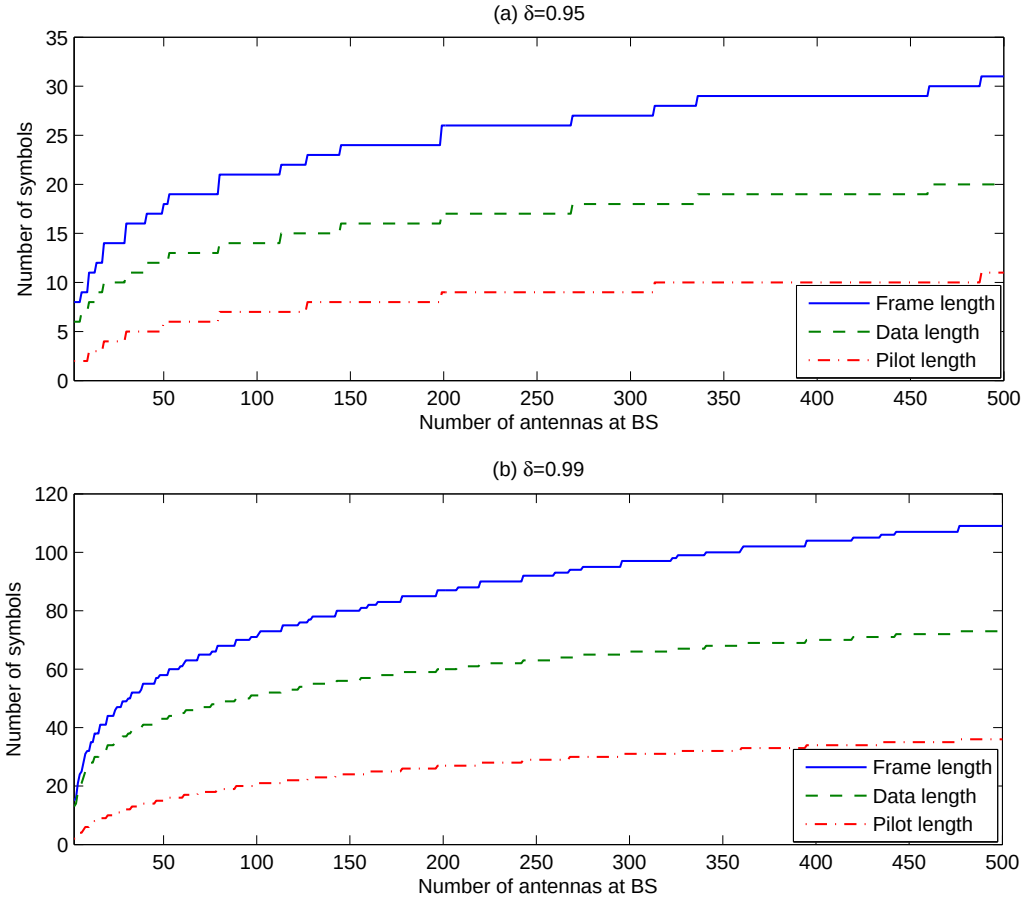

FIGURE 4.2: Optimum frame length, data length and pilot length for the time-selective channels with $\delta = 0.95$ and $\delta = 0.99$.

for optimal pilot and data lengths for the block-fading model as shown in Figure 4.3. As the number of BS antennas increases, the pilot length increases to exploit the spatial multiplexing gain, but the data length reduces in order to ensure the frame length does not grow beyond the coherence time. Comparing the frame size achieved in Figure 4.2 and Figure 4.3 for massive MIMO, we can see that the frame length can be increased beyond the coherence time to achieve optimality in a time-selective channel.

In Figure 4.4, we compare the sum rates for block-fading and time-selective channels. For each correlation value, we investigate 3 different conditions which are

- BL-BL is the sum rate of the block-fading channel using the optimal training for a block-fading channel given in Figure 4.3.
- TS-TS is the sum rate of the time-selective channel using optimal training for a time-selective channel from Figure 4.2.

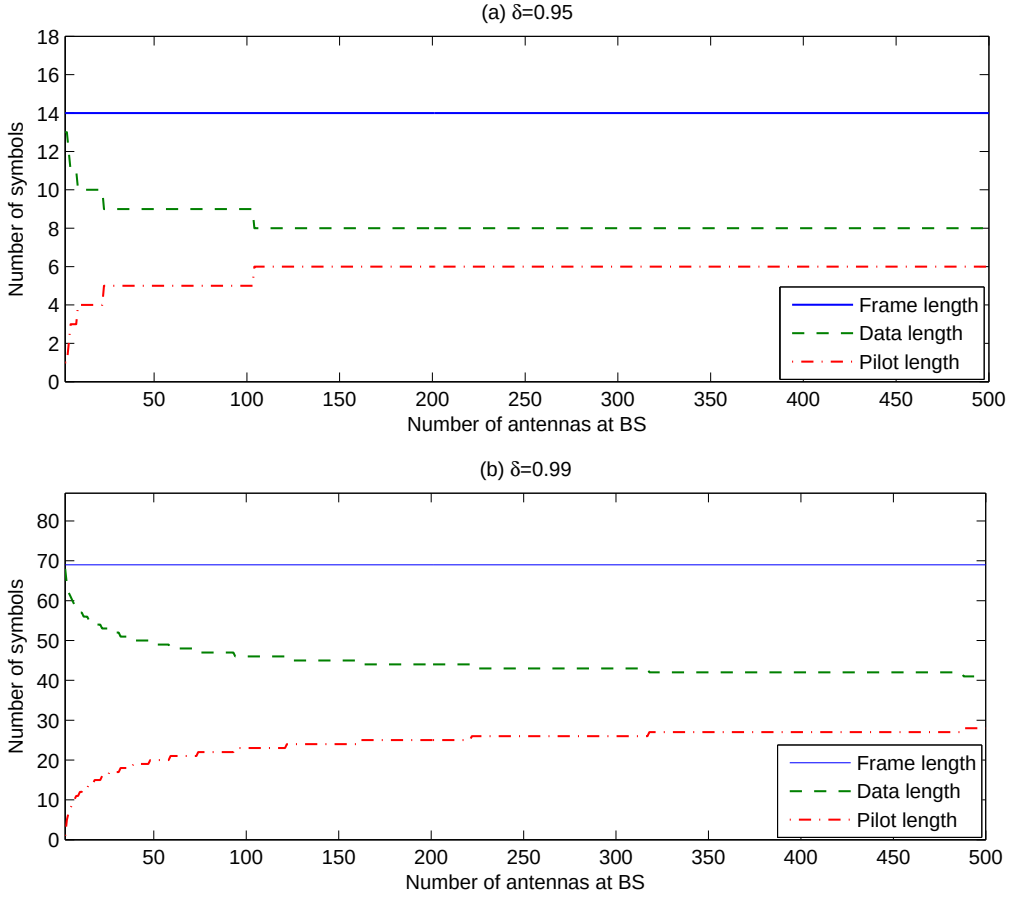

FIGURE 4.3: Optimum frame length, data length and pilot length for the block-fading channels with $\delta = 0.95$ and $\delta = 0.99$.

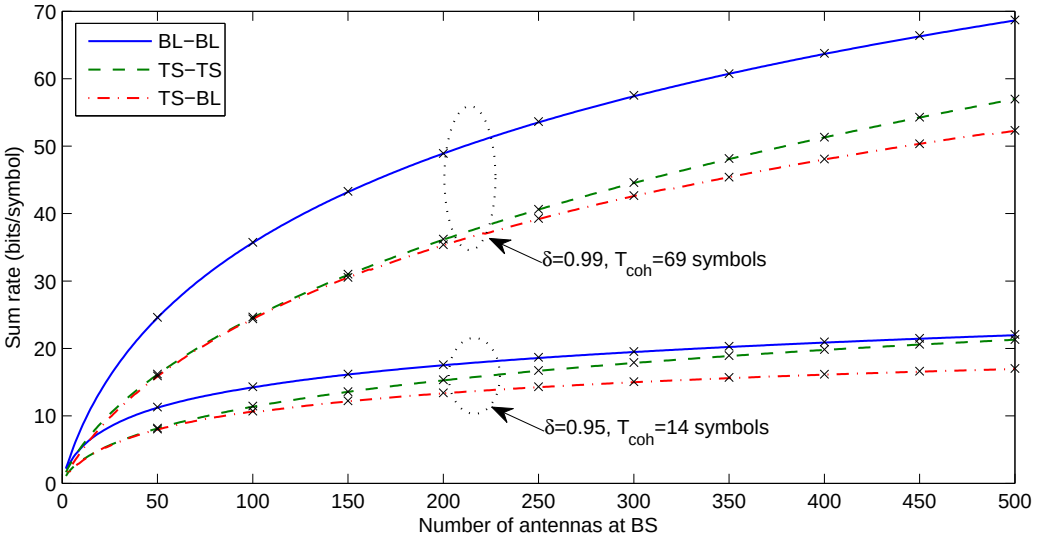

FIGURE 4.4: Performance of block-fading sum rate using optimum training for block-fading (BL-BL), time-selective sum rate using optimum training for block-fading (TS-BL) and time-selective sum rate using optimum training for time-selective (TS-TS). The lines are sum rates from analytical simulations, while each point marked with a '×' is from Monte Carlo simulations.

- TS-BL is the sum rate of the time-selective channel using optimal training for a block-fading channel from Figure 4.3.

We also obtain the Monte Carlo sum rate for each scenario by numerically solving the expected values in (4.18) and (4.20) across many random channel realizations. Results in Figure 4.4 show that the analytical sum rate agrees well with the Monte Carlo sum rate. If we use the same pilot and data length for both channel models, such as in the case of BL-BL and TS-BL, then a much higher sum rate is achieved in a block-fading rather than a time-selective channel. This is expected since the block-fading model excludes the effect of interference from channel time variation. The result also shows that if the optimum training values for the block-fading channel model are used in a time-selective channel (TS-BL), the sum rate is noticeably lower than using the optimum training values for the time-selective channel (TS-TS) for larger numbers of antennas. The training optimization for the block fading model is not optimal in a time-selective fading channel because the channel model approximation method in block fading model is limited by coherence time which can result to inaccurate analysis under certain conditions. For example, in the case of massive MIMO system, having many excess antennas at the BS means the SINR can be increased significantly, which enables the frame length to be expanded beyond the coherence time to achieve optimal length, as shown in Figure 4.2. This contradicts the block fading model, which limits the frame length to the coherence time, thus limit its potential. This shows the importance of considering channel time variation when determining training and data length in massive MIMO.

In order to analyze the impact of channel time correlation on the optimal training, Figures 4.5 and 4.6 show the relationship between δ (correlation between two consecutive symbols) with optimal frame length, data length, pilot length and sum rate. The results show that as the value of δ increases, the value of the optimum frame length, data length, pilot length and sum rate also increase. Note that for higher δ, the increase becomes steeper. This is due to the exponential relationship between the correlation and time.

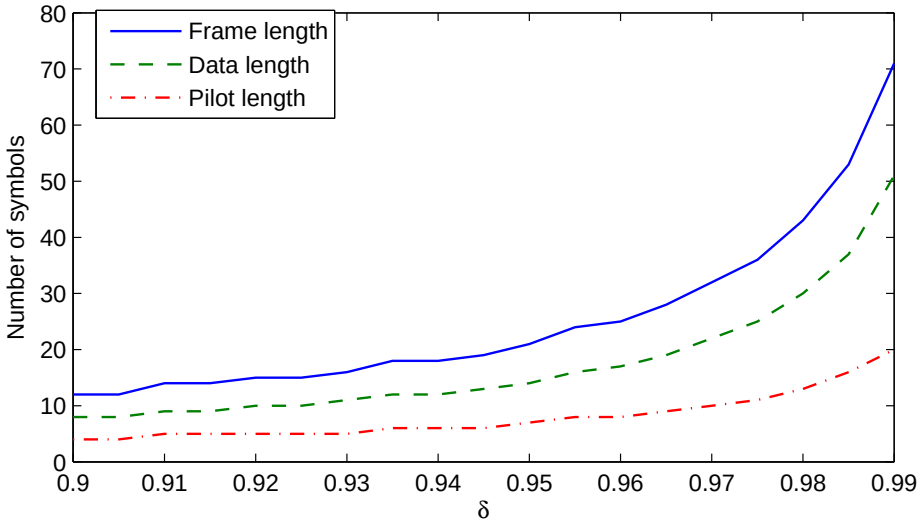

FIGURE 4.5: Optimum frame length, data length and pilot length for the time-selective channels with 100 antennas st BS for various values of δ.

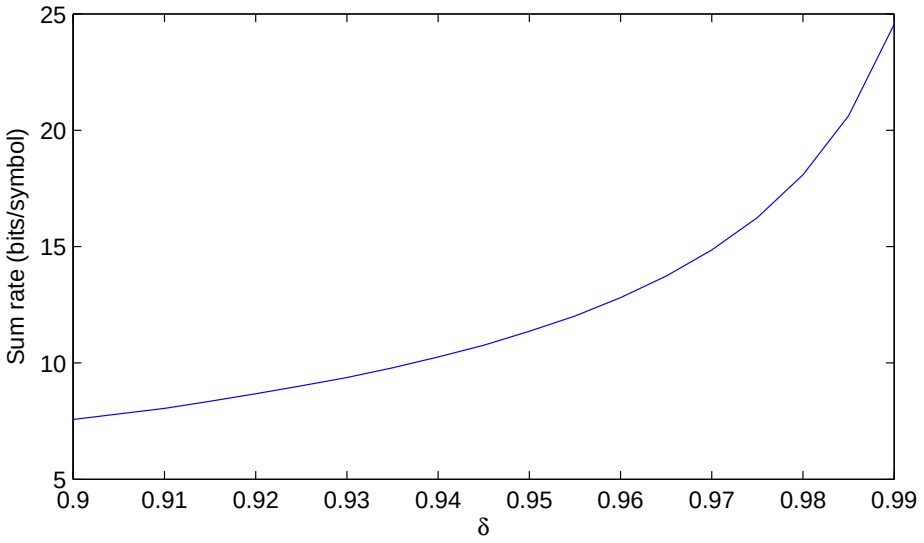

FIGURE 4.6: Sum rate with optimum training for the time-selective channels with 100 antennas st BS for various values of δ.

4.6 Summary

In this chapter, we derived the achievable sum rate for the time-selective channel model. We showed that if we used constant amplitude pilots, the asymptotic SINR performance was limited by the interference effect from the time-selective channel. We also showed that there is a significant difference between the sum rate obtained for the block-fading and time-selective models. In addition, we showed that optimal training in a block-fading channel may not be optimal in a time-selective channel for a very large number of

antennas at the BS. We showed that for time selective channel model, the optimal frame length can grow beyond the coherence time when the number of antennas becomes very large. This finding cannot be obtained using the block fading channel model because of the limitation of the frame length being equal to the coherence time in block fading channel model. We also show that the optimal length for pilot is between 20% to 35% of the frame length for the finite massive MIMO systems considered.

Chapter 5

Massive MIMO systems in time and frequency selective channels

5.1 Introduction

In Chapter 4, it is shown that the optimal training in block-fading channels may not be optimal if applied in time-selective fading channels. As discussed in Chapter 2, other than time variation, the channel may also experience frequency variation due to multipath. The combination of time and frequency channel variations leads to a doubly-selective channel model, where the channel can vary within the coherence block in both time and frequency domains. Hence, in this chapter, we aim to investigate the performance of massive MIMO systems in doubly-selective channels. Various papers have studied the optimization of training in time-selective, frequency-selective and doubly-selective fading channels [55–58, 60–67]. Training optimization in a time-selective channel model is investigated in [55, 58] with a single antenna system. The impact of the number of antennas on optimal training in time-selective channels is investigated in [56, 60, 61]. MIMO orthogonal frequency-division multiplexing (OFDM) in frequency-selective channels is studied in [62–64] by excluding channel selectivity in the time domain. A capacity-based performance analysis is used to study training design in a doubly-selective channel using a single antenna system in [57, 65]. Training design for a multiple antenna system in a doubly-selective channel with a delay tap model is

investigated in [66, 67]. This work only considers a single cell system and excludes the impact of increasing numbers of antennas on the optimal training.

In this research, we analyze the optimal training and data lengths in the time and frequency domains in a massive MIMO system. Compared to conventional MIMO, massive MIMO has the potential to achieve much higher spatial multiplexing gain due to the asymptotically orthogonal nature of the spatial signatures between different UEs as the number of BS antennas increases [1]. Since the spatial multiplexing gain depends on the size of pilot [38], the number of antennas can affect the optimal training size. We use sum rate as our performance metric since we aim to study the performance enhancement provided by the spatial multiplexing gains.

This chapter extends the study of optimal training of massive MIMO systems in Chapter 4 to time-frequency or doubly-selective channels in a multicell system. We also compare two different types of pilot sequence which are a constant amplitude pilot and a zero padded pilot. Similar pilot sequences have been investigated in [55, 61] for constant amplitude pilots and in [63, 66] for zero padded pilots. However, so far there has been no study of the impact of the number of BS antennas and the number of UEs on the optimum training size in time and frequency domains in doubly-selective channels. In addition, different channel conditions provide different statistical properties which can affect the training size. For example, a very high UE speed, but low delay spread means it is more advantageous to have a longer training length in the frequency domain and vice versa. Therefore, in addition to the performance comparison between two pilot schemes, a key contribution of this chapter is to use the novel doubly-selective channel sum rate expression to obtain the optimum training size in time and frequency domains for various numbers of UEs, numbers of BS antennas and channel statistical properties. To achieve this, two methods of time-frequency training optimization are considered, which are average optimal training and adaptive optimal training.

