



VICTORIA UNIVERSITY OF WELLINGTON

MASTERS THESIS

**Optimising the Massive MIMO Downlink
in the Presence of Power Amplifier
Nonlinearities**

Author:
Peter HUMPHREY

Supervisor:
Dr. Pawel Dmochowski

*A thesis submitted in fulfillment of the requirements
for the degree of Master of Science*

in the

School of Engineering and Computer Science

September 14, 2020

VICTORIA UNIVERSITY OF WELLINGTON

Abstract

Faculty of Engineering
School of Engineering and Computer Science

Master of Science

Optimising the Massive MIMO Downlink in the Presence of Power Amplifier Nonlinearities

by Peter HUMPHREY

Massive MIMO is known for its high level of spectral efficiency in multipath rich environments. We present a detailed Massive MIMO cell system using maximum-ratio transmission (MRT) and zero-forcing (ZF) where energy efficiency is taken into account. This is done through the use of a realistic model of moderate performance and hence moderate cost power amplifiers (PAs) for the base station downlink, which could be applied in a practical Massive MIMO system. In the process of detailing the linear aspects of the Massive MIMO system, results for the normalisation factor and array gain are derived, which as far as the author is aware are original. These results are used to derive an expression to optimise the downlink signal-to-interference-and-noise-ratio (SINR) in a linear system, which is also original as far as the author is aware. A process is outlined to optimise the downlink SINR when nonlinear PAs are used and a simulation of a cell system is performed where the benefits of applying the nonlinear optimisation process are demonstrated.

Acknowledgements

Thank you Pawel for your help and input. Thank you Sarah for your support.

Contents

Abstract	iii
Acknowledgements	v
1 Introduction	1
1.1 Literature and Background	1
1.2 Outline and Contributions	6
2 Information and Background	9
2.1 Model for Large Scale Fading	10
2.1.1 Spatial Distribution of Users	11
2.1.2 Effect of Large Scale Fading	11
2.1.3 Countering Effect of Large Scale Fading	11
2.2 Model for Small Scale Fading	12
2.3 Channel Approximation with Pilot Uplink	12
2.3.1 Using Uplink Pilot to Estimate the Channel	12
2.3.2 Magnitude of Elements of Channel Approximation Matrix and Error Matrix	13
2.4 Precoder Output	14
2.4.1 Precoding Normalisation	14
2.4.2 Precoding Matrix Definition for ZF and MRT	15
2.5 Linear Downlink System Model	15
2.5.1 Linear Signal Received by Users	15
2.5.2 Linear Downlink: Gain, Noise and Interference Terms	16
2.5.3 Linear SINR	17
2.6 Matched Filtering	18
2.7 Received Signal with Nonlinearities	19
2.7.1 Distortion Term Sub-components	19
2.7.2 Nonlinear SINR	20
2.8 Power Amplifiers	21
2.8.1 Power Amplifier Model	21
2.8.2 Signal Backoff	22
2.8.3 Amplifier Efficiency	23
2.9 Conclusion	24
3 Contributions	27
3.1 Normalisation Value	27
3.1.1 Normalisation Value for MRT	28
3.1.2 Normalisation Value for ZF	29
3.2 Array Gain	30
3.2.1 Array Gain for MRT	30
3.2.2 Array Gain for ZF	31
3.3 Array Gain Variance	32

3.3.1	MRT Array Gain Variance	32
3.3.2	ZF Array Gain Variance	32
3.4	Channel Estimation Error Variance	33
3.5	Mean SINR Values for Linear System	33
3.5.1	Mean SINR Value MRT, Linear System	33
3.5.2	Mean SINR Value for ZF, Linear System	34
3.6	Linear Optimisation	34
3.6.1	General Expression	34
3.6.2	Linear Optimisation for MRT	35
3.6.3	Linear Optimisation for ZF	36
3.7	Optimisation Process with Nonlinear Amplifiers	37
3.7.1	Nonlinear Optimisation MRT	40
3.7.2	Nonlinear Optimisation ZF	40
3.8	Conclusion	40
4	Results	43
4.1	Nonlinear Amplifier Derived Constants for MRT and ZF	43
4.2	Estimating Level of PA Input Backoff	44
4.2.1	MRT PA Input Backoff	44
4.2.2	ZF PA Input Backoff	45
4.3	Nonlinear Optimisation Results, ZF	45
4.4	Nonlinear Optimisation Results, MRT	49
4.5	Conclusion	52
5	Conclusion and Future Work	55
A	Normalisation Factors Applied to Generate the Precoding Matrix Frobenius Norm	59
A.1	Effect of Changed Normalisation Factor on Frobenius Norm for MRT	59
A.2	Effect of Changed Normalisation Factor on Frobenius Norm for ZF	60
B	Fixed Pilot Power Optimisation	63
B.1	Optimising SINR with Configurable Pilot Uplink Power, MRT	63
B.2	Optimising SINR with Configurable Pilot Uplink Power, ZF	65
B.3	Conclusion	66
	Bibliography	67

List of Figures

4.1	Average SINR of received signal, MRT	46
4.2	Average SINR of received signal, ZF	46
4.3	CDF of SINR for ZF across multiple large scale fades	47
4.4	CDF of SINR for ZF across multiple large scale fades, smaller axis range	48
4.5	C as a function of ξ , ZF	48
4.6	D as a function of ξ , ZF	49
4.7	ρ as a function of ξ , ZF	49
4.8	CDF of SINR for MRT across multiple large scale fades	50
4.9	C as a function of ξ , MRT	50
4.10	D as a function of ξ , MRT	51
4.11	ρ as a function of ξ , MRT	51
4.12	CDF of SINR for MRT across multiple large scale fade, $\eta = 0.1$	52

List of Tables

4.1	Nonlinear Amplifier Derived Constants	44
4.2	Nonlinear SINR constants, fixed β_k	45
4.3	Nonlinear optimisation constants, multiple large scale fades	45

List of Abbreviations

ACLR	Adjacent Channel Leakage Ratio
AM	Amplitude Modulation
AWGN	Additive White Gaussian Noise
CDF	Cummulative Distribution Function
CSI	Channel State Information
DAC	Digital-to-Analogue Converter
DPC	Dirty Paper Coding
EE	Energy Efficiency
FDD	Frequency Division Duplexing
ISI	InterSymbol Interference
MIMO	Multiple-Input Multiple-Output
MMSE	Minimum Mean Square Error
MRT	Maximum Ratio Transmission
MU-MIMO	Multi-User MIMO
OFDM	Orthogonal Frequency Division Multiplexing
PA	Power Amplifier
PM	Phase Modulation
QPSK	Quadrature Phase Shift Keying
RF	Radio Frequency
RRC	Root-Raised-Cosine
SINR	Signal to Noise and Interference Ratio
SNR	Signal to Noise Ratio
SSPA	Solid State Power Amplifier
SU-MIMO	Single-User MIMO
TDD	Time Division Duplexing
UE	User Equipment
ZF	Zero-Forcing

List of Symbols

M	Number of base station antennas
K	Number of single antenna UEs
\mathbf{H}	Downlink channel matrix $\in \mathbb{C}^{K \times M}$
$\tilde{\mathbf{H}}$	MMSE approximation to channel matrix $\in \mathbb{C}^{K \times M}$
\mathbf{V}	Matrix of orthogonal basis functions $\in \mathbb{C}^{B \times K}$
\mathbf{W}	Precoding matrix with normalisation $\in \mathbb{C}^{M \times K}$
$s_k[n]$	Vector of complex message symbols sent K users at time n
g_k	Linear signal array gain of message signal for user k
$r_k[n]$	Signal received by user k if the approximation to channel \mathbf{H} was actual channel
$u_m[n]$	Output of precoder at transmitter m
\mathbf{B}	Diagonal matrix $\in \mathbb{R}^{K \times K}$ where $\mathbf{B}_{kk} = \beta_k$, the large scale fading power coefficients
$y_k[n]$	Received signal for user k
$z_k[n]$	AWGN at the receiver for user k during the data downlink
B	Number of symbols (QPSK) in uplink pilot
R	Cell radius (all users will be $\leq R$ from base station)
\mathbf{N}	Matrix of AWGN at the base station receivers during the pilot uplink $\in \mathbb{C}^{M \times B}$
r	Cell exclusion zone (all users will be $\geq r$ from base station)
d_0	Cell reference distance, where power scaling = 1
d_k	Distance of user k from base station
\mathbf{F}	Error matrix, that is, the difference between channel matrix \mathbf{H} and approximation $\tilde{\mathbf{H}}$
e_k	Error component of the received signal for user k caused by channel approximation error
E_k	Expected value of the power for e_k
\mathbf{Y}	Matrix of received signal at base station during pilot $\in \mathbb{C}^{M \times B}$
P	Total radiated (RF) power from base station
P_{DC}	Total consumed power at base station
A_{max}	Maximum theoretical output amplitude from power amplifier
u_{ref}	Reference signal input level for input to power amplifier
$\mathbf{u}(t)$	Time domain output of filters, vector $\in \mathbb{C}^M$
$\mathbf{x}(t)$	Time domain output of power amplifier $\in \mathbb{C}^M$
N_s	Number of complex message symbols sent to each user in every frame
N_H	Number of small scale fades generated for each large scale fade
N_β	Number of instances of large scale fading
Ψ	Matrix used in calculation MMSE estimator to channel $\in \mathbb{C}^{B \times B}$
ξ_k	Data downlink power allocation to user k
ζ_k	Pilot uplink power allocation to user k , only applies in Appendix B
δ_k	$\mathbb{E} \tilde{h}_{km} ^2$
α	Normalisation constant for \mathbf{W} matrix
β_k	Power gain to user k caused by large scale fading
η	Average SNR at UE during downlink when $\xi_k \beta_k = 1$
ρ	Mean SNR of the pilot at the base station receiver
σ	RRC filter roll-off
κ	Samples per symbol in RRC filter

τ	RRC filter span in symbols
α_s	Path loss exponent
μ	Efficiency of base station power amplifier

Dedicated to Sarah and Emily...

Chapter 1

Introduction

1.1 Literature and Background

In wireless communications there is an ever increasing demand to obtain better utilisation of available system resources, for example, power and bandwidth. Resource pressure comes from greater demand for user data and the desire to reduce power consumption for cost, battery life and environmental reasons. In wireless communications, a common system is a cell where multiple distinct users have bidirectional communication with a single base station, which accumulates and distributes data to and from the cell users to the wider world. Examples of this system are cellular data networks and Wi-Fi. The increased demands placed on wireless networks show no sign of abating.

One very promising technology to increase spectral efficiency (SE) is MIMO (Multiple Input Multiple Output) at the base station. SE is the data throughput per unit of bandwidth. MIMO involves the use of multiple antennas to exploit spatial multiplexing to increase SE via multiplexing gain and provides robustness to fading due to diversity gain [8]. The use of MIMO as a method of increasing SE is well studied [19, 21]. Spatial multiplexing provides gain by sending multiple streams simultaneously at the same frequency and time by exploiting spatial diversity caused by multipath [28]. Multipath gain is greatest in systems without a strong line of sight. This thesis deals exclusively with spatial multiplexing in cells with significant multipath and without a strong line of sight path. Fading caused by multipath is known as small scale fading. If the separate small scale fading paths between different users and the base station are orthogonal, or near orthogonal, it is known as favourable

propagation for a Massive MIMO channel [2, 17, 19, 29]. Favourable propagation conditions reduce the level of inter-user interference experienced within a cell.

MIMO can be divided into three types, single-user MIMO (SU-MIMO), multiuser MIMO (MU-MIMO) and Massive MIMO [19]. SU-MIMO involves point to point communications between two units, both of which have multiple transmit and receive antennas. MU-MIMO involves a base station with multiple antennas communicating to multiple physically separated units which each have one or more antennas. Massive MIMO is a subset of MIMO where only the base station has knowledge of the channel and only linear signal processing techniques are used. By keeping the requirements of only the base station having channel knowledge and using linear processing, the system is scalable, allowing for the potential of a very large number of base station antennas [19]. This thesis deals exclusively with Massive MIMO where each user has a single antenna.

Massive MIMO is a technology that has only recently been applied to physical commercial systems [5]. The scope of growth for Massive MIMO applications is very large. Massive MIMO can be applied in a range of frequency bands [2]. Massive MIMO has potential benefits in the growing millimeter (30 GHz to 300 GHz) wave band where antenna spacing is reduced by smaller wavelengths [15] allowing denser arrangements of base station antennas. Massive MIMO can also be applied at lower frequencies, for instance, a physical demonstration of MIMO at 2.6 GHz, close to current Wi-Fi and cellular bands, was given in [11]. The authors in [11] also demonstrated high performance linear precoding using zero forcing (ZF), achieving 98% of the capacity of the ideal theoretical dirty paper coding (DPC) throughput in a residential environment, with 20 base station antennas and 2 users. The existence of a favourable propagation environment at 2.6 GHz, which is an environment where different users experience orthogonal small scale fading was demonstrated in [16].

Linear signal processing techniques can provide near optimal performance in cases where the number of base station transceivers is twice or more the number of user receivers [27]. This thesis will deal with two forms of linear signal processing, maximum ratio (MR) and ZF. These two forms of linear processing are the most prominent [33] and whose capacity characteristics have been widely studied [2, 16, 27, 33]. Maximum-ratio processing can be split into maximum-ratio combining for

the uplink and maximum-ratio transmission (MRT) for the downlink. It is also sometimes referred to as matched filtering processing.