The rest of the chapter is arranged as follows. Sections 5.2, 5.3 and 5.4 provide the system model, channel estimation procedure and sum rate derivation, respectively. Section 5.5.1 and Section 5.5.2 discuss constant amplitude and zero padded pilots, respectively. Section 5.5.3 investigates two training optimization methods. Numerical results are given in Section 5.6, followed by a chapter summary in Section 5.7.

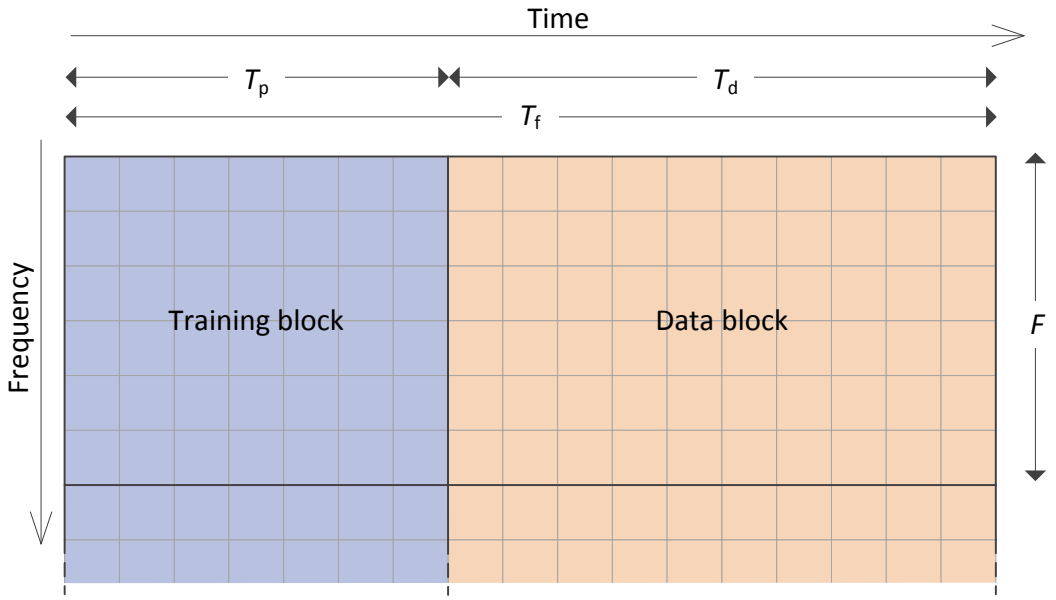

FIGURE 5.1: Time-frequency arrangement of the transmission block. T_f, T_p, T_d are the frame, pilot and data length in the time domain (number of symbols) and F is the length of the training/data block in the frequency domain (number of subcarriers). Each small square represents a resource element which is occupied by a pilot/data symbol.

5.2 System model

We assume OFDM transmission in a multicell system. Based on the time-frequency transmission arrangement in Fig. 5.1, channel estimation uses the uplink pilot during the training block. Then, the estimated channel is used to beamform the downlink data during the data block. Pilot and data transmission for all UEs are completely synchronized. We exclude the performance loss from the usage of a cyclic prefix. We assume channel reciprocity (equal uplink and downlink channels). Each transmission frame or update interval has a length of T_f symbols, consisting of a pilot of length T_p and data of length $T_d = T_f - T_p$. The total length of the pilot is $K = T_p \times F$ symbols. Note that F is not the total number of subcarriers in the transmission, but the number of subcarriers used by the pilot to achieve a unique pilot sequence of length K. The same pilot pattern within the F subcarriers can be repeated across the whole bandwidth. A pilot of length K means that there can be K orthogonal pilot sequences, which also means that K UEs can perform spatial multiplexing transmission using the same time-frequency resources in each cell. Each BS has M antennas while each UE has 1 antenna. Our channel model is similar to the multicell channel model in [8], but in order to include the effect of the doubly-selective channel, we extend the channel dimensions to include

time-frequency dimensions. Therefore, the channel vector between the k-th UE in cell j and the BS in cell l at discrete time t and subcarrier number$f$ is $\sqrt{\beta_{jlk}}\boldsymbol{h}_{jlktf}$, where $\boldsymbol{h}_{jlktf}$ is $1\times M$ small scale fading with $\boldsymbol{h}_{jlktf} \sim \mathcal{CN}(0, \boldsymbol{I}_M)$ and $\sqrt{\beta_{jlk}}$ is large scale fading that is time-frequency invariant. We assume no channel correlation between different antennas. We allow the channel to vary within the transmission block, but stay constant within each resource element. Using the Jakes model [28], the cross-covariance between channels at time-frequency $t_1 f_1$ and $t_2 f_2$ is

$$\mathbb{E}[\boldsymbol{h}^H_{jlkt_1f_1}\boldsymbol{h}_{jlkt_2f_2}] = \frac{J_0\left(\sigma_{jlk}\left|t_1 - t_2\right|\right)}{1 - i\mu_{jlk}\left(f_1 - f_2\right)}\boldsymbol{I}_M, \tag{5.1}$$

where $\sigma_{jlk} = 2\pi f_{\mathrm{d},jlk} T_\mathrm{s}$, $\mu_{jlk} = 2\pi\Delta v \tau_{\mathrm{s},jlk}$. T_s is symbol time, $f_{\mathrm{d},jlk}$ is the maximum Doppler shift between UE k in cell j and BS in cell l that depends on the UE speed and carrier frequency [29], Δv is the frequency spacing between subcarriers and $\tau_{\mathrm{s},jlk}$ is the r.m.s. delay spread between UE k in cell j and BS in cell l.

We assume the BS knows all the statistical properties of channels such as the large scale fading and the time-frequency correlation. In order to simplify the analysis, we assume that there is no inter-carrier interference due to the UEs mobility. The increase of inter-carrier interference from the UEs mobility and its impact on the performance will be an interesting topic for future research.

5.3 Channel estimation

To estimate the channel, we use the pilot signal received from the UEs at the BS. We define the user k pilot within the training block as

$$\boldsymbol{\Psi}_k = \begin{bmatrix} \psi_{k11} & \cdots & \psi_{kT_p1} \\ \vdots & \ddots & \vdots \\ \psi_{k1F} & \cdots & \psi_{kT_pF} \end{bmatrix}, \tag{5.2}$$

where ψ_{ktf} is the pilot sequence for UE k at time-frequency tf. The pilot sequences for different UEs are orthogonal to each other. As a result, the pilot sequences will have

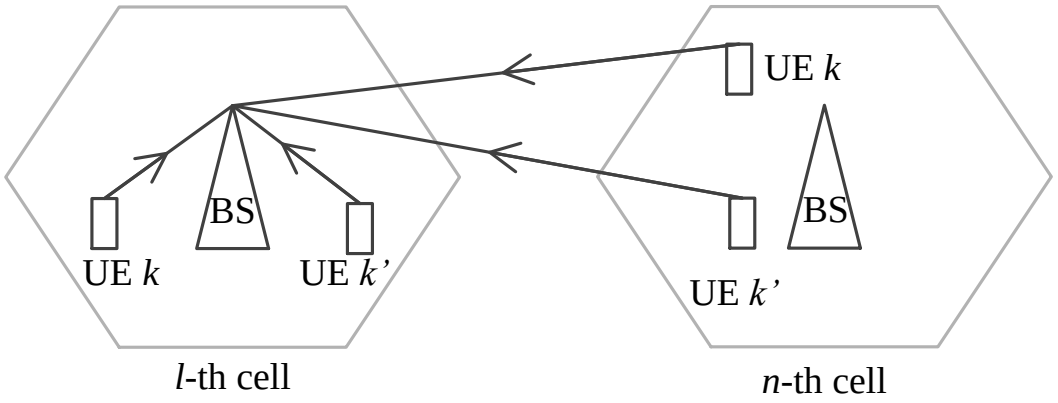

FIGURE 5.2: Uplink pilot signal received at BS in cell l from all UEs including those in nearby cells.

the following properties

$$\sum_{\{t,f\}=1}^{\{T_{\mathrm{p}},F\}} \psi^*_{k'tf}\psi_{k'tf} = K, \tag{5.3}$$

$$\sum_{\{t,f\}=1}^{\{T_{\mathrm{p}},F\}} \psi^*_{k'tf}\psi_{ktf} = 0, \quad \text{for} \quad k \neq k'. \tag{5.4}$$

These properties enable the BS to estimate the channel by eliminating interference between UEs with different pilot sequences during the training.

Fig. 5.2 shows the BS in cell l receiving pilots from all UEs in cell l and nearby cells. If the BS receives pilots from K UEs in each of L nearby cells, then the received pilot signal vector at the BS in cell l from all UEs at time-frequency tf can be expressed as

$$\boldsymbol{y}_{ltf} = \sum_{n=1}^{L}\sum_{k=1}^{K} \sqrt{P_{\mathrm{p}}\beta_{nlk}}\psi_{ktf}\boldsymbol{h}_{nlktf} + \boldsymbol{n}_{ltf}, \tag{5.5}$$

where P_{p} is the UE average transmit pilot power and $\boldsymbol{n}_{ltf}$ is a $1 \times M$ noise vector with a $\mathcal{CN}(0, \boldsymbol{I}_M)$ distribution. Referring to the properties in (5.3) and (5.4), we estimate the channel by correlating the received pilot signal with the known pilot sequence for UE k'

as follows

$$\begin{aligned}\check{\boldsymbol{h}}_{lk'} &= \sum_{\{t,f\}=1}^{\{T_p,F\}} \psi_{k'tf}^{*} \boldsymbol{y}_{ltf} \\ &= \underbrace{\sum_{n=1}^{L} \sum_{\{t,f\}=1}^{\{T_\mathrm{p},F\}} \sqrt{P_\mathrm{p}\beta_{nlk'}} \left|\psi_{k'tf}\right|^2 \boldsymbol{h}_{nlk'tf}}_{\text{Pilots from UEs with the same sequences}} \\ &+ \underbrace{\sum_{n=1}^{L} \sum_{k\neq k'}^{K} \sum_{\{t,f\}=1}^{\{T_\mathrm{p},F\}} \sqrt{P_\mathrm{p}\beta_{nlk}} \psi_{k'tf}^{*} \psi_{ktf} \boldsymbol{h}_{nlktf}}_{\text{Pilots from UEs with different sequences (multiuser interference)}} + \underbrace{\sum_{\{t,f\}=1}^{\{T_\mathrm{p},F\}} \psi_{k'tf}^{*} \boldsymbol{n}_{ltf}}_{\text{Noise}}. \end{aligned} \tag{5.6}$$

In (5.6), the first summation term is the received signal from UEs with the same pilot sequences which contains signals from UEs from other cells in addition to the target cell. This is a result of pilot reuse in other cells which will also cause pilot contamination. The second summation term in (5.6) is an interference term from signals with different pilot sequences. If we assume the channel in the training block is constant, such as in the case of a block fading channel, then we can apply the relationship in (5.4) and this interference term can be eliminated. However, since $\boldsymbol{h}_{nlktf}$ is changing across tf in selective fading, this means that the interference term may not be zero.

Next, we perform LMMSE channel estimation using (5.6). The resulting estimated channel is independent from its error, which will be used to obtain the lower bound sum rate [38]. In addition to obtaining channel estimates between the BS and UEs in the same cell, we also acquire channel estimates between the BS and UEs from different cells in order to analyze the impact of pilot contamination [8]. For example, if we aim to obtain the downlink rate for a UE in cell j, then we need to obtain the LMMSE estimate between the UE and BSs from surrounding cells in order to know the effective downlink interference from other cells. The LMMSE estimation of channel $\boldsymbol{h}_{jlk'tf}$ is acquired using channel information from $\check{\boldsymbol{h}}_{lk'}$ and statistical properties of the channel.

Theorem 5.1. The LMMSE estimate of the small scale fading between UE k' in cell j and the BS in cell l at time-frequency tf can be expressed as

$$\hat{\boldsymbol{h}}_{jlk'tf} = \theta_{jlk'tf} \check{\boldsymbol{h}}_{lk'}, \tag{5.7}$$

where

$$\theta_{jlk'tf} = \frac{\sqrt{P_{\mathrm{p}}\beta_{jlk'}}\sum_{\{t_1,f_1\}=1}^{\{T_{\mathrm{p}},F\}}\left|\psi_{k't_1f_1}\right|^2\frac{J_0\left(\sigma_{jlk'}|t-t_1|\right)}{1-i\mu_{jlk'}(f-f_1)}}{\sum_{n=1}^{L}\sum_{k=1}^{K}\sum_{\{t_1,f_1,t_2,f_2\}=1}^{\{T_{\mathrm{p}},F,T_{\mathrm{p}},F\}}P_p\beta_{nlk}\psi_{k't_1f_1}\psi^*_{kt_1f_1}\psi^*_{k't_2f_2}\psi_{kt_2f_2}\frac{J_0(\sigma_{nlk}|t_1-t_2|)}{1-i\mu_{nlk}(f_1-f_2)}+K}. \tag{5.8}$$

Proof. Using a similar approach to that in [46], the LMMSE estimate of the small scale fading between UE k' cell j and the BS in cell l at time-frequency tf can be expressed as

$$\hat{\boldsymbol{h}}^T_{jlk'tf} = \boldsymbol{R}_{\boldsymbol{h}_{jlk'tf}\check{\boldsymbol{h}}_{lk'}}\left(\boldsymbol{R}_{\check{\boldsymbol{h}}_{lk'}\check{\boldsymbol{h}}_{lk'}}\right)^{-1}\check{\boldsymbol{h}}^T_{lk'}, \tag{5.9}$$

where $\boldsymbol{R}_{\check{\boldsymbol{h}}_{lk'}\check{\boldsymbol{h}}_{lk'}}$ in (5.9) is the covariance of $\check{\boldsymbol{h}}_{lk'}$. Using $\check{\boldsymbol{h}}_{lk'}$ from (5.6), $\boldsymbol{R}_{\check{\boldsymbol{h}}_{lk'}\check{\boldsymbol{h}}_{lk'}}$ can be expressed as

$$\begin{aligned}\boldsymbol{R}_{\check{\boldsymbol{h}}_{lk'}\check{\boldsymbol{h}}_{lk'}} =& \mathbb{E}\left[\check{\boldsymbol{h}}^H_{lk'}\check{\boldsymbol{h}}_{lk'}\right]\\ =& \mathbb{E}\left[\sum_{\{n_1,n_2\}=1}^{\{L,L\}}\sum_{\{k_1,k_2\}=1}^{\{K,K\}}\sum_{\{t_1,f_1,t_2,f_2\}=1}^{\{T_{\mathrm{p}},F,T_{\mathrm{p}},F\}}\sqrt{P^2_{\mathrm{p}}\beta_{n_1lk_1}\beta_{n_2lk_2}}\right.\\ &\left.\psi_{k't_1f_1}\psi^*_{k_1t_1f_1}\psi^*_{k't_2f_2}\psi_{k_2t_2f_2}\boldsymbol{h}^H_{n_1lk_1t_1f_1}\boldsymbol{h}_{n_2lk_2t_2f_2}\right]\\ &+\sum_{\{t,f\}=1}^{\{T_{\mathrm{p}},F\}}\left|\psi_{k'tf}\right|^2\mathbb{E}\left[\boldsymbol{n}^H_{tf}\boldsymbol{n}_{tf}\right].\end{aligned} \tag{5.10}$$

Using (5.1), (5.3) and since $\boldsymbol{h}_{nlktf}$ is independent for different values of n and k, we simplify (5.10) as

$$\begin{aligned}\boldsymbol{R}_{\check{\boldsymbol{h}}_{lk'}\check{\boldsymbol{h}}_{lk'}} =& \sum_{n=1}^{L}\sum_{k=1}^{K}\sum_{\{t_1,f_1,t_2,f_2\}=1}^{\{T_{\mathrm{p}},F,T_{\mathrm{p}},F\}}\left(P_p\beta_{nlk}\psi_{k't_1f_1}\psi^*_{kt_1f_1}\psi^*_{k't_2f_2}\psi_{kt_2f_2}\right.\\ &\left.\mathbb{E}\left[\boldsymbol{h}^H_{nlkt_1f_1}\boldsymbol{h}_{nlkt_2f_2}\right]\right)+K\boldsymbol{I}_M\\ =& \sum_{n=1}^{L}\sum_{k=1}^{K}\sum_{\{t_1,f_1,t_2,f_2\}=1}^{\{T_{\mathrm{p}},F,T_{\mathrm{p}},F\}}\left(P_p\beta_{nlk}\psi_{k't_1f_1}\psi^*_{kt_1f_1}\psi^*_{k't_2f_2}\psi_{kt_2f_2}\right.\\ &\left.\frac{J_0\left(\sigma_{nlk}\left|t_1-t_2\right|\right)}{1-i\mu_{nlk}\left(f_1-f_2\right)}\boldsymbol{I}_M\right)+K\boldsymbol{I}_M.\end{aligned} \tag{5.11}$$

$\boldsymbol{R}_{\boldsymbol{h}_{jlk'tf}\check{\boldsymbol{h}}_{lk'}}$ in (5.9) is the cross-covariance between the true channel, $\boldsymbol{h}_{jlk'tf}$ and $\check{\boldsymbol{h}}_{lk'}$. Hence, we eliminate the terms in $\check{\boldsymbol{h}}_{lk'}$ (see (5.6)) that are not correlated with $\boldsymbol{h}_{jlk'tf}$ and obtain

$$\begin{aligned}\boldsymbol{R}_{\boldsymbol{h}_{jlk'tf}\check{\boldsymbol{h}}_{lk'}} =& \mathbb{E}\left[\boldsymbol{h}_{jlk'tf}^H \check{\boldsymbol{h}}_{lk'}\right] \\ =& \sqrt{P_{\mathrm{p}}\beta_{jlk'}} \sum_{\{t_1,f_1\}=1}^{\{T_{\mathrm{p}},F\}} \left|\psi_{k't_1f_1}\right|^2 \mathbb{E}\left[\boldsymbol{h}_{jlk'tf}^H \boldsymbol{h}_{jlk't_1f_1}\right] \\ =& \sqrt{P_{\mathrm{p}}\beta_{jlk'}} \sum_{\{t_1,f_1\}=1}^{\{T_{\mathrm{p}},F\}} \left|\psi_{k't_1f_1}\right|^2 \frac{J_0\left(\sigma_{jlk'}\left|t-t_1\right|\right)}{1-i\mu_{jlk'}\left(f-f_1\right)} \boldsymbol{I}_M. \end{aligned} \tag{5.12}$$

Substituting (5.11) and (5.12) into (5.9), gives (5.7) as required, since (5.11) and (5.12) are scaled identity matrices. □

Using (5.7) and (5.11), the covariance of $\hat{\boldsymbol{h}}_{jlk'tf}$ can be expressed as

$$\begin{aligned}\boldsymbol{R}_{\hat{\boldsymbol{h}}_{jlk'tf}\hat{\boldsymbol{h}}_{jlk'tf}} =& \mathbb{E}\left[\hat{\boldsymbol{h}}_{jlk'tf}^H \hat{\boldsymbol{h}}_{jlk'tf}\right] \\ =& \left|\theta_{jlk'tf}\right|^2 \boldsymbol{R}_{\check{\boldsymbol{h}}_{lk'}\check{\boldsymbol{h}}_{lk'}} \qquad (5.13) \\ =& \alpha_{jlk'tf}\boldsymbol{I}_M, \qquad (5.14)\end{aligned}$$

where $\alpha_{jlk'tf}$ is given in (5.19).

The channel estimation error at time-frequency tf is $\tilde{\boldsymbol{h}}_{jlk'tf} = \boldsymbol{h}_{jlk'tf} - \hat{\boldsymbol{h}}_{jlk'tf}$ and $\tilde{\boldsymbol{h}}_{jlk'ft}$ is uncorrelated with $\hat{\boldsymbol{h}}_{jlk'tf}$. The covariance of $\tilde{\boldsymbol{h}}_{jlk'tf}$ is

$$\begin{aligned}\boldsymbol{R}_{\tilde{\boldsymbol{h}}_{jlk'tf}\tilde{\boldsymbol{h}}_{jlk'tf}} =& \boldsymbol{R}_{\boldsymbol{h}_{jlk'tf}\boldsymbol{h}_{jlk'tf}} - \boldsymbol{R}_{\hat{\boldsymbol{h}}_{jlk'tf}\hat{\boldsymbol{h}}_{jlk'tf}} \\ =& \boldsymbol{I}_M - \boldsymbol{R}_{\hat{\boldsymbol{h}}_{jlk'tf}\hat{\boldsymbol{h}}_{jlk'tf}} \\ =& \left(1-\alpha_{jlk'tf}\right)\boldsymbol{I}_M. \end{aligned} \tag{5.15}$$

The covariances in (5.14) and (5.15) will be used to derive the closed form transmission rate.

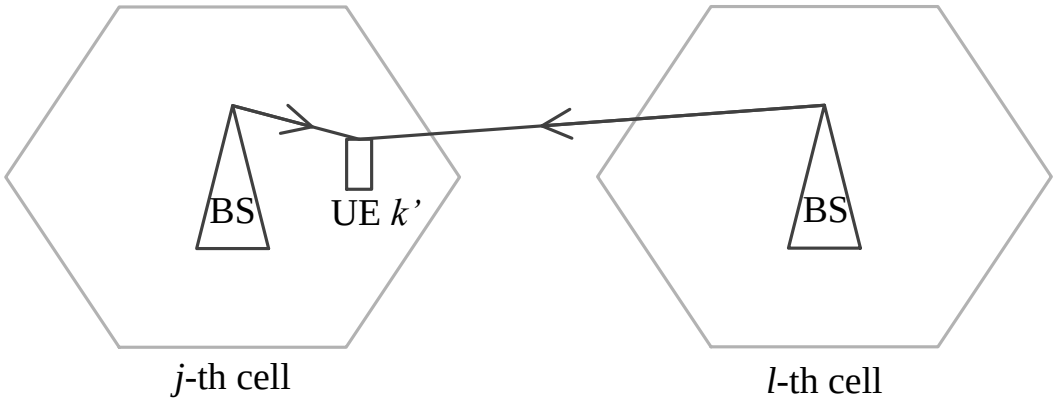

FIGURE 5.3: Downlink data signal received at UE k' in cell j from all BSs in surrounding cells.

5.4 Achievable sum rate

Fig. 5.3 shows that UE k' in cell j receives downlink data from BSs in cell j and surrounding cells. The received downlink data signal at time-frequency tf can be expressed as

$$x_{jk'tf} = \sum_{l=1}^{L} \sum_{k=1}^{K} \sqrt{P_{\mathrm{d}} \beta_{jlk'}} \boldsymbol{h}_{jlk'tf} \boldsymbol{a}_{lktf} q_{lktf} + v_{jk'tf}, \tag{5.16}$$

where

– P_{d} is the average downlink data power;

– $\boldsymbol{a}_{lktf}$ is an $M \times 1$ precoding vector for user k in cell l at time-frequency tf;

– $v_{jk'tf}$ is noise at UE k' in cell j at time-frequency tf with $v_{k'tf} \sim \mathcal{CN}(0,1)$;

– q_{lktf} is downlink data to user k in cell l at time-frequency tf with $q_{lktf} \sim \mathcal{CN}(0,1)$.

The matched filter (MF) precoding is defined as [68]

$$\boldsymbol{a}_{lktf} = \frac{\hat{\boldsymbol{h}}_{llktf}^{H}}{\sqrt{K} \left\| \hat{\boldsymbol{h}}_{llktf} \right\|}. \tag{5.17}$$

Theorem 5.2. A lower bound on the mean sum rate of the doubly-selective channel for K users in cell j is given by

$$R_j = \frac{1}{T_{\mathrm{f}}F} \sum_{\{k',f\}=1}^{\{K,F\}} \sum_{t=T_{\mathrm{p}}+1}^{T_{\mathrm{f}}} \log_2 \left(1 + \frac{\frac{P_{\mathrm{d}}\beta_{jjk'}\xi(M)^2\alpha_{jk'tf}}{K}}{\left(\begin{array}{c} \sum_{l=1}^{L} \frac{P_{\mathrm{d}}\beta_{jlk'}}{K}\left(1+(M-1)\alpha_{jlk'tf}\right) \\ +\sum_{l=1}^{L}\sum_{k\neq k'}^{K} \frac{P_{\mathrm{d}}\beta_{jlk'}}{K}\left(1+(M-1)\phi^2_{jlk'ktf}\right) \\ +1-\frac{P_{\mathrm{d}}\beta_{jjk'}\xi(M)^2\alpha_{jk'tf}}{K} \end{array} \right)} \right), \tag{5.18}$$

where

$$\alpha_{jlk'tf} = \frac{P_{\mathrm{p}}\beta_{jlk'} \left| \sum_{\{t_1,f_1\}=1}^{\{T_{\mathrm{p}},F\}} \left|\psi_{k't_1f_1}\right|^2 \frac{J_0\left(\sigma_{jlk'}|t-t_1|\right)}{1-i\mu_{jlk'}(f-f_1)} \right|^2}{\sum_{n=1}^{L}\sum_{k=1}^{K}\sum_{\{t_1,f_1,t_2,f_2\}=1}^{\{T_{\mathrm{p}},F,T_{\mathrm{p}},F\}} \left(P_{\mathrm{p}}\beta_{nlk}\psi_{k't_1f_1}\psi^*_{kt_1f_1}\psi^*_{k't_2f_2}\psi_{kt_2f_2}\frac{J_0(\sigma_{nlk}|t_1-t_2|)}{1-i\mu_{nlk}(f_1-f_2)}\right) + K}, \tag{5.19}$$

and

$$\phi^2_{jlk'ktf} = \frac{P_{\mathrm{p}}\beta_{jlk'} \left| \sum_{\{t_1,f_1\}=1}^{\{T_{\mathrm{p}},F\}} \psi^*_{kt_1f_1}\psi_{k't_1f_1}\frac{J_0\left(\sigma_{jlk}|t-t_1|\right)}{1-i\mu_{jlk}(f-f_1)} \right|^2}{\sum_{n=1}^{L}\sum_{q=1}^{K}\sum_{\{t_1,f_1,t_2,f_2\}=1}^{\{T_{\mathrm{p}},F,T_{\mathrm{p}},F\}} \left(P_{\mathrm{p}}\beta_{nlq}\psi_{kt_1f_1}\psi^*_{qt_1f_1}\psi^*_{kt_2f_2}\psi_{qt_2f_2}\frac{J_0(\sigma_{nlq}|t_1-t_2|)}{1-i\mu_{nlq}(f_1-f_2)}\right) + K}. \tag{5.20}$$

Proof. To find the sum rate bound, we use a similar approach to that in [8]. However, since the SINR for the doubly-selective channel model will be time and frequency dependent, this also means that there will be different achievable rates at different times and frequencies. Therefore, we sum the rate equation within the data block interval, which is from $t = T_{\mathrm{p}} + 1$ to $t = T_{\mathrm{f}}$ in the time domain and from $f = 1$ to $f = F$ in the frequency domain and then divide by the total block size, $T_{\mathrm{f}}F$, to obtain the average rate per symbol per subcarrier. As a result, the lower bound ergodic rate for UE k' in cell j is

$$r_{jk'} = \frac{1}{T_{\mathrm{f}}F} \sum_{t=T_{\mathrm{p}}+1}^{T_{\mathrm{f}}} \sum_{f=1}^{F} \log_2\left(1+\frac{S_{jk'tf}}{N_{jk'tf}}\right), \tag{5.21}$$

where $S_{jk'tf}$ in (5.21) is the effective (desired) power for UE k' in cell j at time-frequency tf. Using a similar method as in [8], we assume that the effective channel of the transmission is $\sqrt{P_{\mathrm{d}}\beta_{jjk'}}\mathbb{E}\left[\boldsymbol{h}_{jjk'tf}\boldsymbol{a}_{jk'tf}\right]$. As a result,

$$S_{jk'tf} = P_{\mathrm{d}}\beta_{jjk'}\left|\mathbb{E}\left[\boldsymbol{h}_{jjk'tf}\boldsymbol{a}_{jk'tf}\right]\right|^2. \tag{5.22}$$

Using a similar method as in [8, eq. (13)] and (3.27), but with the precoding in (5.17), we can solve the expectation in (5.22) to give

$$\begin{aligned} S_{jk'tf} =& \frac{P_{\mathrm{d}}\beta_{jjk'}}{K}\left|\mathbb{E}\left[\left\|\hat{\boldsymbol{h}}_{jjk'tf}\right\|\right]\right|^2 \\ =& \frac{P_{\mathrm{d}}\beta_{jjk'}\xi\left(M\right)^2\alpha_{jjk'tf}}{K}, \end{aligned} \tag{5.23}$$

where $\xi\left(M\right)$ is defined in (3.27) in Chapter 3.

$N_{jk'tf}$ in (5.21) is the interference and noise for UE k' in cell j at time-frequency tf which equals the total received power minus the effective power, $N_{jk'tf} = \mathbb{E}\left[\left|x_{jk'tf}\right|^2\right] - S_{jk'tf}$. Using $x_{jk'tf}$ from (5.16), $N_{jk'tf}$ can be expanded as

$$\begin{aligned} N_{jk'tf} =& \sum_{l=1}^{L} P_{\mathrm{d}}\beta_{jlk'}\mathbb{E}\left[\left|\boldsymbol{h}_{jlk'tf}\boldsymbol{a}_{lk'tf}\right|^2\right] + \sum_{l=1}^{L}\sum_{k\neq k'}^{K} P_{\mathrm{d}}\beta_{jlk'}\mathbb{E}\left[\left|\boldsymbol{h}_{jlk'tf}\boldsymbol{a}_{lktf}\right|^2\right] + 1 \\ & - S_{jk'tf}. \end{aligned} \tag{5.24}$$

The first expectation term in (5.24) relates to downlink data signal power for UEs with the same pilot sequence. Other than the desired signal, there will also be downlink data transmissions from nearby cell BSs that will result in pilot contamination. The closed form of the expectation term can be obtained using a similar method to [8, eq. (14)], but using the precoding defined in (5.17). A similar derivation can also be seen in (3.29) in Chapter 3. The precoding $\boldsymbol{a}_{lk'tf}$ is correlated with $\boldsymbol{h}_{jlk'tf}$. Since the channel and its estimate are Gaussian, we can separate $\boldsymbol{h}_{jlk'tf}$ into terms that are correlated and uncorrelated with $\boldsymbol{a}_{lk'tf}$. Therefore, the second expectation term in (5.24) can be

written as

$$\begin{aligned}\mathbb{E}\left[\left|\boldsymbol{h}_{jlk'tf}\boldsymbol{a}_{lk'tf}\right|^2\right] =&\mathbb{E}\left[\left|\boldsymbol{h}_{jlk'tf}\frac{\hat{\boldsymbol{h}}_{llk'tf}^H}{\sqrt{K}\left\|\hat{\boldsymbol{h}}_{llk'tf}\right\|}\right|^2\right]\\ =&\mathbb{E}\left[\left|\hat{\boldsymbol{h}}_{jlk'tf}\frac{\hat{\boldsymbol{h}}_{llk'tf}^H}{\sqrt{K}\left\|\hat{\boldsymbol{h}}_{llk'tf}\right\|}\right|^2\right]+\mathbb{E}\left[\left|\tilde{\boldsymbol{h}}_{jlk'tf}\frac{\hat{\boldsymbol{h}}_{llk'tf}^H}{\sqrt{K}\left\|\hat{\boldsymbol{h}}_{llk'tf}\right\|}\right|^2\right]. \quad (5.25)\end{aligned}$$

Using (5.7) and (5.17), it can be shown that $\hat{\boldsymbol{h}}_{llk'tf} = \sqrt{\frac{\beta_{llk'}}{\beta_{jlk'}}}\hat{\boldsymbol{h}}_{jlk'tf}$. This also means that $\frac{\hat{\boldsymbol{h}}_{llk'tf}^H}{\left\|\hat{\boldsymbol{h}}_{llk'tf}\right\|} = \frac{\hat{\boldsymbol{h}}_{jlk'tf}^H}{\left\|\hat{\boldsymbol{h}}_{jlk'tf}\right\|}$. Using this, along with (5.14) and (5.15), (5.25) becomes

$$\begin{aligned}\mathbb{E}\left[\left|\boldsymbol{h}_{jlk'tf}\boldsymbol{a}_{lk'tf}\right|^2\right] =&\mathbb{E}\left[\left|\hat{\boldsymbol{h}}_{jlk'tf}\frac{\hat{\boldsymbol{h}}_{jlk'tf}^H}{\sqrt{K}\left\|\hat{\boldsymbol{h}}_{jlk'tf}\right\|}\right|^2\right]+\mathbb{E}\left[\left|\tilde{\boldsymbol{h}}_{jlk'tf}\frac{\hat{\boldsymbol{h}}_{jlk'tf}^H}{\sqrt{K}\left\|\hat{\boldsymbol{h}}_{jlk'tf}\right\|}\right|^2\right]\\ =&\frac{1}{K}\mathrm{tr}\left(\boldsymbol{R}_{\hat{\boldsymbol{h}}_{jlk'tf}\hat{\boldsymbol{h}}_{jlk'tf}}\right)+\frac{1}{K}\frac{\mathrm{tr}\left(\boldsymbol{R}_{\tilde{\boldsymbol{h}}_{jlk'tf}\tilde{\boldsymbol{h}}_{jlk'tf}}\right)}{M}\\ =&\frac{\alpha_{jlk'tf}}{K}M+\frac{1-\alpha_{jlk'tf}}{K}. \quad (5.26)\end{aligned}$$