The standard Massive MIMO system operates in time-division duplex mode (TDD) [2], due to the simpler channel estimation procedure compared to frequency-division duplex (FDD). Each Massive MIMO TDD frame can be divided into three parts as described in [1]: pilot uplink, data uplink, and data downlink. This frame structure for Massive MIMO has been used in the literature for a long time [18]. Uplink refers to transmission from the users to the base station, while downlink refers to transmission from the base station to the users. It is assumed that the base station antennas are connected to radio frequency (RF) paths that allow for half-duplex operation. That is, if a base station has M antennas, it has M distinct receive paths during uplink and M distinct transmit paths during downlink. The pilot refers to the simultaneous transmission of fixed orthogonal data streams known by both the users and the base station to help the base station learn the effects of small scale fading on the uplink. The effects of small scale fading are assumed to be symmetric for uplink and downlink, so the base station also simultaneously learns the effects of small scale fading on the downlink. Small scale fading effects are assumed to be constant within a TDD frame and uncorrelated between TDD frames. The extent to which the base station has knowledge of the cell's small scale fading and the effect of incomplete knowledge or imperfect channel state information (CSI) forms a significant part of this thesis. This thesis uses the minimum mean square error (MMSE) approximation for small scale fading outlined in [1]. The MMSE estimator in [1] assumes a cell where users experience a Rayleigh fading channel for small scale fading, while users also experience different large scale fading, which is known to the base station.

Large scale fading is due to path loss from the base station to a particular user and is dependent on the user's physical location within the cell, that is, distance from the base station. The combined result of the multipath fading between the users and the base station transceivers describes the communications channel.

This thesis primarily deals with optimisation of downlink signal to noise and interference ratio (SINR), where the aim is to maximise the minimum mean SINR

experienced by any user in the cell. This is subject to the constraint of constant total radiated power with ideal base station hardware and constant consumed power with non-ideal base station hardware. The optimisation involves allocating average downlink powers to users based on the large scale fading they are experiencing. Optimising over large scale fading makes sense, because as explained in [30], optimising over small scale fading is computationally expensive, as the small scale fading environment changes rapidly and it is unnecessary to compensate for variations in capacity due to small scale fading, as users are primarily concerned with average capacity not instantaneous. Another reason given to base optimisation on large scale fading is channel hardening which is explained in [7]. Channel hardening in Massive MIMO channels means that the effects of small scale fading and frequency dependence disappear when the number of base station transceivers is large [19]. A practical physical demonstration of channel hardening at 2.6 GHz is given in [12].

The sole hardware imperfection taken into account in the thesis is PA nonlinearity. While the use of Massive MIMO represents a method of increasing spectral efficiency [19], that is, data throughput per unit of bandwidth, in order to fully characterise a MIMO system, energy efficiency (EE) must be taken into account [9]. According to [8], power amplifiers consume approximately 57% of the energy in macro base stations and cooling contributes 10%, which would also be reduced if power dissipation in the power amplifier is reduced. Power amplifiers (PAs) exhibit a trade off between efficiency and nonlinearity, so the ability to use PAs operating with substantial nonlinearity could greatly reduce overall power consumption at the base station, increasing EE.

The effects of non-ideal PAs on MIMO systems is an active field of research. Some papers model the effects of non-ideal PAs as either a fixed power additive complex Gaussian signal [6, 34] or proportional additive complex Gaussian signal [1, 4], added to the wanted transmit signal at the output of the PAs. Another approach used was to model the input to the PAs as a complex Gaussian signal input to an amplifier described by a nonlinear function [3, 10, 25]. In these cases, by using the Bussgang Theorem, the effect of nonlinearity can be modelled as the combination of a real multiplicative scalar and an additive noise term. These are statistical models for the effect of nonlinearity, which the author found to be unsatisfactory.

In this thesis, we use the approach used in [20], which involved using actual MRT and ZF waveforms run through a nonlinear function. The use of real waveforms is explained in [13], where the limitation of using uncorrelated interference terms is outlined. In reality interference is highly correlated to the transmit waveform. Also as any practical wireless system would use a band-limited passband [14], a Gaussian signal input is not a realistic model. This thesis exclusively uses band-limited waveforms.

There are several papers which optimise capacity over short term fading taking into account amplifier efficiency assuming Rayleigh channels [6, 9, 10, 34]. Also, [34] showed that increasing both the transmit and the receive antennas in a SU-MIMO system means performance does not have a capacity ceiling but a fixed performance offset. While [25] showed that by using multiple receivers for each user, it is possible to compensate for transmit nonlinearities through correcting the received constellation. However, this is not an option for Massive MIMO which relies on single antenna users.

The authors of [4] demonstrated a fixed capacity ceiling for Massive MIMO downlink when hardware impairments are taken into account. While [13] provided a more detailed breakdown of hardware imperfections, including a detailed power amplifier model which included nonlinear memory polynomials for the individual transmitters, mutual coupling between the base station power amplifiers and the impact of digital to analogue (DAC) conversion noise in Ricean and Rayleigh small scale fading channels. The authors of [13] showed a greater resilience to power amplifier coupling in Rayleigh channels. While the authors of [33] provided a detailed comparison of the performance of MRT and ZF for the downlink in a single cell TDD system with Rayleigh fading and imperfect channel knowledge, where channel knowledge is obtained via an MMSE estimate. Note these papers do not take into account large scale fading or imperfect CSI.

1.2 Outline and Contributions

This thesis deals with EE optimisation over large scale fading in systems with imperfect CSI and non-ideal amplifiers. Combining all of those factors together is relatively unusual. Papers which combine all these factors include [1, 20, 31, 35]. Only one of those papers [20], used real waveforms to provide the most realistic model of a Massive MIMO system and it is the paper most heavily referenced in this thesis.

This thesis also focuses on the effects of incomplete channel knowledge and base station PAs nonlinearity on simulated Massive MIMO cells. It empirically determines approximate values of several nonlinear system parameters which are then used to optimise the system capacity subject to maximum-minimum criteria. This is also known as an "egalitarian" criteria [20], where SINR is maximised, subject to the constraint that all users have the same SINR.

Chapter 2 provides a detailed breakdown of the system model used in simulations for both MRT and ZF. This breakdown includes: the definition of cell parameters for small and large scale fading, MRT and ZF signal processing, the MMSE estimator, the nonlinear PA model and the definition of the elements of SINR which can be used to calculate SINR_k , the SINR for user k in both the linear and nonlinear cases. The elements of SINR which are identified for the linear case include: array gain, array gain variance and channel estimation error. While the additional elements of SINR introduced due to saturation in the nonlinear case include: gain loss, interference loss and additional uncorrelated distortion.

Chapter 3 derives explicit values for the elements of SINR for MRT and ZF which were introduced in Chapter 2. This includes what the author believes to be original values for the array gains for MRT and ZF. Having derived explicit values for the elements of SINR, Chapter 3 derives an expression to optimise the maximum-minimum SINR by varying the downlink power allocation in a linear Massive MIMO system. Finally, the chapter introduces a process to optimise the maximum-minimum SINR by varying the downlink power allocation in a nonlinear system, which integrates the explicit value of the elements of SINR for the linear system, as well as the nonlinear factors outlined in Chapter 2. Appendix A applies the normalisation factors to derive values of the Frobenius norms for MRT and ZF. While Appendix B applies

the normalisation factors and linear downlink optimisation expression to a situation where the total pilot uplink power is fixed.

Chapter 4 applies the linear optimisation expression and the nonlinear optimisation process introduced Chapter 3 to a specific set of cells. Each instance of a cell is chosen from a distribution of large scale fading and experiences many instances of small scale fading as introduced in Chapter 2. An optimal level of input backoff for the nonlinear amplifiers is estimated through simulation for MRT and ZF and this level of backoff is applied to the nonlinear process simulations. Estimates of various parameters for the nonlinear amplifier outlined in Chapter 2 are generated through simulation in the process of finding the optimal level of backoff. The nonlinear process relies on calculated estimations of various nonlinear factors introduced in Chapter 3 for MRT and ZF as a function of downlink power. These estimations are plotted and their correlation to allocated downlink power outlined. These estimations are applied to the nonlinear optimisation process to generate cumulative distribution functions (CDFs). These CDFs show the effect of nonlinearity on the minimum average SINR and the benefits of applying the nonlinear optimisation process.

Chapter 5 summarises the thesis and outlines potential areas for future work.

Chapter 2

Information and Background

This chapter provides a detailed description of the Massive MIMO model used in Chap.3 and Chap.4. We begin with the definition of the parameters which make up a cell.

Each cell simulation consists of a single M antenna base station and K single antenna users. The set of simulations take into account large and small scale fading. Initially, an instance of a set of users experiencing large scale fading values is selected from a statistical distribution described in Sec.2.1.1. For this single instance of large scale fading, multiple instances of small scale fading are selected from a different statistical distribution described in Sec.2.2. For each instance of small scale fading, a TDD frame simulation is run. It is assumed that each TDD frame would experience independent small scale fading and that the base station would need to learn the unique channel for each frame. The channel is assumed to stay constant for the duration of each frame, that is, the frame duration does not exceed the channel coherence time. It is assumed the base station has perfect knowledge of large scale fading, while it only has complete knowledge of the statistical distribution of small scale fading. The assumption of independent channels for each frame is based on the observation that small scale fading occurring quickly, while generating multiple instances of small scale fading for each large scale fade is based on large scale fading occurring slowly. The TDD frame is divided into pilot uplink and data downlink. The channel is assumed to be frequency flat for all channel instantiations, which is a valid assumption for orthogonal frequency-division multiplexing (OFDM) based transmissions.

This chapter then introduces the MMSE channel estimation matrix in Sec.2.3.1

and defines some associated parameters in Sec.2.3.2. The base station can only derive an approximation to the small scale fading based on the signal it receives during the pilot uplink combined with its knowledge of the statistical distribution of small scale fading. During the pilot uplink, each of the users simultaneously transmits a sequence of symbols which are orthogonal to other users. The uplink pilots are scaled by the large scale fading gain. The base station makes an MMSE estimation to the channel, whose accuracy is limited by the presence of additive white Gaussian noise (AWGN) at its receivers. The base station uses its approximation to the channel to generate a precoding matrix for the data downlink. This thesis uses ZF and MRT precoding. Simulations only deal with a single cell, that is, we neglect inter-cell interference.

The users receive data in the presence of multiple sources of noise and interference. The ability to isolate and separate the sources of noise and interference is essential to optimising the downlink power distribution to maximise minimum SINR as outlined in Chap.3. The sources can be divided into linear given in Sec.2.5 and nonlinear given in Sec.2.7.1. The linear sources are independent of nonlinearities in the base station PAs, while the nonlinear sources are caused by the PAs' nonlinearities. The linear sources of noise and interference are gain variance, channel approximation error and AWGN at the receivers. In the nonlinear case, the data downlink experiences all of the linear effects and also array gain loss due to saturation of transmit signal, interference loss due to saturation and additional uncorrelated distortion.

For the linear case, the aim is to maximise the received average SINR for a given total transmit power at the base station. The definition of SINR is given in Sec.2.5.3, and it is separated into its linear components. The additional components of SINR in the nonlinear case are given in Sec.2.7.1. For the nonlinear case, the aim is to maximise the received average SINR for a power consumed at the base station PAs. A detailed breakdown of the model used for the PAs is given in Sec.2.8.1.

2.1 Model for Large Scale Fading

Every instance of large scale fading involves a set of K independent large scale fading values described in Sec.2.1.2. For each instance of large scale fading, many instances

of small scale fading are generated from a model explained in Sec.2.2.

2.1.1 Spatial Distribution of Users

The large scale channel effect used in this thesis depends on the spatial distribution of users. User k is distance d_k from the base station. The dimensions of the cell are an annulus shape where $r < d_k < R$. The distribution of users within the cell is uniform by area. Using the inverse sampling method, d_k is computed by

$$d_k = \sqrt{u_k(R^2 - r^2) + r^2}, \quad (2.1)$$

where $u_k \sim \mathcal{U}(0, 1)$ i.e., is taken from a uniform distribution between 0 and 1.

2.1.2 Effect of Large Scale Fading

The distribution of users given in (2.1) is then combined with a path loss exponent to give a total power scaling factor for user k ,

$$\beta_k = d_k^{\alpha_s}, \quad (2.2)$$

where α_s is the path loss exponent, which can vary depending on the type of environment (e.g., rural or urban) in which the cell is located. In this thesis, the β_k values are often grouped together in a matrix $\mathbf{B} \in \mathbb{R}^{K \times K}$ where

$$\mathbf{B} = \begin{bmatrix} \beta_1 & 0 & \dots & 0 \\ 0 & \beta_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \beta_K \end{bmatrix}. \quad (2.3)$$

2.1.3 Countering Effect of Large Scale Fading

To compensate for the effect of large scale fading, this thesis uses ζ_k as a parameter to be optimised. ζ_k is the proportion of the downlink power that the base station allocates to user k during downlink data transmissions.

In this thesis, ζ_k is normalised such that

$$\sum_{i=1}^K \zeta_k \equiv 1, \quad 0 < \zeta_k \leq 1 \quad \forall k, \quad (2.4)$$

that is, there is a fixed total downlink power to be distributed between the users.

2.2 Model for Small Scale Fading

Small scale fading is taken into account in the downlink channel matrix $\mathbf{H} \in \mathbb{C}^{K \times M}$. The distribution of channel matrices is Rayleigh where each element is i.i.d. and $h_{km} \sim \mathcal{CN}(0, 1)$. Assuming TDD channel reciprocity, the uplink and downlink for each channel instance are symmetrical, that is, the uplink channel matrix is \mathbf{H}^* .