The second expectation term in (5.24) is downlink data interference from BSs to UEs with different pilot sequences ($k \neq k'$). For the block fading assumption, the channel and the precoding in this expectation term are uncorrelated. However, in a selective fading channel, these variables can be correlated. This is because the extra interference term which arises from selective fading during the training (specifically, the second summation term in (5.6)) has made $\boldsymbol{a}_{lktf}$ becomes a function of $\boldsymbol{h}_{jlk'tf}$ and creates correlation between these variables. Since $\boldsymbol{h}_{jlk'tf}$ and $\hat{\boldsymbol{h}}_{jlktf}$ are Gaussian, we can express $\boldsymbol{h}_{jlk'tf}$ in terms of $\hat{\boldsymbol{h}}_{jlktf}$ as

$$\boldsymbol{h}_{jlk'tf} = \phi_{jlk'ktf}\frac{\hat{\boldsymbol{h}}_{jlktf}}{\sqrt{\left\|\boldsymbol{R}_{\hat{\boldsymbol{h}}_{jlktf}\hat{\boldsymbol{h}}_{jlktf}}\right\|}}+\sqrt{1-\phi_{jlk'ktf}^2}\boldsymbol{\epsilon}_{jlk'ktf}, \quad (5.27)$$

where $\phi_{jlk'ktf}$ is defined as the correlation between $\boldsymbol{h}_{jlk'tf}$ and $\hat{\boldsymbol{h}}_{jlktf}$. Both $\frac{\hat{\boldsymbol{h}}_{jlktf}}{\sqrt{\left\|\boldsymbol{R}_{\hat{\boldsymbol{h}}_{jlktf}\hat{\boldsymbol{h}}_{jlktf}}\right\|}}$ and $\boldsymbol{\epsilon}_{jlk'ktf}$ in (5.27) have $\mathcal{CN}(0, \boldsymbol{I}_M)$ distributions and are uncorrelated with each other. To obtain $\phi_{jlk'ktf}$, we first find the covariance between these two variables, $\mathbb{E}\left[\boldsymbol{h}_{jlk'tf}^H\hat{\boldsymbol{h}}_{jlktf}\right]$. $\hat{\boldsymbol{h}}_{jlktf}$ has a similar definition as in (5.7), but for UE $k \neq k'$. Using

(5.1) and eliminating uncorrelated terms, we can simplify the covariance to

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{h}_{jlk'tf}^{H}\hat{\boldsymbol{h}}_{jlktf}\right] =&\mathbb{E}\left[\theta_{jlktf}\boldsymbol{h}_{jlk'tf}^{H}\check{\boldsymbol{h}}_{lk}\right]\\ =&\theta_{jlktf}\sqrt{P_{\mathrm{p}}\beta_{jlk'}}\sum_{\{t_1,f_1\}=1}^{\{T_{\mathrm{p}},F\}}\psi_{kt_1f_1}^{*}\psi_{k't_1f_1}\mathbb{E}\left[\boldsymbol{h}_{jlk'tf}^{H}\boldsymbol{h}_{jlk't_1f_1}\right]\\ =&\theta_{jlktf}\sqrt{P_{\mathrm{p}}\beta_{jlk'}}\sum_{\{t_1,f_1\}=1}^{\{T_{\mathrm{p}},F\}}\psi_{kt_1f_1}^{*}\psi_{k't_1f_1}\frac{J_0\left(\sigma_{jlk}\left|t-t_1\right|\right)}{1-i\mu_{jlk}\left(f-f_1\right)}\boldsymbol{I}_M. \end{aligned}\tag{5.28}$$

The correlation is obtained using the following expression

$$\phi_{jlk'ktf}=\frac{\left\|\mathbb{E}\left[\boldsymbol{h}_{jlk't}^{H}\hat{\boldsymbol{h}}_{jlktf}\right]\right\|}{\sqrt{\left\|\mathbb{E}\left[\boldsymbol{h}_{jlk't}^{H}\boldsymbol{h}_{jlk't}\right]\right\|\left\|\mathbb{E}\left[\hat{\boldsymbol{h}}_{jlktf}^{H}\hat{\boldsymbol{h}}_{jlktf}\right]\right\|}}.\tag{5.29}$$

Using (5.28) and (5.13), $\phi_{jlk'ktf}$ becomes

$$\phi_{jlk'ktf}=\frac{\left\|\theta_{jlktf}\sqrt{P_{\mathrm{p}}\beta_{jlk'}}\sum_{\{t_1,f_1\}=1}^{\{T_{\mathrm{p}},F\}}\frac{\psi_{kt_1f_1}^{*}\psi_{k't_1f_1}J_0\left(\sigma_{jlk'}\left|t-t_1\right|\right)}{1-i\mu_{jlk'}\left(f-f_1\right)}\boldsymbol{I}_M\right\|}{\sqrt{\left\|\boldsymbol{I}_M\right\|\left|\theta_{jlktf}\right|^2\left\|\boldsymbol{R}_{\check{\boldsymbol{h}}_{lk}\check{\boldsymbol{h}}_{lk}}\right\|}}.\tag{5.30}$$

Using (5.11) and (5.30), $\phi_{jlk'ktf}^2$ becomes (5.20).

From (5.7) and (5.17), it can be shown that $\boldsymbol{a}_{lktf}=\frac{\hat{\boldsymbol{h}}_{jlktf}^{H}}{\sqrt{K}\left\|\hat{\boldsymbol{h}}_{jlktf}\right\|}$. Using this and (5.27), we expand the second expectation interference in (5.24) as

$$\begin{aligned}\mathbb{E}\left[\left|\boldsymbol{h}_{jlk'tf}\boldsymbol{a}_{lktf}\right|^2\right] =&\frac{\phi_{jlk'ktf}^2}{K\left\|\boldsymbol{R}_{\hat{\boldsymbol{h}}_{jlktf}\hat{\boldsymbol{h}}_{jlktf}}\right\|}\mathbb{E}\left[\left|\hat{\boldsymbol{h}}_{jlktf}\frac{\hat{\boldsymbol{h}}_{jlktf}^{H}}{\left\|\hat{\boldsymbol{h}}_{jlktf}\right\|}\right|^2\right]\\ &+\frac{1-\phi_{jlk'ktf}^2}{K}\mathbb{E}\left[\left|\boldsymbol{\epsilon}_{jlk'ktf}\frac{\hat{\boldsymbol{h}}_{jlktf}^{H}}{\left\|\hat{\boldsymbol{h}}_{jlktf}\right\|}\right|^2\right].\end{aligned}\tag{5.31}$$

Since $\hat{\boldsymbol{h}}_{jlktf}$ and $\boldsymbol{\epsilon}_{jlk'ktf}$ are Gaussian and uncorrelated with each other, it follows that

$$\begin{aligned}\mathbb{E}\left[\left|\boldsymbol{h}_{jlk'tf}\boldsymbol{a}_{lktf}\right|^2\right] =&\frac{\phi_{jlk'ktf}^2\mathrm{tr}\left(\boldsymbol{R}_{\hat{\boldsymbol{h}}_{jlktf}\hat{\boldsymbol{h}}_{jlktf}}\right)}{K\left\|\boldsymbol{R}_{\hat{\boldsymbol{h}}_{jlktf}\hat{\boldsymbol{h}}_{jlktf}}\right\|}+\frac{1-\phi_{jlk'ktf}^2}{K}\frac{\mathrm{tr}\left(\mathbb{E}\left[\boldsymbol{\epsilon}_{jlk'tf}^{H}\boldsymbol{\epsilon}_{jlk'tf}\right]\right)}{M}\\ =&\frac{\phi_{jlk'ktf}^2}{K}M+\frac{1-\phi_{jlk'ktf}^2}{K}.\end{aligned}\tag{5.32}$$

Substituting (5.26) and (5.31) into the first and second expectation term in (5.24), respectively, then (5.24) becomes

$$N_{jk'tf} = \sum_{l=1}^{L} \frac{P_{\mathrm{d}}\beta_{jlk'}}{K}\left(1+(M-1)\alpha_{jlk'tf}\right) + \sum_{l=1}^{L}\sum_{k\neq k'}^{K} \frac{P_{\mathrm{d}}\beta_{jlk'}}{K}\left(1+(M-1)\phi^2_{jlk'ktf}\right) + 1 - S_{jk'tf}, \tag{5.33}$$

where $\alpha_{jlk'tf}$ and $\phi^2_{jlk'ktf}$ are given in (5.19) and (5.20), respectively.

From (5.21), we obtain the sum rate for K UEs in cell j as

$$R_j = \sum_{k'=1}^{K} r_{jk'}. \tag{5.34}$$

By substituting (5.21), (5.23) and (5.33) into (5.34), the final sum rate becomes (5.18). □

Note that the result in (5.18)-(5.20) give the general form for the sum rate where the pilot values, ψ_{ktf}, are generic. Two specific pilot sequence methods are considered, which are discussed in the next section.

5.5 Training sequences and optimizations

5.5.1 Constant amplitude pilot

The use of constant amplitude (CA) pilots is a training method where the pilot amplitude is the same for every resource element during training, i.e. $|\psi_{ktf}|^2 = 1$ for all tf. A CA pilot approach is also considered in Chapter 4 for a time-selective channel. Here, we consider Zadoff-Chu sequences, which can be written as

$$\psi_{ktf} = \exp\left(\frac{-(t+T_{\mathrm{p}}f-k)\,(t+T_{\mathrm{p}}f-k+(K \mod 2))\,u\pi i}{K}\right), \tag{5.35}$$

where u is a constant integer with $0 < u < K$ and $\gcd(K,u) = 1$. Using (5.35), the pilot sequence will follow the properties in (5.3) and (5.4).

Using a CA pilot in (5.18) creates a non-zero $\phi^2_{jlk'ktf}$. Hence, for CA pilots, all interference terms from downlink data signal in (5.18) increase linearly with the number of

antennas. This means that pilot contamination-like interference can still occur with CA pilots even if all UEs have orthogonal pilot sequences.

Due to multiuser interference during the training phase for CA pilots, the total sum rate can be severely affected if some of the UEs have significantly higher fading compared to the other UEs. This condition naturally occurs due to differences in link distance and shadowing. If the pilot power received at the BS from certain UEs is much higher than other UEs, this will create high channel estimation error for the other UEs. To reduce this problem, we can either increase the power for UEs with low fading or reduce the power for UEs with high fading. The latter method will be more practical since it will reduce the power consumption. If the transmit pilot power from UE k in cell j is $P_{\mathrm{p},jk}$, then the received pilot power at the BS in cell j from this UE is $P_{\mathrm{p},jk}\beta_{jjk}$. We propose to use a simple power control method where the received pilot power for all UEs must not exceed a threshold value, $P_{\mathrm{threshold}}$. After $P_{\mathrm{p},jk}$ is initialized to a predetermined value, we perform the following procedure

$$\text{If} \quad P_{\mathrm{p},jk}\beta_{jjk} > P_{\mathrm{threshold}}, \text{ then } P_{\mathrm{p},jk} = \frac{P_{\mathrm{threshold}}}{\beta_{jjk}}. \tag{5.36}$$

Procedure (5.36) enables the received pilot power from each UE to be equal or less than $P_{\mathrm{threshold}}$.

5.5.2 Zero padded pilot

With zero padded (ZP) pilots, each UE will only use one resource element in each training block and leave the remaining area in the training block empty. A similar training sequence approach has been applied in [63] in a frequency selective channel. We assume that UE k' transmits a pilot signal at time-frequency $t'f'$ and transmits no signal for the other part of the training block as follows

$$\begin{aligned} \psi_{k'tf} &= \sqrt{K} \quad \text{for} \quad tf = t'f', \\ \psi_{k'tf} &= 0 \quad \text{for} \quad tf \neq t'f'. \end{aligned} \tag{5.37}$$

The purpose of setting the pilot value at time-frequency $t'f'$ to $\sqrt{K}$ is to fulfill the property in (5.3), which ensures that the average pilot power for a ZP pilot will be the same as that of a CA pilot. To ensure the pilots between UEs are orthogonal and

follow the property in (5.4), each UE will transmit at different time-frequencies. In other words, the combination of t' and f' must be different for each value of k' in order to avoid pilots overlapping. We set the values of t' and f' based on the value of k', as follows

$$t' = \big((k'-1) \bmod T_{\mathrm{p}}\big) + 1,$$
$$f' = \frac{k'-t'}{T_{\mathrm{p}}} + 1.$$

The advantage of ZP pilots compared to CA pilots is that there are no same cell UE pilots overlapping during training so that $\psi^*_{k'tf}\psi_{ktf} = 0$ for $k \neq k'$ at any tf. As a result, there will be no multiuser interference during training with ZP pilots and the second summation term in (5.6) will be zero. In addition, if we apply (5.37) to obtain $\phi_{jlk'ktf}$ in (5.20), we find that $\phi_{jlk'ktf} = 0$. This will simplify the sum rate equation. Such a simplification cannot be applied with CA pilots since $\phi_{jlk'ktf} \neq 0$ in this case. Hence, there can be more interference with CA pilots compared to ZP pilots.

Note that in the block fading model, the sum rate for both types of pilot sequence methods can be simplified to the same expression. Hence, a more complete comparison of pilot sequences can be achieved if selective fading in both time and frequency is considered.

5.5.3 Training optimization

Using the sum rate equation in (5.18), the training can be optimized by finding the training and data block sizes that yield the highest sum rate. The optimal training depends on the UE statistical properties such as large scale fading and channel correlation. Even though these statistical properties are assumed to be constant within a small scale time range, in the long term this values may vary due to the changes in UE position and speed. As a result of these variations, we consider two methods of finding the optimal training which are average optimum training and adaptive optimum training. For average optimum training, we find the optimal training size in order to obtain the highest

average sum rate. This can be expressed as

$$R_j^{\mathrm{av}} = \underset{\{T_{\mathrm{p}},F,T_{\mathrm{f}}\}}{\arg\max} \left(\bar{\mathbb{E}}\left[R_j\right]\right), \tag{5.38}$$

where $\bar{\mathbb{E}}\left[\cdot\right]$ is the average value across various possible large scale fading and channel correlation values. This approach requires the historical values of the statistical properties to obtain the long term average sum rate. This method does not require the training size to be updated regularly. However, since the total length of the pilot is affected by the number of UEs, this means that the average optimal training also needs to be updated if the total number of UEs changes.

For adaptive optimum training, the optimization depends on the current statistical property. Therefore, instead of obtaining the highest average sum rate as in (5.38), this method finds an optimal training size that yields the highest sum rate for every channel condition. This can be expressed as

$$R_j^{\mathrm{ad}} = \bar{\mathbb{E}}\left[\underset{\{T_{\mathrm{p}},F,T_{\mathrm{f}}\}}{\arg\max}\left(R_j\right)\right]. \tag{5.39}$$

Since (5.39) is adaptive to the channel conditions, it is expected to result in a better sum rate performance than the average optimum training method. However, this method requires the optimum training size to be computed and updated based on the current channel statistical values, which requires considerable extra processing and control overheads. This approach is far less practical and is essentially a benchmark of what can be achieved with no constraint on the overheads. Both the average and adaptive optimum training are found using an exhaustive search approach.

5.5.4 Frequency and pilot reuse

Frequency reuse is a method of reducing intercell interference by allocating different frequency band at different cells [11]. As we discussed in Chapter 2, if we use frequency reuse factor 1, all cells will use the same frequency band and this will create interference from all nearby cells. We can increase the frequency reuse factor to avoid this interference. The reduction in interference enables the transmission to increase the training block size and achieve higher spatial multiplexing gain. However, there will be a trade-off effect where the bandwidth that can be used in each cell is equal to the total bandwidth

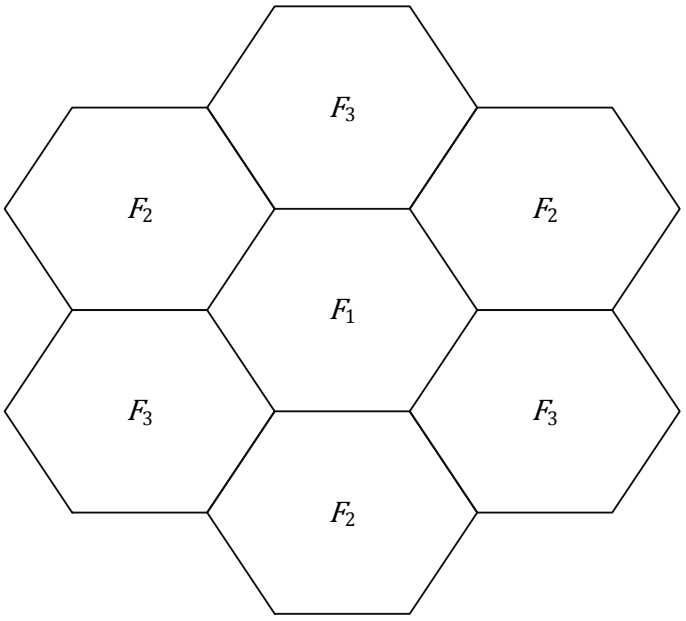

FIGURE 5.4: Cell arrangement with frequency reuse factor 3. F_1, F_2 and F_3 represents different frequency band that is used by each cells.

divides by the frequency reuse factor. Therefore, we aim to investigate this trade-off effect and study the impact of frequency reuse on the doubly selective channel. In the case of a hexagonal cell system, the interference from the nearest cells can be avoided by using frequency reuse factor 3, as shown in the Fig 5.4.

In addition to frequency reuse, intercell interference can also be reduced by increasing the length of pilot. This method does not eliminate the interference from the nearest cells unlike the frequency reuse method, but it can avoid pilot contamination from the nearest cell. To achieve this, the UEs from different cells must use different pilot sequences, which means the pilot size (training block) needs to be longer than the number of spatial multiplexed UEs per cell. In the case of a hexagonal cells system, we can avoid pilot contamination from the nearest cells by increasing the pilot size to at least 3 times the number of spatial multiplexed UEs per cell. This uses a similar cell arrangement to a frequency reuse factor of 3 in Fig 5.4. Note that the pilot reuse method will also experience a trade-off where the number of spatial multiplexed UEs can only be one-third the size of the pilot.

5.6 Numerical results

We use a 7 cell hexagonal cluster. All the large-scale fading values are obtained using the path-loss equation $\beta = zd^{-\gamma}$, where β is large scale fading, z is log-normal shadow fading, γ is the path-loss exponent and d is the distance between transmitter and receiver, normalized to the inner radius of the cell [6]. The path-loss exponent is set to 3.8. Transmission variables are given in Table 5.1, which are based on the LTE system [10], along with high speed UEs [57] and high delay spread [11]. In Table 5.1, we assume all UEs have the same speed and delay spread. Hence, the variables σ_{jlk} and μ_{jlk} in (5.1) are assumed to have constant values of σ_{c} and μ_{c} for all UEs. Using information from Table 5.1, we can obtain $\sigma_{\mathrm{c}} \approx 0.12$ and $\mu_{\mathrm{c}} \approx 0.28$.

TABLE 5.1: Transmission variables

Frequency spacing between subcarrier, Δv	15 kHz
OFDM symbol time, T_{s}	66.7 μs
Carrier frequency	2 GHz
Speed of UE	150 km/hour
Delay spread, τ_s	3 μs

We compare the closed form sum rate in (5.18) with Monte Carlo simulation. In this simulation, we consider equal distance UEs, where the UEs are located at 2/3 of the cell radius away from the BS, and use 0 dB log-normal shadowing ($z = 1$) [6]. The transmit power for all UEs and BSs is set to $P_{\mathrm{p}} = 10$ dB and $P_{\mathrm{d}} = 20$ dB [8]. We use $T_{\mathrm{p}} = 3$, $T_{\mathrm{d}} = 4$ and $F = 3$. To obtain the sum rate using Monte Carlo simulation, we first generate the complex random time-frequency channel with the correlation structure given in (5.1). Note that the generation of the channel with a predetermined correlation can be realized using a Cholesky decomposition of the channel correlation matrix [47]. For example, to generate a channel vector $\boldsymbol{v} \sim \mathcal{CN}(0, \boldsymbol{R_{vv}})$, we can use $\boldsymbol{v} = \boldsymbol{Au}$, where $\boldsymbol{A}$ is the lower triangular output of the Cholesky decomposition of the covariance matrix, $\boldsymbol{R_{vv}}$, and $\boldsymbol{u}$ is an i.i.d. $\mathcal{CN}(0, \boldsymbol{I}_N)$ vector. To generate the pilot sequences, we use (5.35) for CA pilots and (5.37) for ZP pilots. For every channel realization, we obtain the LMMSE channel estimate and the rate is found by averaging values in the expectations in (5.22) and (5.24) across many channel realizations. We then compare the Monte Carlo simulation with the closed form sum rate obtained from (5.18). Note that the sum rate in (5.18) is in bits/symbol/subcarrier. Values in Table 5.1 can be used to convert the sum rate to bits/second/Hz. Results are given in Fig. 5.5 and show that

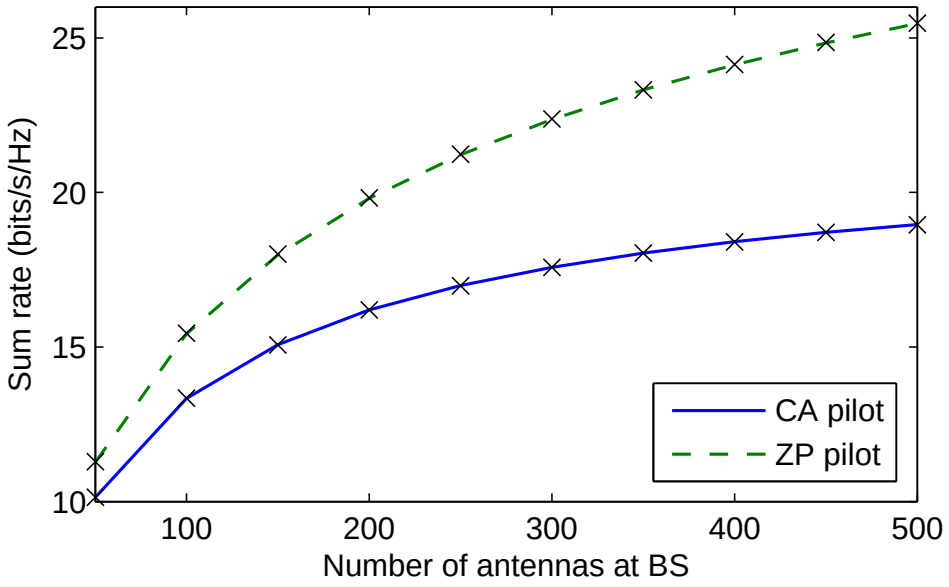

FIGURE 5.5: Sum rate performance using CA and ZP pilots. The lines are obtained using closed form expressions in (5.18) while points marked '×' are obtained using Monte Carlo simulation.

the closed form sum rate agrees well with Monte Carlo simulation. Also notice that ZP pilot has noticeably higher sum rate than CA pilot. This is because, as explained in Section 5.5.2, there are pilots overlapping between different UEs during CA training which create an extra interference term. Since there are no pilots overlapping during ZP pilot transmission, this means that there is less interference with ZP pilots compared to CA pilots.

For training optimization, a more complex path loss model with random UE locations and shadowing is considered. The sum rate is obtained using the closed form expression in (5.18). The UEs are located at uniform random locations within each cell excluding the area inside a 0.1 normalized distance from the BSs. The log-normal shadow fading has a standard deviation of 8 dB [5]. The final sum rate is calculated by averaging the sum rate across many possible UE locations and log-normal shadowing values. The UE cell allocation will be based on the power level between the UE and the BS [11], so that each UE will be located in the cell which has the largest received power. In addition to using the transmit pilot power value defined previously, we also use power control for the CA pilot where the transmit pilot power is adjusted using (5.36). The initial power is set to $P_{\mathrm{p},jk'} = 10$ dB and the threshold power, $P_{\mathrm{threshold}}$, is set to the optimum value (which yields the highest sum rate) using an exhaustive search. The values for T_{d}, T_{p} and F are

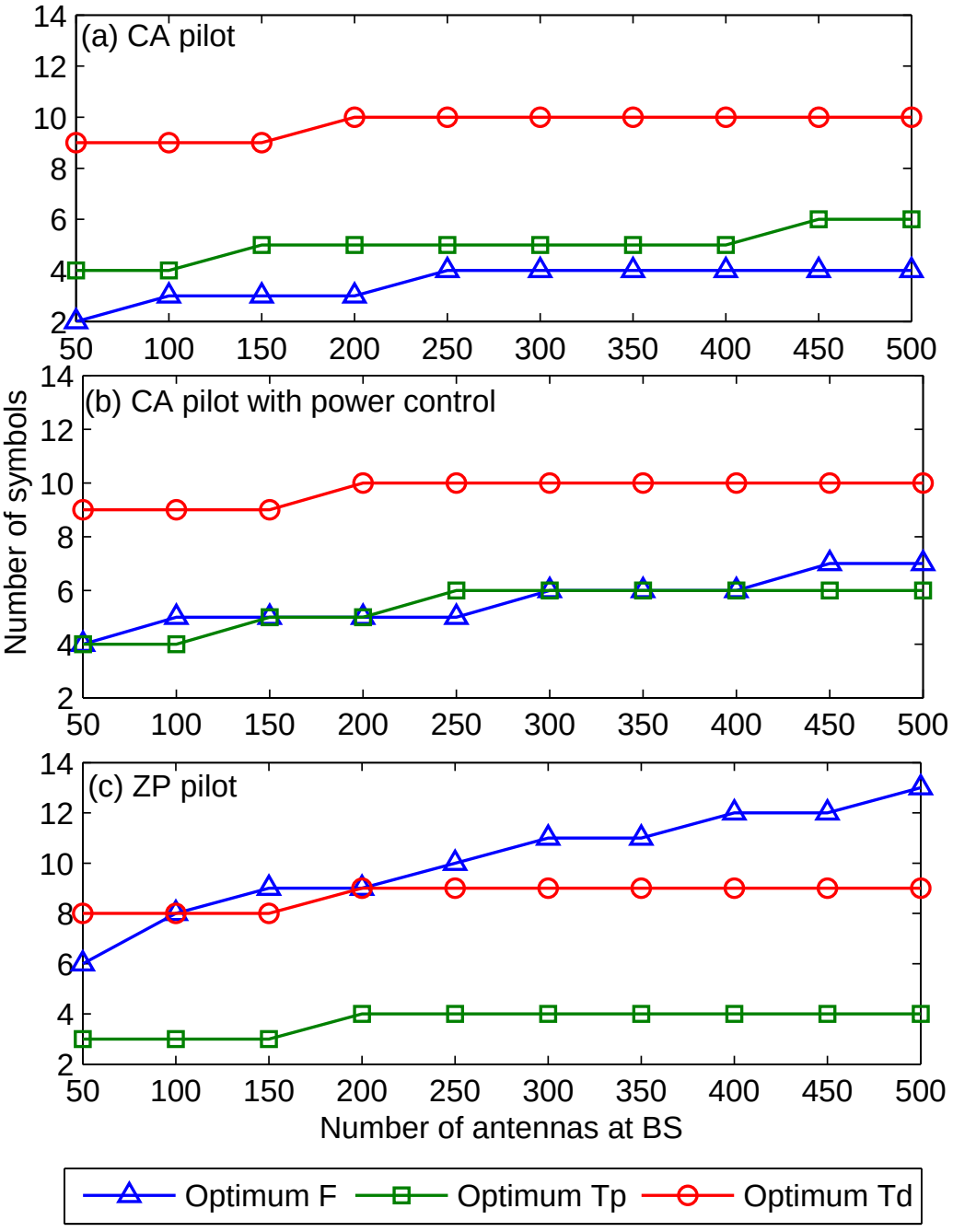

FIGURE 5.6: Optimum F, T_{p}, and T_{d} values for CA pilots, CA pilots with power control and ZP pilots.

optimized using the average optimum training method as in (5.38). We also assume the number of UEs is flexible, which means we can add or reduce UEs to achive the highest sum rate. Results are given in Fig. 5.6. In this figure, we can see that the optimal training and data block size increases as the number of antennas increases. This means that the higher the number of antennas, the more we can exploit the channel correlation to expand the pilot and data size in both time and frequency domains. This results

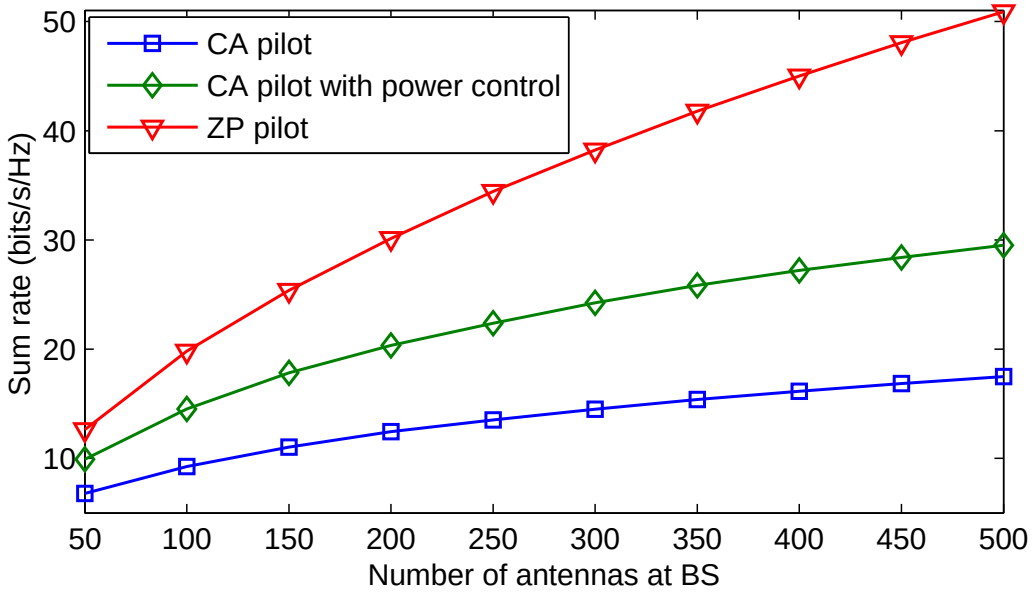

FIGURE 5.7: Sum rate performance for CA pilot, CA pilot with the power control and ZP pilot for various number of antennas.

in higher spatial multiplexing gains. For example, when the number of BS antennas increases from 50 to 500, the optimal number of spatial multiplexed UEs increases from 8 to 24 for CA pilots, from 16 to 42 for CA pilots with power control and from 18 to 52 for ZP pilots. The ZP approach shows a higher spatial multiplexing gain than the CA approach because there are no pilots overlapping during the training phase ZP pilots and this allows the scheme to expand the pilot with less interference compared to CA pilots.

Using the optimum training in Fig. 5.6, we obtain the sum rate results given in Fig. 5.7. The results show that power control can significantly improve the performance of the CA pilot. However, CA pilots have a much lower sum rate compared to ZP pilots even with power control. This is because, as we discussed in Section 5.5.2, there are no pilots overlapping during training with the ZP pilot scheme, which means there will be less interference from channel selectivity compared to CA pilots. Note that there is potential trade-off between the ZP and CA approaches. With ZP pilots, different UEs have a different time-frequency gap between their data and the pilot. The larger the pilot-data gap, the bigger the mismatch between the estimated channel and the true channel during the data phase. In order to reduce the pilot-data gap, an intermediate approach can be proposed where, for example, two UEs occupy two time-frequency slots using a Zadoff-Chu sequence and the remaining slots are zero padded. The other slots

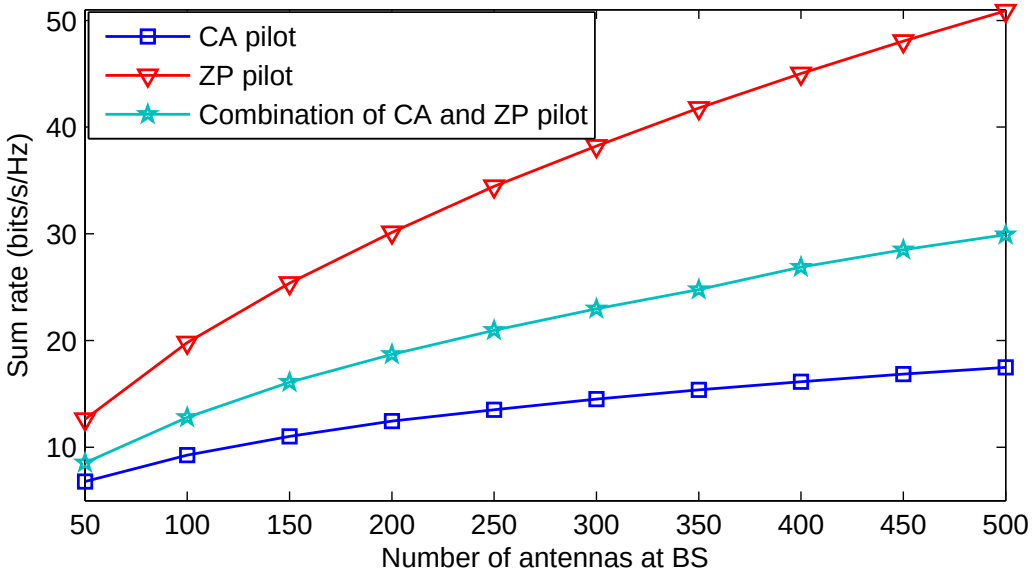

FIGURE 5.8: Sum rate performance for CA pilot, ZP pilot and combination of CA and ZP pilots for various number of antennas.

are then used by the pilots of other UE pairs. This allows the two users in each pair to spread their pilots over the pilot block and reduce the pilot-data gap. However, it also introduces pilot contamination-like interference as in the full CA approach. To realize the combined CA and ZP pilot method, we consider the use of only two resource elements in each training block and leave the remaining area empty (zero padded) for pilot transmission for each UE. As a result, two UEs must use the same two resource elements in each training block. To ensure orthogonality, these two UE pilots are valued based on Zadoff-Chu sequences in (5.35) with length of two. These pilots are positioned in the training block so that the total time-frequency distance between the pilot and the nearest data will be approximately the same for each UE. Results in Figure 5.8 show that combining CA and ZP pilot results in a sum rate performance that lies between that of a CA pilot and a ZP pilot. This means the trade-off is not beneficial and the extra interference outweighs the reduced pilot-data gap leading to worse performance than the ZP approach.