In this thesis

$$\mathbb{E}[|h_{km}|^2] = 1 \quad \forall k, m, \quad (2.5)$$

which is a standard normalisation [20], causing small scale fading to have no effect on the mean signal power.

2.3 Channel Approximation with Pilot Uplink

The system uses TDD, where each cell consists of a fixed number K of single antenna user equipments (UEs) and one base station with M antennas. Each frame consists of a pilot uplink and a data downlink. Note that this thesis does not consider a possible data uplink. This section outlines the pilot uplink and the resulting channel approximation matrix.

2.3.1 Using Uplink Pilot to Estimate the Channel

During the pilot uplink, the UEs simultaneously send a sequence of B orthogonal quadrature phase shift keying (QPSK) signals where $B \geq K$. The received signal at the base station is given by

$$\mathbf{Y} = \mathbf{H}^* \sqrt{\mathbf{B}} \mathbf{V}^* + \mathbf{N}, \quad (2.6)$$

where $\mathbf{Y} \in \mathbb{C}^{M \times B}$ is the matrix of the received signal at the base station at antenna m due to symbol b , \mathbf{H}^* is the uplink channel matrix $\in \mathbb{R}^{M \times K}$, \mathbf{V} is the matrix of pilot signals $\in \mathbb{C}^{B \times K}$ where each of the K users broadcast an orthogonal sequence of B QPSK symbols, \mathbf{N} is a matrix of AWGN $\in \mathbb{C}^{M \times B}$ where $\mathbf{N}_{mb} \sim \mathcal{CN}(0, \rho)$, ρ is the power ratio inverse of signal to noise ratio (SNR) during the pilot from user k at the base station normalised for $\beta_k = 1$ and $\mathbf{B} \in \mathbb{R}^{K \times K}$ is the large scale fading matrix defined in (2.3).

The base station generates an MMSE approximation $\tilde{\mathbf{H}}$ to the channel \mathbf{H} . The MMSE estimator for a Rayleigh channel under AWGN is given in [1] for a multicell case. This thesis only deals with a single cell case which reduces the estimate to

$$\tilde{\mathbf{H}} = (\mathbf{Y}\Psi^{-1}\mathbf{V})^*, \quad (2.7)$$

where

$$\Psi = \mathbf{V}\mathbf{B}\mathbf{V}^* + \frac{1}{\rho}\mathbf{I}_B. \quad (2.8)$$

\mathbf{I}_B is the $B \times B$ identity matrix. The uplink approximation is done the same way and using the same equations for both MRT and ZF.

2.3.2 Magnitude of Elements of Channel Approximation Matrix and Error Matrix

The channel matrix \mathbf{H} can be defined as the sum of the MMSE estimation matrix $\tilde{\mathbf{H}}$ and an error matrix \mathbf{F} ,

$$\mathbf{H} = \tilde{\mathbf{H}} + \mathbf{F}, \quad (2.9)$$

where $\mathbf{F} \in \mathbb{C}^{K \times M}$ is the error matrix between the approximation to the channel matrix $\tilde{\mathbf{H}} \in \mathbb{C}^{K \times M}$ and \mathbf{H} . Note that all the elements of \mathbf{F} are uncorrelated to all elements in \mathbf{H} . Introducing δ_k for the variance of the elements of the estimator, that is

$$\delta_k \equiv \mathbb{E}[|\tilde{h}_{km}|^2]. \quad (2.10)$$

For an MMSE estimator for a Rayleigh fading channel [20] gives the variance of the elements of the estimator as

$$\delta_k = \frac{B\rho\beta_k}{1 + B\rho\beta_k}, \quad (2.11)$$

where B is the pilot length, β_k is the power scaling factor for user k and ρ is the mean SNR during the pilot uplink at the base station receiver when $\beta_k = 1$. Note that [20] also considers for the number of taps in the channel, L , allowing for the possibility the channel is not frequency flat. This thesis only deals with the case $L = 1$ where the channel is frequency flat.

Similarly the power in the elements of the error matrix \mathbf{F} defined in (2.9) is

$$\mathbb{E}[|f_{km}|^2] = \frac{1}{1 + B\rho\beta_k}. \quad (2.12)$$

Note that as the power of the elements of \mathbf{H} are defined as in (2.5) and the elements of \mathbf{H} and \mathbf{F} are uncorrelated

$$\mathbb{E}[|f_{km}|^2] = 1 - \delta_k, \quad \forall k, m. \quad (2.13)$$

2.4 Precoder Output

This section outlines the precoding matrices used for MRT and ZF, \mathbf{W}_{ZF} and \mathbf{W}_{MRT} , respectively and defines the power value at the output of the precoder. When it is not necessary to specify a precoder type, i.e., the statement or equation applies to both, \mathbf{W} will be used.

2.4.1 Precoding Normalisation

We denote the output of the precoder at discrete time n by [20]

$$\mathbf{u}[n] = \mathbf{W}\mathbf{s}[n], \quad (2.14)$$

where \mathbf{W} is the precoding matrix and $\mathbf{s}[n]$ is the message vector at time n . We normalise the signal so that

$$\sum_{m=1}^M \mathbb{E}[|\mathbf{u}_m[n]|^2] = 1 \quad \forall m, n, \quad (2.15)$$

or alternatively

$$\mathbb{E}[|\mathbf{u}_m[n]|^2] = \frac{1}{M} \quad \forall m, n. \quad (2.16)$$

2.4.2 Precoding Matrix Definition for ZF and MRT

The precoding matrix \mathbf{W}_{MRT} for MRT is defined as [20]

$$\mathbf{W}_{\text{MRT}} \equiv \alpha_{\text{MRT}} \tilde{\mathbf{H}}^*. \quad (2.17)$$

The precoding \mathbf{W}_{ZF} matrix for ZF is defined as

$$\mathbf{W}_{\text{ZF}} \equiv \alpha_{\text{ZF}} \tilde{\mathbf{H}}^* (\tilde{\mathbf{H}} \tilde{\mathbf{H}}^*)^{-1}, \quad (2.18)$$

where an expression for the value of α_{ZF} required to satisfy (2.16) will be derived in Sec.3.1.1 while the value of α_{MRT} required to satisfy (2.16) is derived in Sec.3.1.2.

2.5 Linear Downlink System Model

This section defines the signal received by the users during the data downlink. It then divides this signal into various gain, noise and interference terms to allow for optimisation of downlink SINR described in Chap.3.

2.5.1 Linear Signal Received by Users

The signal received by the UEs in a linear model can be described as follows, which is taken from [20]

$$\mathbf{y}[n] = (\sqrt{P} \sqrt{\mathbf{B}} \mathbf{H})(\mathbf{W} \mathbf{s}[n]) + \mathbf{z}[n] \quad (2.19)$$

where P is the total radiated (RF) power, \mathbf{B} is a diagonal matrix $\in \mathbb{R}^{K \times K}$ defined in (2.3), \mathbf{H} is the downlink channel matrix, \mathbf{W} is the normalised precoding matrix $\in \mathbb{C}^{M \times K}$, $\mathbf{s}[n] \in \mathbb{C}^K$ is the vector of QPSK message symbols sent to user k at time n

and $\mathbf{z}[n]$ is the vector of AWGN at the UEs, where $z_k[n] \sim \mathcal{CN}(0, Z) \forall n$. There are two types of precoding matrices used in this thesis, MRT and ZF outlined in Sec.2.4.2.