The results in Fig. 5.7 use the correlation values that are calculated from Table 5.1. Different UE speeds and delay spreads will result in different time-frequency correlation. Hence, we now analyze the performance of the pilot methods for various correlation values. We vary σ_c (with a constant $\mu_c = 0.28$) to analyze the impact of correlation variation in the time domain and vary μ_c (with a constant $\sigma_c = 0.12$) to analyze the

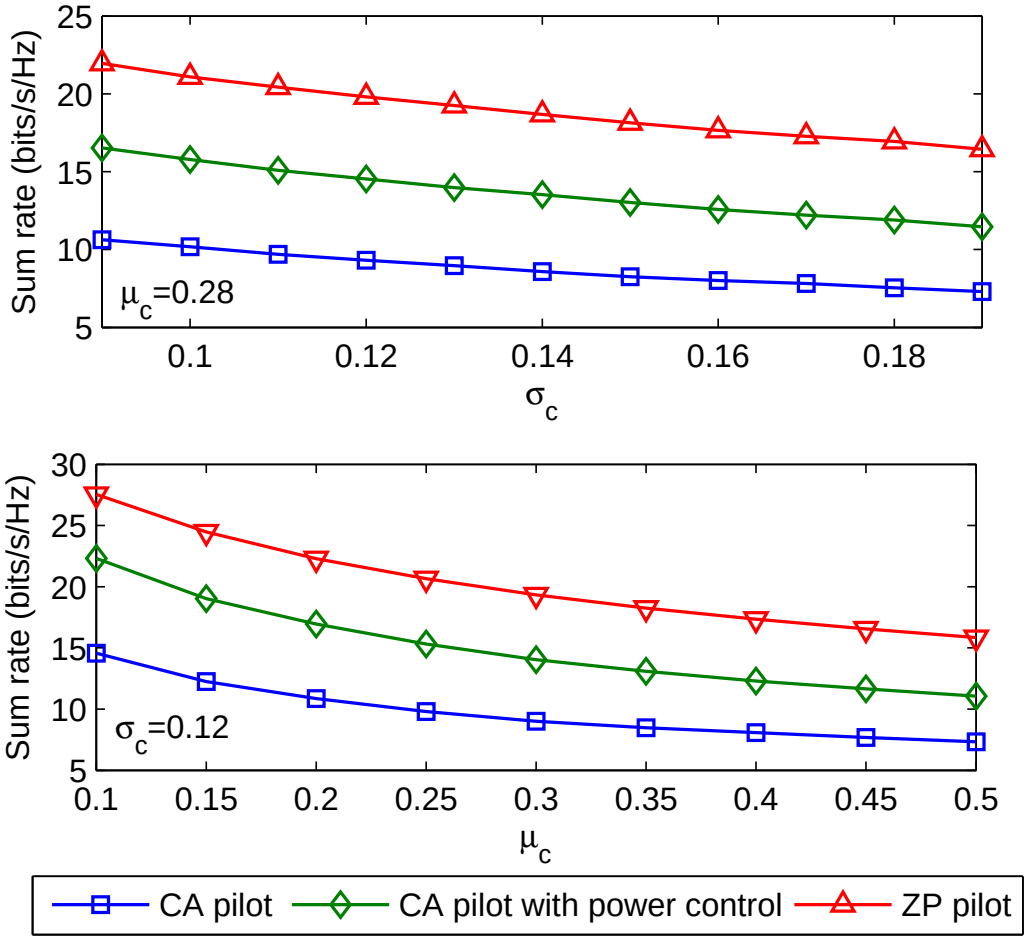

FIGURE 5.9: Sum rate performance for CA pilots, CA pilots with power control and ZP pilot for various σ_c and μ_c values using 100 BS antennas.

impact in the frequency domain. We use the optimum training and data length for each correlation value. The results using 100 BS antennas are given in Fig. 5.9. The results confirm that the sum rate for ZP pilots still outperforms that for CA pilots with power control, for all correlation values considered.

The results in Figs. 5.7 and 5.9 use average optimal training. As discussed in Section 5.5.3, we can also perform optimization using adaptive optimum training. Therefore, next we compare the performance between these two optimization methods. Note that Fig. 5.7 and 5.9 uses flexible number of UEs to obtain the highest average sum rate. In the next test we investigate the impact of the number of UEs with random speed on the training optimization. Each UE has a speed generated from a Gaussian distribution with 84 km/h mean and 19 km/h standard deviation [69]. ZP pilots and 100 BS antennas are used for this test. The result in Fig. 5.10 shows that adaptive optimum training has a slightly higher sum rate than average optimum training, but the difference is not

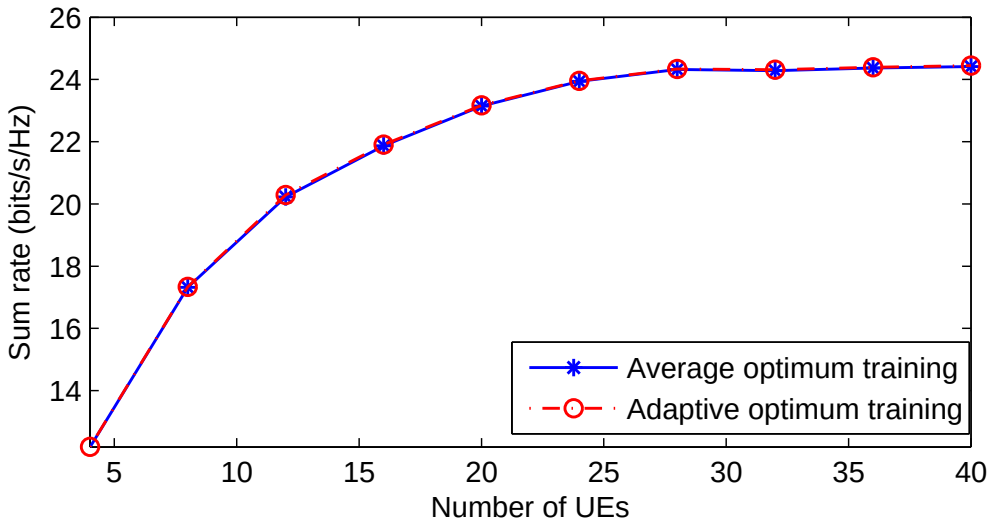

FIGURE 5.10: Sum rate performance for average optimal training and adaptive optimal training for various number of UEs using 100 BS antennas.

noticeable. Hence, we can use the average optimum training approach and still achieve a similar performance to the adaptive optimal training. This has a considerable impact on the system design since average optimum training is far more convenient and since it is a long term selection, an exhaustive search is a feasible solution.

Next, we analyze the impact of changing the frequency and pilot reuse factor on the performance, as discussed in Section 5.5.4. Previous results depended on using frequency and pilot reuse factor of 1, which means all cells use the same bandwidth and same pilot sequences. In this test, we use the ZP pilot method and the variable settings from Table 5.1. To obtain the performance of a frequency reuse factor of 3, we just eliminate the interference from the nearest cell and divide the final sum rate by 3. For a pilot reuse factor 3, we just need to ensure that there will be no pilot sequence reuse from the nearest cell. For each frequency and pilot reuse condition, we find the optimum training and data block size, which is given in Fig. 5.11. The results show that increasing the frequency and pilot reuse factor can also increase the optimum training and data size of the transmission. This is because a higher frequency and pilot reuse factor can reduce intercell interference which enables the training and data to expand in both time and frequency domains. However, this does not necessarily mean that such methods will result in a higher sum rate, as shown in Fig. 5.12. This result shows that using a frequency and a pilot reuse factor of 1 has a higher sum rate performance compared to a frequency reuse factor of 3 and pilot reuse factor of 3. This is because, in the case

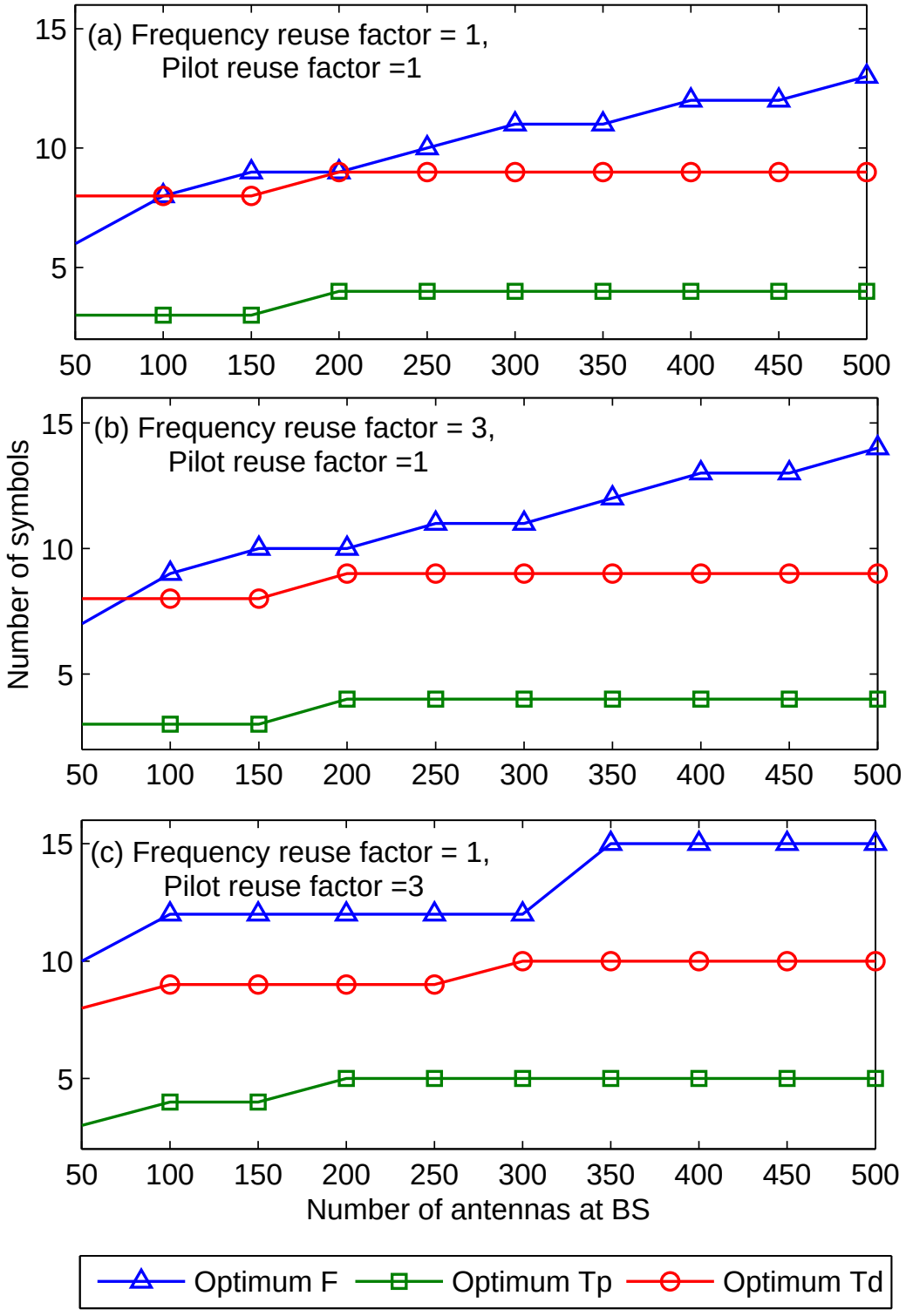

FIGURE 5.11: Optimum F, T_{p}, and T_{d} values for various frequency and pilot reuse factor.

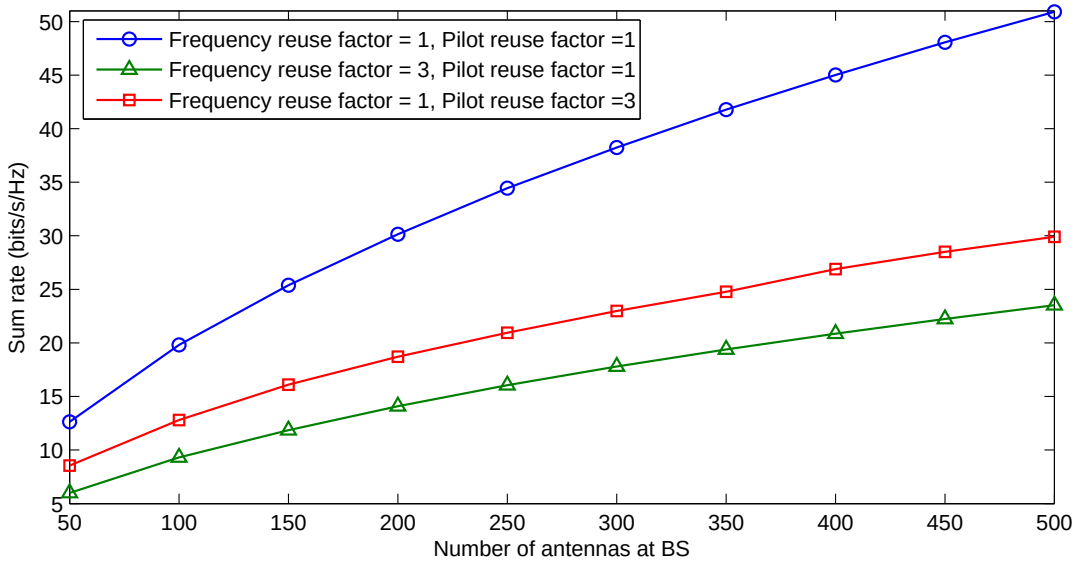

FIGURE 5.12: Sum rate performance for various frequency and pilot reuse factor.

of a frequency reuse factor of 3, even though there is no nearby intercell interference, each cell only uses one-third of the available bandwidth, which reduces the spectral efficiency. In the case of a pilot reuse factor of 3, even though pilot contamination from nearby cells is avoided, the number of spatial multiplexed UEs is only one-third of the training block size. In other words, both intercell interference reduction methods create trade-offs which can reduce the final sum rate of the transmission system.

5.7 Summary

In this chapter, a closed form achievable rate for a time and frequency selective channel model in a massive MIMO system had been derived. Two types of pilot sequence were studied, namely CA and ZP pilots. The results for both pilot methods showed that as the number of antennas increased, optimum training and data length in time and frequency domain also increased. A sum rate comparison shows that ZP pilots always have a higher sum rate than CA pilot for all numbers of antennas and time-frequency correlation values considered. Two different optimization methods, namely adaptive optimal training and average optimal training were also investigated. A sum rate comparison shows that both methods had similar performance. This chapter also investigated the impact of intercell interference reduction methods such as increasing the frequency reuse and pilot reuse

factors. However, these methods were shown to reduce the sum rate due to inefficient use of time and frequency resources.

Chapter 6

Conclusions and future works

6.1 Conclusions

In Chapter 3, we investigated the performance of synchronized and unsynchronized pilots in finite massive MIMO systems. We considered TShdown and TShup unsynchronized pilot or time-shifted pilot schemes. TShdown is where the pilots are overlapped with downlink data from nearby cells while TShup is where the pilots are overlapped with uplink data from nearby cells. Results showed that if there are a small numbers of UEs, TShdown generally had a higher sum rate than other methods. However, for higher numbers of UEs, TShup with power optimization had advantages over the other methods. The results also showed that compared to TShdown, the TShup with power optimization sum rate performance improvement with respect to TSync was more consistent when the number of UEs was increased. We also showed that in the case of a path-loss models with shadowing, that TSync is not necessarily the worst case scenario in terms of sum rate performance.

In Chapter 4, we studied the impact of channel variation in the time domain on the performance of massive MIMO systems. Using the first order Gauss-Markov Rayleigh fading channel model, we derived a closed form achievable rate for time-selective channels. Constant amplitude (CA) pilot was used. Asymptotic analysis showed that even if there was no pilot contamination from other cells, the intracell interference effect in time-selective channels didn't diminish when the number of antennas approached infinity. The numerical results showed that there is a significant difference between the sum

rate obtained using block-fading and time-selective models. We also showed that the optimum training for block-fading may not be optimal for a time-selective channel, particularly for large numbers of antennas at the BS. Due to these differences, it is important to consider the impact of channel selectivity in the time domain on the performance of massive MIMO systems. We showed that for time selective channel model, when the number of antennas becomes very large, the optimal frame length can grow beyond the coherence time. This result cannot be observed in block fading channel model because the frame length in this model is limited by the coherence time. We also show that the optimal length for pilot is between 20% to 35% of the frame length.

In Chapter 5, we analyzed the impact of time and frequency selective channels on the design of training for massive MIMO systems. To achieve this, we derived a closed form achievable rate for the doubly selective channel model. Two different pilot sequences were considered, namely the CA pilot and a zero padded (ZP) pilot. The results showed that in general, as the number of antennas increased, the optimum training block size and spatial multiplexing gain increased. Comparing the two pilot sequence schemes, the results also showed that the ZP pilot has a better sum rate performance than the CA pilot for a wide range of numbers of BS antennas and time-frequency correlation values. In addition to comparing the pilot schemes, Chapter 5 also compared the sum rate performance of two different training optimization methods, namely adaptive optimal training and average optimal training. Results showed that both methods have a similar sum rate performance. Chapter 5 also studied the effect of two intercell interference reduction methods, namely increasing the frequency reuse and pilot reuse factors. However, due to inefficient use of time-frequency resources, these methods lower the sum rate.

6.2 Future works

6.2.1 Spatial correlation

Throughout this thesis, we assume that there is no spatial correlation between antennas. This assumption comes from the Jakes' channel model [28] which shows that when there are an infinite number of multipath channels circularly surrounding the receiver, then zero spatial correlation between antennas can be achieved by using an appropriate

antenna spacing (see Section 2.2.1 in Chapter 2). Using this assumption, along with certain time-frequency channel correlation assumptions, the number of spatial multiplexing UEs can be upper bounded by the number of BS antennas (see Section 2.4 in Chapter 2). As there can be hundreds of BS antennas in massive MIMO, this means that there can also be hundreds of UEs to transmit parallel streams of data using the same time-frequency resources using spatial multiplexing or SDMA. Since we can potentially support hundreds of spatial multiplexed UEs under these conditions, this means that other multiple access schemes, namely OFDMA, will be unnecessary because dividing the time-frequency resources among the UEs will be inefficient compared to SDMA where the UEs can share the same time-frequency resources. However, this is only true when there is no spatial correlation between antennas. Realistically, there will be spatial correlation between antennas, such as in the case of limited numbers of multipath channels [7]. In addition, if the time-frequency channel correlation values are low due to high speed UEs and high channel delay spread, the channel training size can also be limited. As a result, it may not be possible to support hundreds of spatial multiplexed UEs in one cell. This means that other multiple access schemes such as OFDMA need to be included. To perform this analysis, we can use the training-based sum rate equation (5.18) developed for the doubly-selective channel in Chapter 5 and extend it to include the spatial correlation effect. The optimum combination of resource allocation schemes such as SDMA and OFDMA under different spatial, time and frequency channel correlation values will be an interesting topic for future research. In relation to spatially correlated channels, recent measurement of massive MIMO systems shows that the channel propagation is sparse in nature, which means there are few significant paths of propagation. Although this can reduce the number of spatial multiplexed UEs, this can also simplify the channel estimation process. Therefore, extension in this specific area can provide important insights on massive MIMO performance.

6.2.2 Millimeter wave

Other than massive MIMO systems, millimeter wave (mmWave) has also been proposed as a key component of next generation wireless technology [12]. The idea is to use a very high carrier frequency for wireless transmission in the range of tens to hundreds of Gigahertz. As spectrum is becoming a scare resource in microwave frequencies, researchers are looking into the untapped higher frequency range, namely the mmWave

realm. The mmWave and massive MIMO systems do not have to be two competing wireless technologies. In fact, both of them can be used to improve each other's limitations. For example, in the case of mmWave, one of the issues using frequencies at this range is high signal attenuation due to energy absorption from the surrounding medium. Since massive MIMO has been shown to be able to achieve higher SINR by directing the signal energy into a more concentrated area [27], this enables mmWave to compensate for its high power attenuation. In the case of massive MIMO, one of the issue is array size because the antennas need to be set with certain spacing based on the signal wavelength in order to reduce correlation. Since the wavelength in mmWave is within the millimeter range, this means that we can pack a lot of antennas within small physical space to achieve the massive MIMO range. For mmWave, high mobility UEs will be an issue because the small cell radius means there will be frequent handover between cells for fast moving UEs. One way to solve this is to use microwave (larger cell radius) for high mobility UE and use mmWave (smaller cell radius) for low mobility UEs. Another issue with mmWave is that since its cell radius is small, more BSs need to be constructed per unit area. This can increase the complexity and the cost of backhaul links between BSs. This issue can potentially be solved by using massive MIMO in the microwave range that acts as a backhaul connection between mmWave BSs. Note that since we are using mmWave for the small cells, there will be no interference issues between the small cells and the microwave links. The addition of small cells means there will be fewer UEs per BS. Furthermore, unlike the regular BS to UEs links, the wireless backhaul links between BS and the smaller cell BSs do not require regular channel training due to the static nature of the BSs. As a result, the addition of mmWave cells can potentially reduce the channel training size. To analyze its performance, we can extend the training-based analytical results from this thesis by including the mmWave cells. The idea of combining the conventional microwave, massive MIMO systems, small cells and mmWave is still not widely being investigated and therefore can be a worthy area for future research.

6.2.3 Cell radius

Even though there are many studies on massive MIMO systems in multicell scenarios, so far there has been no detailed study on the impact of cell radius length. Determining the size of a cell is an important part of designing the cellular network. A longer cell

radius is beneficial in terms of construction cost because there will be fewer base station towers to be built per unit area. There will also be a smaller concentration of backhaul links and this further reduces the cost of construction. Increasing the cell radius can also reduce the loss of performance due to handover between cells as the average length of the cell boundary per unit area is also reduced. However, a larger cell radius will also result in certain negative issues. For example, due to the exponential decay of power against distance, the cell edge UEs will suffer low SINR. Another issue is that if the number of UEs per unit area is constant, as the cell becomes bigger then there will be an increase in the number of UEs per BS, which consequently increases intracell interference and reduces time-frequency resources per UE. Such issues can be reduced by using massive MIMO which has the ability to improve the SINR at the receiver and support larger number of UEs. Since the changes in radius size can also effect the total number of UEs in a cell, this means that the radius size can also affect the channel training size. To study the impact of the cell radius on massive MIMO performance, we can use the training-based sum rate expression developed in this thesis. This allow us to study the relationship between the cell radius, the sum rate, the number of antennas at the BS and the channel training size.

6.2.4 Coding scheme

In this thesis, we depend on an information-theoretical approach to measure the transmission performance. As we discussed in Section 2.2 Chapter 2, this approach excludes the coding and modulation schemes in its performance evaluation by assuming that the data rate is able to achieve its theoretical upper limit (channel capacity). This approach has certain benefits such as the reduction of data rate calculation complexity. However, it also excludes an investigation of coding scheme design. Since our research involves channel variation in both time and frequency domains, it is interesting to extend this research into the design of coding schemes that are optimized for the channel condition in massive MIMO systems. We can also investigate how to include the time-frequency training optimization methods that are studied in this thesis into this research extension. In this thesis, we assume that the BS knows the statistical channel information in order to perform training optimization. Practically, we still need to acquire the statistical knowledge in some ways. Therefore, it can be beneficial to research methods of acquiring the statistical channel information. Specifically, we can study how much of

the time-frequency resource is needed to obtain the statistical information (i.e. through channel training) and how much performance improvement it can contribute to the wireless transmission.

Another way to make our analysis more realistic is to include inter-carrier interference. So far, in order to simplify the analysis, we assume that there is no inter-carrier interference due to the UEs mobility. The increase of inter-carrier interference from the UEs mobility and its impact on the performance will be an interesting topic for future research.

6.3 Summary

In this thesis, we have provided various novel contributions related to training in massive MIMO systems. Specifically, we provided a detailed analysis of the impact of unsynchronizing the pilots in massive MIMO. Furthermore, we developed a mathematical tool which allowed us to compare the performance of block-fading and time-selective models. We also extend the research to a doubly selective channel, where we studied the impact of different pilot schemes and training optimization methods. In addition to enhancing our understanding of massive MIMO systems through the numerical results, this thesis has provided various analytical results which can be used as tools for future work.

Appendix A

Proof of Lemma 3.2

If we condition on $\boldsymbol{y}$, then $\boldsymbol{w}$ and $\boldsymbol{Z}$ are both complex Gaussian with zero mean. Consider $\boldsymbol{w}$ and $\boldsymbol{z}_i$ (the i-th row of $\boldsymbol{Z}$). Conditional on $\boldsymbol{y}$, both $\boldsymbol{w}$ and $\boldsymbol{z}_i$ are correlated zero mean Gaussian vectors and it is straightforward to show

$$\boldsymbol{z}_i = u_i \boldsymbol{w} + v_i \boldsymbol{e}, \tag{A.1}$$

where $u_i = \frac{\sqrt{b}}{a+b} \frac{y_i}{\|\boldsymbol{y}\|}$, $v_i = \sqrt{1 - \frac{b}{a+b} \left| \frac{y_i}{\|\boldsymbol{y}\|} \right|^2}$ and $\boldsymbol{e}$ is a $\mathcal{CN}(0, \boldsymbol{I}_M)$ vector independent of $\boldsymbol{w}$ and y_i is the i-th element of $\boldsymbol{y}$.

Let S_{ij} be the i,j-th element of $\frac{\boldsymbol{Z}\boldsymbol{w}^H \boldsymbol{w}\boldsymbol{Z}^H}{\boldsymbol{w}\boldsymbol{w}^H}$ or $S_{ij} = \frac{\boldsymbol{z}_i \boldsymbol{w}^H \boldsymbol{w}\boldsymbol{z}_j^H}{\boldsymbol{w}\boldsymbol{w}^H}$. Since $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$ are independent, for non-orthogonal elements, we obtain

$$\mathbb{E}\left[S_{ij}\right] = 0, \text{ for } i \neq j. \tag{A.2}$$

Next, consider the expected value of a diagonal element, S_{ii}, which can be written as

$$\begin{aligned}
\mathbb{E}\left[S_{ii}\right] =& \mathbb{E}\left[\mathbb{E}\left[S_{ii}|\boldsymbol{y}\right]\right] \\
=& \mathbb{E}\left[\mathbb{E}\left[\frac{\left(u_i \boldsymbol{w} + v_i \boldsymbol{e}\right) \boldsymbol{w}^H \boldsymbol{w} \left(u_i^* \boldsymbol{w}^H + v_i^* \boldsymbol{e}^H\right)}{\boldsymbol{w}\boldsymbol{w}^H} | \boldsymbol{y}\right]\right] \\
=& \mathbb{E}\left[\mathbb{E}\left[\left|u_i\right|^2 \boldsymbol{w}\boldsymbol{w}^H + \frac{\left|v_i\right|^2 \operatorname{tr}\left(\boldsymbol{w}^H \boldsymbol{w}\right)}{\boldsymbol{w}\boldsymbol{w}^H} | \boldsymbol{y}\right]\right] \\
=& \mathbb{E}\left[\left|u_i\right|^2 \boldsymbol{w}\boldsymbol{w}^H\right] + \mathbb{E}\left[\left|v_i\right|^2\right].
\end{aligned} \tag{A.3}$$

By expanding $\boldsymbol{w}$ using (3.9), we obtain

$$\begin{aligned}\mathbb{E}\left[S_{ii}\right] =& \frac{b}{(a+b)^2}\mathbb{E}\left[\left|\frac{y_i}{\|\boldsymbol{y}\|}\right|^2 a\boldsymbol{x}\boldsymbol{x}^H + b\left|\frac{y_i}{\|\boldsymbol{y}\|}\right|^2 \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|}\boldsymbol{Z}\boldsymbol{Z}^H\frac{\boldsymbol{y}^H}{\|\boldsymbol{y}\|}\right] \\ &+ 1 - \frac{b}{a+b}\left|\frac{y_i}{\|\boldsymbol{y}\|}\right|^2. \end{aligned} \tag{A.4}$$

To solve (A.4), first we find the closed form solution of $\mathbb{E}\left[\left|\frac{y_i}{\|\boldsymbol{y}\|}\right|^2\right]$. This expectation can be written as

$$\begin{aligned}\mathbb{E}\left[\left|\frac{y_i}{\|\boldsymbol{y}\|}\right|^2\right] =& \mathbb{E}\left[\frac{|y_i|^2}{\boldsymbol{y}\boldsymbol{y}^H}\right] \\ =& \mathbb{E}\left[\frac{|y_i|^2}{|y_i|^2 + \sum_{n\neq i}^{M}|y_n|^2}\right] \\ =& \mathbb{E}\left[\frac{X}{X+Y}\right], \end{aligned} \tag{A.5}$$

where X and Y have a Chi-squared distribution with 2 and $2(M-1)$ degree of freedom, respectively. It is well-known that $\frac{X}{X+Y} \sim \beta(1, M-1)$, where $\beta(\cdot)$ represents a Beta distribution [70]. This distribution has an expected value of $\frac{1}{M}$. This means that

$$\mathbb{E}\left[\left|\frac{y_i}{\|\boldsymbol{y}\|}\right|^2\right] = \frac{1}{M}. \tag{A.6}$$

Using (A.6), along with the identities $\mathbb{E}\left[\boldsymbol{Z}\boldsymbol{Z}^H\right] = M\boldsymbol{I}_M$, $\mathbb{E}\left[\frac{\boldsymbol{y}}{\|\boldsymbol{y}\|}\frac{\boldsymbol{y}^H}{\|\boldsymbol{y}\|}\right] = 1$ and $\mathbb{E}\left[\boldsymbol{x}\boldsymbol{x}^H\right] = M$, (A.4) becomes

$$\mathbb{E}\left[S_{ii}\right] = \frac{b}{a+b} + 1 - \frac{b}{(a+b)\,M}. \tag{A.7}$$

Combining results in (A.2) and (A.7), we get (3.10).

Appendix B

Proof of uncorrelated channel estimation error

The cross covariance between the LMMSE channel estimate and its error can written as

$$\begin{aligned}\boldsymbol{R}_{\hat{\boldsymbol{h}}_{ljk'}\tilde{\boldsymbol{h}}_{ljk'}} =& \mathbb{E}\left[\hat{\boldsymbol{h}}_{ljk'}^{H}\tilde{\boldsymbol{h}}_{ljk'}\right] \\ =& \mathbb{E}\left[\hat{\boldsymbol{h}}_{ljk'}^{H}\left(\boldsymbol{h}_{ljk'}-\hat{\boldsymbol{h}}_{ljk'}\right)\right] \\ =& \mathbb{E}\left[\hat{\boldsymbol{h}}_{ljk'}^{H}\boldsymbol{h}_{ljk'}\right]-\mathbb{E}\left[\hat{\boldsymbol{h}}_{ljk'}^{H}\hat{\boldsymbol{h}}_{ljk'}\right]. \end{aligned} \tag{B.1}$$

Using (3.4), then (B.1) becomes

$$\boldsymbol{R}_{\hat{\boldsymbol{h}}_{ljk'}\tilde{\boldsymbol{h}}_{ljk'}} = \frac{\sqrt{p_{\mathrm{r}}\beta_{ljk'}}}{\alpha_{jk'}^{\mathrm{TD}}}\mathbb{E}\left[\check{\boldsymbol{h}}_{jk'}^{H}\boldsymbol{h}_{ljk'}\right]-\boldsymbol{R}_{\hat{\boldsymbol{h}}_{ljk'}\hat{\boldsymbol{h}}_{ljk'}}. \tag{B.2}$$

Using (3.3), and eliminating uncorrelated terms in $\check{\boldsymbol{h}}_{jk'}$, then (B.2) becomes

$$\begin{aligned}\boldsymbol{R}_{\hat{\boldsymbol{h}}_{ljk'}\tilde{\boldsymbol{h}}_{ljk'}} =& \frac{\sqrt{p_{\mathrm{r}}\beta_{ljk'}}}{\alpha_{jk'}^{\mathrm{TD}}}\mathbb{E}\left[\sqrt{p_{\mathrm{r}}\beta_{ijk'}}K\boldsymbol{h}_{ljk'}^{H}\boldsymbol{h}_{ljk'}\right]-\boldsymbol{R}_{\hat{\boldsymbol{h}}_{ljk'}\hat{\boldsymbol{h}}_{ljk'}} \\ =& \frac{p_{\mathrm{r}}\beta_{ljk'}K}{\alpha_{jk'}^{\mathrm{TD}}}\boldsymbol{I}_{M}-\frac{p_{\mathrm{r}}\beta_{ljk'}K}{\alpha_{jk'}^{\mathrm{TD}}}\boldsymbol{I}_{M} \\ =& \boldsymbol{0}_{M}. \end{aligned} \tag{B.3}$$

Since $\boldsymbol{R}_{\hat{\boldsymbol{h}}_{ljk'}\tilde{\boldsymbol{h}}_{ljk'}} = \boldsymbol{0}_{M}$, this means $\hat{\boldsymbol{h}}_{ljk'}$ is uncorrelated with $\tilde{\boldsymbol{h}}_{ljk'}$.