The downlink power allocated to user k is ζ_k defined by

$$\zeta_k \equiv \mathbb{E}[|s_k[n]|^2] \quad \forall n, k, \quad (2.20)$$

where ζ_k was introduced in Sec.2.1.3. The values of ζ_k optimised for maximum-minimum SINR in Chap.3 and Chap.4.

2.5.2 Linear Downlink: Gain, Noise and Interference Terms

In order to perform linear SINR optimisation in Chap.3 and Chap.4 it is necessary to identify and separate the sources of signal gain, interference and noise that the users receive.

One potential source of interference is due to channel estimation error. By treating the channel as the sum of an estimation and error matrix as in (2.9) and substituting into (2.19) gives

$$\mathbf{y}[n] = \sqrt{P}\sqrt{\mathbf{B}}(\tilde{\mathbf{H}}\mathbf{W}\mathbf{s}[n] + \mathbf{F}\mathbf{W}\mathbf{s}[n]) + \mathbf{z}[n]. \quad (2.21)$$

Defining the received signal in the absence of approximation error as $\mathbf{r}[n] \in \mathbb{C}^K$ where

$$\mathbf{r}[n] \equiv \tilde{\mathbf{H}}\mathbf{W}\mathbf{s}[n] \quad (2.22)$$

and the received signal error from channel approximation error $\mathbf{e}[n] \in \mathbb{C}^K$ where

$$\mathbf{e}[n] \equiv \mathbf{F}\mathbf{W}\mathbf{s}[n], \quad (2.23)$$

then substituting (2.22) and (2.23) into (2.21) gives

$$\mathbf{y}[n] = \sqrt{\mathbf{B}}(\mathbf{r}[n] + \mathbf{e}[n]) + \mathbf{z}[n]. \quad (2.24)$$

Once approximation error is removed to create $r_k[n]$, an important deterministic constant can be derived from r_k , the array gain g_k which is defined as [20]

$$g_k \equiv \frac{\mathbb{E}[s_k^*[n]r_k[n]]}{\sqrt{\delta_k}\zeta_k}. \quad (2.25)$$

The array gain term for user k , g_k is calculated using the correlation between the signal transmitted to user k , $s_k[n]$ and $r_k[n]$. The gain is normalised by $\sqrt{\delta_k}$ to remove the effect of multiplication by $\tilde{\mathbf{H}}$ and ζ_k to remove the effect of power allocation at the base station. Note that in Chap.3, original values for g_k for MRT and ZF are derived.

Using the deterministic array gain term g_k , the $r_k[n]$ can be divided further into array gain acting on the message symbol $s_k[n]$ which is then scaled by the message channel gain for the approximation $\sqrt{\delta_k}$ and $i_k[n]$ which is the rest of the received signal. Being uncorrelated to the message symbols, $i_k[n]$ is a interference source.

From [20], the interference term i_k is

$$i_k[n] \equiv r_k[n] - \sqrt{\delta_k}g_k s_k[n]. \quad (2.26)$$

Rearranging (2.26) for r_k and substituting into (2.24) gives

$$y_k[n] = \sqrt{\beta_k}(\sqrt{\delta_k}g_k s_k[n] + i_k[n] + e_k[n]) + z_k. \quad (2.27)$$

For later use, the variance of the interference for user k is I_k ,

$$I_k \equiv \mathbb{E}[|i_k|^2]. \quad (2.28)$$

In Sec.3.3.1, the value of I_k is derived for MRT, while in Sec.3.3.2 it is derived for ZF, while the variance of the channel error for user k is E_k , that is,

$$E_k \equiv \mathbb{E}[|e_k|^2], \quad (2.29)$$

where e_k is defined in (2.23).

2.5.3 Linear SINR

In [20] the SINR for user k is defined by

$$\text{SINR}_k = \frac{|\mathbb{E}[y_k^*[n]s_k[n]]|^2/\zeta_k}{\mathbb{E}[|y_k[n]|^2] - |\mathbb{E}[y_k^*[n]s_k[n]]|^2/\zeta_k} \quad (2.30)$$

Note that this definition applies to both linear and nonlinear systems. By substituting the definition of y_k given in (2.27), and using the fact that s_k , i_k and e_k are uncorrelated and using the definition of ζ_k given in (2.20), substituting into (2.30) gives [20]

$$\text{SINR}_k = \frac{P\beta_k\delta_k\zeta_k g_k^2}{P\beta_k(I_k + E_k) + Z}. \quad (2.31)$$

Introducing mean downlink SNR value at UE when $\beta = 1$

$$\eta = \frac{P}{Z} \quad (2.32)$$

Then (2.31) can be written

$$\text{SINR}_k = \frac{\eta\beta_k\delta_k\zeta_k g_k^2}{\eta\beta_k(I_k + E_k) + 1} \quad (2.33)$$

In Chap.3, values for g_k , I_k and E_k for MRT and ZF are derived. Also in Chap.3, once the values of g_k , I_k and E_k are known, they are used to optimise (2.33) by varying the downlink power allocation ζ_k .

2.6 Matched Filtering

The output signal from the precoder $\mathbf{u}[n]$ given in (2.14) is then upsampled and passed through a root-raised-cosine (RRC) filter. The purpose of the filter is to produce a band-limited analogue signal (or in simulation, upsampled digital domain signal) to pass to the PAs. This is the same as in [20]. All M transmitters at the base station have their own individual RRC filter, which all have identical coefficients. The UEs have matched receive filters. If the PA is perfectly linear and in the absence of channel frequency selectivity, the output of the receiver matched filter will have resulting zero intersymbol interference (ISI) [14]. In this thesis, the use of nonlinearity in the PAs causes ISI at the receivers.

The matched filter takes the M output signals from the precoder $u_m[n]$ in the discrete domain to M continuous domain functions $u_m(t)$ via the filtering operation [20]

$$u_m \equiv \sum_n u_m[n]p(t - nT). \quad (2.34)$$

2.7 Received Signal with Nonlinearities

Having detailed the linear elements of the downlink system in Sec.2.5 the remainder of Chap.2 deals with the nonlinear system elements. The in-band distortion caused to user k due to base station PA nonlinearity is the difference between the linear signal and the received signal after matched filtering at the receiver followed by downsampling. As the output of the nonlinear amplifier no longer matched to the receive filter, it has to be represented in the time domain (and simulated in the upsampled digital domain) and can be written

$$d_k(t) = \mathbf{H}\mathbf{x}(t) - \mathbf{H}(\sqrt{P}\mathbf{u}(t)). \quad (2.35)$$

Note that $\sqrt{P}\mathbf{u}(t)$ is the equivalent output power from a perfectly linear amplifier.