Appendix C

Lemma C.1

Lemma C.1. Let $\boldsymbol{v}$ and $\boldsymbol{w}$ be $1 \times M$ vectors with $\mathcal{CN}(0, \boldsymbol{I}_M)$ distributions. If $\boldsymbol{v}$ and $\boldsymbol{w}$ are independent from each other, then

$$\mathbb{E}\left[\left|\boldsymbol{v}\frac{\boldsymbol{w}^H}{\|\boldsymbol{w}\|}\right|^2\right] = \frac{1}{M}\mathrm{tr}\left(\boldsymbol{R}_{\boldsymbol{vv}}\right). \tag{C.1}$$

Proof.

$$\begin{aligned}\mathbb{E}\left[\left|\boldsymbol{v}\frac{\boldsymbol{w}^H}{\|\boldsymbol{w}\|}\right|^2\right] =& \mathbb{E}\left[\boldsymbol{v}\frac{\boldsymbol{w}^H}{\|\boldsymbol{w}\|}\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}\boldsymbol{v}^H\right] \\ =& \mathbb{E}\left[\mathrm{tr}\left(\boldsymbol{v}\frac{\boldsymbol{w}^H}{\|\boldsymbol{w}\|}\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}\boldsymbol{v}^H\right)\right].\end{aligned} \tag{C.2}$$

Using known properties of the trace operator, (C.2) becomes

$$\mathbb{E}\left[\left|\boldsymbol{v}\frac{\boldsymbol{w}^H}{\|\boldsymbol{w}\|}\right|^2\right] = \mathbb{E}\left[\mathrm{tr}\left(\boldsymbol{v}^H\boldsymbol{v}\frac{\boldsymbol{w}^H}{\|\boldsymbol{w}\|}\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}\right)\right].$$

Since $\boldsymbol{v}$ is independent from $\boldsymbol{w}$, and since $\mathbb{E}\left[\frac{\boldsymbol{w}^H}{\|\boldsymbol{w}\|}\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}\right] = \frac{1}{M}\boldsymbol{I}_M$ (see (A.6) in Appendix A), we get

$$\begin{aligned}\mathbb{E}\left[\left|\boldsymbol{v}\frac{\boldsymbol{w}^H}{\|\boldsymbol{w}\|}\right|^2\right] =& \mathrm{tr}\left(\mathbb{E}\left[\boldsymbol{v}^H\boldsymbol{v}\right]\mathbb{E}\left[\frac{\boldsymbol{w}^H}{\|\boldsymbol{w}\|}\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}\right]\right) \\ =& \frac{1}{M}\mathrm{tr}\left(\boldsymbol{R}_{\boldsymbol{vv}}\right).\end{aligned} \tag{C.3}$$

□

Appendix D

Power optimization

D.1 Power optimization for TShdown

Since $R_{jk'}^{\mathrm{TD,DL}} = \frac{T_{\mathrm{d}}}{T} \times \mathrm{C}\left(\frac{S_{\mathrm{a}}}{N_{\mathrm{a}}}\right)$, the maximization of $R_{jk'}^{\mathrm{TD,DL}}$ is the same as the maximization of the SNIR, $\frac{S_{\mathrm{a}}}{N_{\mathrm{a}}}$, with respect to p_{f}. Therefore, we can find the optimum value of p_{f} by solving $\frac{d}{dp_{\mathrm{f}}}\left(\frac{S_{\mathrm{a}}}{N_{\mathrm{a}}}\right) = 0$ with $p_{\mathrm{f}} > 0$. However, the SINR for downlink transmission includes the signal terms related to the precoding vector from other cells, which is $\alpha_{lk'}^{\mathrm{TD}}$ for $l \neq j$. Since $\alpha_{lk'}^{\mathrm{TD}}$ contains the term p_{f}, this makes the differentiation process complicated. To solve this, we assume that $\alpha_{lk'}^{\mathrm{TD}}$ in the interfering cells is equal to that of the target cell, j ($\alpha_{lk'}^{\mathrm{TD}} = \alpha_{jk'}^{\mathrm{TD}}$). We also perform the approximation, $\alpha_{jk'}^{\mathrm{TD}} \approx \sum_{i\in A_1} p_{\mathrm{r}}\beta_{ijk'}K + \sum_{i\notin A_1}\sum_{k=1}^{K}\frac{p_{\mathrm{f}}\delta_{ji}}{K}\left(1+p_{\mathrm{f}}\delta_{ji}\right)+1$. The results in Figure 3.16 show that this approximation is a reliable method to obtain the optimum downlink power. Using these approximations along with (3.21) and (3.22), we find the optimum power as the solution of

$$\frac{d}{dp_{\mathrm{f}}}\left(\frac{\frac{p_{\mathrm{f}}p_{\mathrm{r}}\beta_{jjk'}^2\xi(M)^2}{\alpha_{jk'}^{\mathrm{TD}}}}{\left(\begin{array}{c}\sum_{l\in A_1}\frac{p_{\mathrm{f}}\beta_{jlk'}}{K}\left(1+\frac{p_{\mathrm{r}}\beta_{jlk'}K(M-1)}{\alpha_{jk'}^{\mathrm{TD}}}\right)\\ -\frac{p_{\mathrm{f}}p_{\mathrm{r}}\beta_{jjk'}^2\xi(M)^2}{\alpha_{jk'}^{\mathrm{TD}}}+\sum_{i\in A_2}\sum_{k=1}^{K}p_{\mathrm{r}}\sigma_{jik'k}\\ +\sum_{i\in A_3}\sum_{k=1}^{K}\frac{p_{\mathrm{f}}\beta_{jik'}}{K}+1\end{array}\right)}\right)=0, \tag{D.1}$$

where $\alpha_{jk'}^{\mathrm{TD}} \approx \sum_{i\in A_1} p_{\mathrm{r}}\beta_{ijk'}K + \sum_{i\notin A_1}\sum_{k=1}^{K}\frac{p_{\mathrm{f}}\delta_{ji}}{K}\left(1+p_{\mathrm{f}}\delta_{ji}\right)+1$ and $p_{\mathrm{f}} > 0$. Solving (D.1), we get (3.50).

D.2 Power optimization for TShup

Similar to Appendix D.1, we can obtain the optimum power for TShup using the associated SINR. However, since TShup uses the uplink transmit data power optimization (see Section 3.4.4), we use the uplink SINR, which is $\frac{S_{\mathrm{d}}}{N_{\mathrm{d}}}$. Similar to Appendix D.1, the optimum power can be obtained by solving $\frac{d}{dp_{\mathrm{u}}}\left(\frac{S_{\mathrm{d}}}{N_{\mathrm{d}}}\right)=0$ with $p_{\mathrm{u}} > 0$. Using (3.70) and (3.71), we get the optimum power as the solution of

$$\frac{d}{dp_{\mathrm{u}}}\left(\frac{\frac{p_{\mathrm{u}}p_{\mathrm{r}}\beta_{jjk'}^{2}K(M-1)}{\alpha_{jk'}^{\mathrm{TU}}}}{\begin{array}{c}p_{\mathrm{u}}\beta_{jjk'}\left(1-\frac{p_{\mathrm{r}}\beta_{jjk'}K}{\alpha_{jk'}^{\mathrm{TU}}}\right)+\sum_{l\in A_1\cap l\neq j}p_{\mathrm{u}}\beta_{ljk'}\left(1+\frac{p_{\mathrm{r}}\beta_{ljk'}K(M-2)}{\alpha_{jk'}^{\mathrm{TU}}}\right)\\ +\sum_{i\in A_1}\sum_{k\neq k'}^{K}p_{\mathrm{u}}\beta_{ijk}+\sum_{f\in A_2}\sum_{k=1}^{K}p_{\mathrm{r}}\beta_{fjk}\left(1+\frac{p_{\mathrm{u}}\beta_{fjk}(M-2)}{\alpha_{jk'}^{\mathrm{TU}}}\right)\\ +\sum_{f\in A_3}\sum_{k=1}^{K}p_{\mathrm{u}}\beta_{fjk}\left(1+\frac{p_{\mathrm{u}}\beta_{fjk}(M-2)}{\alpha_{jk'}^{\mathrm{TU}}}\right)+1\end{array}}\right)=0, \quad \text{(D.2)}$$

where $\alpha_{jk'}^{\mathrm{TU}} = \sum_{i\in A_1} p_{\mathrm{r}}\beta_{ijk'}K + \sum_{i\notin A_1}\sum_{k=1}^{K} p_{\mathrm{u}}\beta_{ijk}+1$ and $p_{\mathrm{u}} > 0$. Solving (D.2), we get (3.79).

Bibliography

[1] F. Rusek, D. Persson, B. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays", *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40-60, Jan. 2013.

[2] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 139-157, Jan. 1999.

[3] F. Fernandes, A. Ashikhmin, and T. L. Marzetta, "Inter-cell interference in noncooperative TDD large scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no.2, pp. 192-201, Feb. 2013.

[4] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590-3600, Nov. 2010.

[5] H. Ngo, E. Larsson, and T. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436-1449, Apr. 2013.

[6] J. Hoydis, S. ten Brink, M. Debbah, "Massive MIMO in UL/DL of cellular networks: How many antennas do we need?," *IEEE J. Sel. Areas Commun.*, vol. 31, pp. 160-171, Feb. 2013.

[7] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "The multicell multiuser MIMO uplink with very large antenna arrays and a finite-dimensional channel," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2350-2361, Jun. 2013.

[8] J. Jose, A. Ashikhmin, T. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640-2651, Aug. 2011.

[9] Cisco, "Global Mobile Data Traffic Forecast Update," *2014-2019 White Paper*, Feb. 2015.

[10] E. Dahlman, S. Parkvall, J. Sköld, *4G: LTE/LTE-Advanced for Mobile Broadband*, Elsevier Science & Technology, 2011.

[11] A. F. Molisch, *Wiley: Wireless Communications, 2nd Edition*, John Wiley and Sons, 2011.

[12] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, pp. 74–80, Feb. 2014.

[13] J. G. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065-1082, Jun. 2014.

[14] C.-X. Wang, F. Haider, X. Gao, X.-H. You, Y. Yang, D. Yuan, H. Aggoune, H. Haas, S. Fletcher, and E. Hepsaydir, "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 122-130, Feb. 2014.

[15] M. D. Renzo, H. Haas, A. Ghrayeb, S. Sugiura, and L. Hanzo, "Spatial Modulation for Generalized MIMO: Challenges, Opportunities, and Implementation," in *Proc.of the IEEE*, vol.102, no.1, pp.56-103, Jan. 2014.

[16] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. T. Sukhavasi, C. Patel, and S. Geirhofer, "Network densification: The dominant theme for wireless evolution into 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 82-89, Feb. 2014.

[17] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," *Bell Syst. Tech. J.*, vol. 1, no. 2, pp. 41-59, 1996.

[18] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, no. 3, pp. 311-335, Mar. 1998.

[19] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bolcskei, "An overview of MIMO communications - A key to gigabit wireless," *Proc. IEEE*, vol. 92, no. 2, pp. 198–218, Feb. 2004.

[20] J. Winters, "On the capacity of radio communication systems with diversity in a rayleigh fading environment," *IEEE J. Sel. Areas Commun.*, vol. 5, pp. 871-878, Jun. 1987.

[21] E. Telatar, "Capacity of multi-antenna gaussian channels," *AT&T-Bell Labs Internal Memo.*, pp. 585–595, Jun. 1995.

[22] G. G. Rayleigh and J. M. Cioffi, "Spatio-Temporal Coding for Wireless Communication," *IEEE Trans. Commun.*, vol. 46, no. 3, pp. 357-366, 1998.

[23] Bernard Sklar "Rayleigh fading channels in mobile digital communication systems part I: Characterization," *IEEE Commun. Mag.* vol. 35, no. 9, pp. 136-146, 1997.

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

[25] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, pp. 1451-1458, Oct. 1998.

[26] T. Marzetta, "How much training is required for multiuser MIMO?", in *Proc. Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, pp. 359-363, Oct. 2006.

[27] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[28] W. C. Jakes, D. C. Cox, *Microwave Mobile Communications*. Hoboken, NJ: Wiley-IEEE Press, 1994.

[29] T. S. Rappaport, *Wireless Communications: Principles and Practice*, Upper Saddle River, NJ: Prentice-Hall Inc., 2002.

[30] C. Shannon, "Communication in the presence of noise," *Proc. Institute of Radio Engineers*, vol. 37, pp. 10-21, Jan. 1949.

[31] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding," in *Proc. of ICC*, Geneva, Switzerland, pp. 1064-1070, May 1993.

[32] A. Goldsmith, *Wireless Communications*, Cambridge, U.K.: Cambridge Univ. Press, 2004.

[33] C. Shepard, H. Yu, N. Anand, L. E. Li, T. L. Marzetta, R. Yang, and L. Zhong, "Argos: Practical many-antenna base stations", *in ACM Int. Conf.Mobile Computing and Networking (MobiCom)*, Istanbul, Turkey, Aug. 2012.

[34] Z. Pi, F. Khan, "An introduction to millimeter-wave mobile broadband systems," in *IEEE Commun. Mag.*, vol.49, no.6, pp.101-107, Jun. 2011.

[35] D. Gesbert, M. Kountouris, R. W. Heath Jr., C. Chae, and T. Salzer, "From single user to multiuser communications: shifting the MIMO paradigm," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 36-46, Sep. 2007.

[36] A. Molisch, M. Win, and J. H. Winters, "Reduced-complexity transmit/receive-diversity systems," IEEE Trans. Signal Proc., vol. 51, pp. 2729–2738, November 2003.

[37] A. Scaglione, G.B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers. I. Unification and optimal designs," *IEEE Trans. Sign. Proc*, Vol. 47, No. 7, pp. 1988 - 2006, July 1999.

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

[39] R. Müller, M. Vehkaperä, and L. Cottatellucci, "Blind pilot decontamination," *Proc. ITG Wksp. Smart Antennas*, Stuttgart, Mar. 2013.

[40] J. Jaldén and B. Ottersten, "On the complexity of sphere decoding in digital communications," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1474-1484, Apr. 2005.

[41] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, pp. 742–758, Oct. 2014.

[42] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inform. Theory*, vol. 60, no. 11, pp. 7112–7139, Nov. 2014.

[43] G. S. Smith, "A direct derivation of a single-antenna reciprocity relation for the time domain," in *IEEE Trans. Antennas and Propagation*, vol.52, no.6, pp. 1568-1577, Jun. 2004.

[44] G. Jiann-Ching and L. D. Larsson, "Modeling and evaluation of MIMO systems exploiting channel reciprocity in TDD mode", in *IEEE Veh.Technol, Conf.*, Los Angeles, CA, vol. 6, pp. 4265-4269, Sep. 2004.

[45] S. Sesia, I. Toufik, and M. Baker, Eds., *LTE: The UMTS Long Term Evolution*, West Sussex, UK: John Wiley and Sons, 2009.

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

[47] Y. S. Cho, J. Kim, W. Young, and C. G. Kang, *MIMO-OFDM Wireless Communications with MATLAB*, Hoboken, NJ: John Wiley & Sons, Oct. 2010.

[48] J. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Commun. Mag.* Vol. 28, No. 5, pp. 5-14, May 1990.

[49] S. Jin, X. Wang, Z. Li, and K. K. Wong, "Zero-forcing beamforming in massive MIMO systems with time-shifted pilots", in *Proc. of ICC*, Sydney, Australia, pp. 4801-4806, Jun. 2014.

[50] N. L. Johnson, S. Kotz and N. Balakrishnan, *Continuous Univariate Distributions*, Vol 1. New York: John Wiley and Sons, 1994.

[51] A. Tulino and S. Verdú, *Random matrix theory and wireless communications*, Delft, Netherlands: Now Publishers Inc., 2004.

[52] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264-273, Feb. 2013.

[53] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "The multicell multiuser MIMO uplink with very large antenna arrays and a finite-dimensional channel," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2350–2361, Jun. 2013.

[54] N. Shariati, E. Björnson, M. Bengtsson, and M. Debbah, "Low-complexity polynomial channel estimation in large-scale MIMO with arbitrary statistics," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 815-830, Oct. 2014.

[55] M. Dong, L. Tong, and B. Sadler, "Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1403-1418, May 2004.

[56] V. Pohl, P. H. Nguyen, V. Jungnickel, and C. von Helmolt, "Continuous flat-fading MIMO channels: Achievable rate and optimal length of the training and data phases," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 1889-1900, Jul. 2005.

[57] X. Ma, G. B. Giannakis, and S. Ohno, "Optimal training for block transmissions over doubly selective wireless fading channels," *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1351-1366, May 2003.

[58] S. Savazzi and U. Spagnolini, "Optimizing training lengths and training intervals in time-varying fading channels," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1098-1112, Mar. 2009.

[59] M. Medard, "The effect upon channel capacity in wireless communication of perfect and imperfect knowledge of the channel," *IEEE Trans. Inform. Theory*, vol. 46, pp. 933–946, May 2000.

[60] K. T. Truong, A. Lozano, and R. W. Heath Jr., "Optimal training in continuous flat-fading massive MIMO systems" *in Proc. European Wireless Conference*, Barcelona, Spain, pp. 219-224, May 2014.

[61] W. A. W. M. Mahyiddin, P. A. Martin, P. J. Smith "Massive MIMO Systems in Time Selective Channels," *IEEE Commun. Lett.*, pp. 1-4, Sep. 2015.

[62] J. Hoydis, M. Kobayashi, and M. Debbah, "Optimal channel training in uplink network MIMO systems," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2824-2833, 2011.

[63] X. Ma, L. Yang, and G. B. Giannakis, "Optimal training for MIMO frequency-selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 453–466, Mar. 2005.

[64] H. Minn and N. Al-Dhahir, "Optimal training signals for MIMO OFDM channel estimation," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1158-1168, May 2006.

[65] A. P. Kannu and P. Schniter, "Design and analysis of MMSE pilot-aided cyclic-prefixed block transmissions for doubly selective channels," *IEEE Trans. Signal Process.*, vol. 56, pp. 1148-1160, Mar. 2008.

[66] L. Dai, Z. Wang, Z. Yang, "Spectrally efficient time-frequency training OFDM for mobile large-scale MIMO systems," *IEEE J. Sel. Areas Commun.* , vol.31, no.2, pp. 251-263, Feb. 2013.

[67] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1615-1624, Jun. 2003.

[68] W. A. W. M. Mahyiddin, P. A. Martin, P. J. Smith, "Performance of synchronized and unsynchronized pilots in finite massive MIMO systems," *IEEE Trans. Wireless Commun.*, no. 99, Jul. 2015.

[69] E. T. Donnell, S. C. Hines, K. M. Mahoney, R. J. Porter, H. McGee, *Speed Concepts: Informational Guide*, U.S. Department of Transportation, Federal Highway Administration, Washington D.C., 2009.

[70] M. Abramovitz and I. A. Stegun, *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables, 9th Ed.*, New York: Dover, pp. 944, 1972.