Once $d_k(t)$ is filtered at the UE receiver and sampled it gives

$$d_k[n] = d_k(\tau) * p^*(-\tau). \quad (2.36)$$

Note as $p(\tau)$ is real and symmetric this reduces to

$$d_k[n] = d_k(\tau) * p(\tau). \quad (2.37)$$

When the distortion term is added, the linear system given in (2.19) becomes

$$\mathbf{y}[n] = (\sqrt{P}\sqrt{\mathbf{B}})(\mathbf{H}\mathbf{W}\mathbf{s}[n] + \mathbf{d}[n]) + \mathbf{z}[n]. \quad (2.38)$$

2.7.1 Distortion Term Sub-components

Once the distortion term for user k , d_k , has been found, it is possible to separate it into sub-components as is done in [20].

We denote the power of the distortion term which is correlated to the sent signal to user k by c_k . It is normally negative as it represents the reduction in gain of the wanted signal due to saturation. It is normalised by $\sqrt{\delta_k \bar{\zeta}_k}$ to remove the effect of the channel and the power allocation at the base station [20]

$$c_k[n] = \frac{\mathbb{E}[s_k^*[n]d_k[n]]}{\sqrt{\delta_k \bar{\zeta}_k}}. \quad (2.39)$$

The term ρ_k , represents the reduction in gain variance and channel estimation error due to saturation [20]. It is normalised by the power of the gain variance I_k and the channel estimation error power E_k . This term is also generally negative.

$$\rho_k[n] = \frac{\mathbb{E}[(i_k^* + e_k^*)d_k[n]]}{I_k + E_k}. \quad (2.40)$$

Finally the term d'_k is used for the additional distortion which is uncorrelated to s_k , i_k and e_k [20]

$$d'_k[n] = c_k \sqrt{\delta_k} s_k[n] + \rho_k (i_k[n] + e_k[n]) - d_k[n]. \quad (2.41)$$

The power of the uncorrelated distortion term d'_k is denoted as

$$D_k = \mathbb{E}[|d'_k[n]|^2]. \quad (2.42)$$

Once the array gain saturation term c_k , the interference and channel error saturation term ρ_k and the uncorrelated distortion term d_k are known, the received signal for user k can be written [20]

$$y_k[n] = \sqrt{P} \beta_k (\sqrt{\delta_k} (g_k + c_k) s_k[n] + (1 + \rho_k) (i_k[n] + e_k[n]) + d'_k[n]) + z_k[n]. \quad (2.43)$$

Having separated the distortion elements, it is possible to write a general term for the downlink SINR for user k in the presence of PA nonlinearity.

2.7.2 Nonlinear SINR

The term for received signal, $y_k[n]$, in the presence of PA nonlinearity which includes g_k , ρ_k and D_k is given by (2.43). By using the general equation for SINR_k , (2.30), substituting (2.43) for y_k gives [20]

$$\text{SINR}_k = \frac{\delta_k \xi_k \eta \beta_k |g_k + c_k|^2}{\eta \beta_k ((I_k + E_k) |1 + \rho_k|^2 + D_k) + 1} \quad (2.44)$$

which is the general term for SINR_k in a system with nonlinearities. The values of g_k , I_k and E_k as a function of system parameters are derived in Chap.3. The values of the distortion related terms c_k , ρ_k and D_k are empirically calculated from cell simulations in Chap.4 using the PA model explained in Sec.2.8. In Chap.3 a process for maximising the minimum SINR using (2.44) with empirically calculated c_k , ρ_k and D_k is outlined.

2.8 Power Amplifiers

This section details the specific PA model used in this thesis. After passing through the matched filter, the base station transmit signal is passed through a nonlinear PA. We introduce the elements of the system dealing with the non-ideal PAs: the power amplifier model, the signal backoff and the amplifier efficiency.

2.8.1 Power Amplifier Model

All M transmitters have identical independent PAs which introduce nonlinear distortion. Given the large number of transmitters involved in Massive MIMO, it is desirable to use low cost, low power consumption PAs. This thesis considers class B amplifiers. As outlined in [22], class B amplifiers are more efficient than class A but produce more distortion. This thesis will use the Rapp model, which is a common model for solid state power amplifiers (SSPAs). It is a type of memoryless model [23]. The Rapp model for SSPAs is widely used as an approximation to the behaviour of low to medium linearity PAs in MIMO systems, for instance [10, 20, 25, 26]. The Rapp model assumes only AM-AM distortion with no AM-PM distortion. AM-AM distortion is a nonlinear effect on output amplitude due to a change in input amplitude. While AM-PM refers to a phase distortion or shift resulting from a change in input amplitude. In a physical system, the PAs are an RF device, however, using a complex base band equivalent model as given in (2.45) provides an accurate representation.

The complex scalar output of the PA, $g(u)$, for a complex scalar input u using the Rapp model is [22]

$$g(u) = A_{\max} \frac{\frac{u}{u_{\text{ref}}}}{\left(1 + \left(\frac{|u|}{u_{\text{ref}}}\right)^{2p}\right)^{\frac{1}{2p}}}, \quad (2.45)$$

where A_{\max} is the maximum output amplitude, u_{ref} is a reference input magnitude and p is positive parameter which control smoothness [22]. As in [20], we set

$$u_{\text{ref}} = \sqrt{M}. \quad (2.46)$$

For small signals

$$\lim_{|u| \rightarrow 0} g(u) = u \left(\frac{A_{\max}}{u_{\text{ref}}} \right), \quad (2.47)$$

which makes the small signal gain

$$g'(0) = \frac{A_{\max}}{u_{\text{ref}}}. \quad (2.48)$$

Note the the nonlinear PA reduces to a linear PA in the limit of a small signal.

2.8.2 Signal Backoff

The input signal to the PA is scaled by a real constant \sqrt{b} to change the level of saturation the signal experiences. The backoff value is measured from the 1dB input compression point, that is the input level at which the output is 1dB below where it would be for a perfectly linear amplifier. The 1dB input compression point for the Rapp model derived below.

Note the for a perfectly linear amplifier, the output is the small signal gain times the input, i.e., $g'(0)u$. The ratio of the actual PA output to the perfectly linear output is set as equal to -1dB and solved for $|u_{1\text{dB}}|$

$$-1 = 20 \log_{10} \left(\frac{g(u_{1\text{dB}})}{g'(0)u_{1\text{dB}}} \right) \quad (2.49)$$

which using the definition of $g(u)$ given in (2.45) and $g'(0)$ given in (2.48), rearranged for $|u_{1\text{dB}}|$ gives

$$|u_{1\text{dB}}| = u_{\text{ref}} \left(10^{\frac{-2p}{20}} - 1 \right)^{\frac{1}{2p}} \quad (2.50)$$

As in [20], we consider $p = 2$, for a model of a moderate cost PA.

Substituting $p = 2$ and (2.46) into (2.50) gives

$$|u_{1\text{dB}}| \approx 0.8745 \sqrt{M}, \quad (2.51)$$

which is consistent with [22]. Using the fact that the power of each of the precoder matrix for all M transmitters is defined in (2.16), the input power scalar b can then be defined as

$$b \approx \frac{10^{\text{IBO}_{\text{dB}}/10}}{0.8745^2}, \quad (2.52)$$

where IBO_{dB} is the relative backoff to the 1dB input saturation point.

The input to the power amplifier M is given by

$$\frac{u_m(t)}{\sqrt{b}}, \quad (2.53)$$

resulting in the output of power amplifier

$$x_m(t) = g \left(\frac{u_m(t)}{\sqrt{b}} \right) = \frac{A_{\text{max}} \left(\frac{u_m(t)}{\sqrt{Mb}} \right)}{\left(1 + \left(\frac{|u_m(t)|}{\sqrt{Mb}} \right)^4 \right)^{\frac{1}{4}}}. \quad (2.54)$$

A_{max} is a scaling factor for $x_m(t)$ which is derived via simulation to give the required total radiated power P for various IBO_{dB} in Sec.4.1. b is also calculated via simulation for various IBO_{dB} in section Sec.4.1.

Finally, the total radiated power from the base station is defined as [20]

$$P = \lim_{t_0 \rightarrow \infty} \sum_{m=1}^M \mathbb{E} \left[\frac{1}{t_0} \int_{\frac{t_0}{2}}^{\frac{t_0}{2}} |x_m(t)|^2 dt \right]. \quad (2.55)$$

2.8.3 Amplifier Efficiency

The efficiency of an amplifier is the ratio of consumed power to transmitted RF power. The efficiency for a class B amplifier whose output is defined by $g(u)$ is given

by [22]

$$\mu = \frac{\pi \mathbb{E}[g(|u(t)|)^2]}{4A_{\max} \mathbb{E}[g(|u(t)|)]}. \quad (2.56)$$

In Sec.4.2.1, the resulting efficiency of using differing levels of backoff constant b are derived through simulation. Note these simulations involve substituting (2.54) into (2.56), which results in A_{\max}^2 terms on both the numerator and the denominator, which cancel, leaving μ independent of A_{\max} .

In Chap.3 and Chap.4, it is assumed that the consumed power is constant. If the consumed power P_{DC} is constant, then the radiated power

$$P = \mu P_{\text{DC}}, \quad (2.57)$$

where μ is the efficiency of the nonlinear amplifier as a ratio. Without loss of generality it is assumed that $P_{\text{DC}} = 1$, giving

$$P = \mu, \quad (2.58)$$

which can be substituted into (2.38) to give a general equation for the received signal in a nonlinear system of

$$\mathbf{y}[n] = (\sqrt{\mu} \sqrt{\mathbf{B}})(\mathbf{H}\mathbf{W}\mathbf{s}[n]) + \mathbf{d}[n] + \mathbf{z}[n]. \quad (2.59)$$

(2.59) is used in Chap.4 to calculate SINR_k and the nonlinear components of SINR, c_k , ρ_k and D_k .

2.9 Conclusion

In this section a detailed breakdown of components of the Massive MIMO system was provided. This included small and large scale fading, the channel approximation matrix, the PAs, the definition of downlink SINR for user k and an expansion of the linear and the nonlinear contributions to that SINR. Having separated the contributions to SINR, it is then possible to optimise the maximum-minimum SINR by controlling the downlink power allocation ξ_k . Chap.3 will provide explicit values for

the linear elements of SINR, including a original result for array gain g_k . Chap.3 will also provide solutions to optimising the maximum-minimum SINR in both linear and nonlinear systems. For a linear system, a closed form solution is found, using the explicit values found for the linear system elements. For a nonlinear system, a solution is found which requires empirical values based on simulations performed in Chap.4.

Chapter 3

Contributions

This chapter will expand on the Massive MIMO system outlined in Chap.2. The aim of this chapter is to introduce what the author believes to be original results for maximum-minimum SINR optimisation in a linear and nonlinear systems.

Having previously introduced the linear system components: array gain g_k , array gain variance I_k , and channel approximation error variance E_k , this chapter will provide explicit values for a given cell in both the ZF and MRT cases. The result for array gains are original as far as the author is aware. As an intermediate step to finding g_k , I_k and E_k , Sec.3.1 derives normalisation factors for ZF and MRT, α_{ZF} and α_{MRT} respectively, which when applied, give constant mean power output from the precoders, independent of channel estimation error and downlink power allocation. As far as the author is aware these normalisation factors are also original. Once the normalisation factors are found, they are applied to generate the array gains for MRT and ZF in Sec.3.2. In Sec.3.3 the gain variance is derived for ZF and MRT and in Sec.3.4 the channel estimation error is found which is the same for ZF and MRT. Having found explicit values of g_k , I_k and E_k , a closed form solution to optimise maximum-minimum SINR for MRT and ZF is found in Sec.3.6. As far as the author is aware this is also original, given the original values of g_k . This chapter will then apply the explicit values of g_k , I_k and E_k to outline a process to optimise the maximum-minimum SINR in a system with nonlinear PAs in Sec.3.7.

3.1 Normalisation Value

This section derives the normalisation values for MRT and ZF which provide constant mean power at the precoder outputs, independent of the channel estimator

matrix variance δ_k and downlink power distribution ζ_k . As far as the author is aware, these values are original. A derivation of the effect of these normalisation values on the Frobenius norm of the precoding matrices for MRT and ZF are given in Appendix A.

3.1.1 Normalisation Value for MRT

To find the value of the normalisation factor for MRT, α_{MRT} , introduced in the definition of MRT (2.17), it is necessary to find the value of α_{MRT} which satisfies the required total precoder output power defined in (2.15). Substituting the definition of precoder output (2.14) into (2.15) gives

$$\sum_{m=1}^M \mathbb{E}[|\mathbf{u}_m[n]|^2] = \mathbb{E}[\mathbf{u}^*[n]\mathbf{u}[n]] = \mathbb{E}[\mathbf{s}^*[n]\mathbf{W}_{\text{MRT}}^* \mathbf{W}_{\text{MRT}} \mathbf{s}[n]] = 1. \quad (3.1)$$

Using the definition of \mathbf{W}_{MRT} given in (2.17)

$$\mathbf{W}_{\text{MRT}}^* \mathbf{W}_{\text{MRT}} = \alpha_{\text{MRT}}^2 (\tilde{\mathbf{H}} \tilde{\mathbf{H}}^*). \quad (3.2)$$

Then substituting into (3.1)

$$\mathbb{E}[\mathbf{s}^*[n]\mathbf{W}_{\text{MRT}}^* \mathbf{W}_{\text{MRT}} \mathbf{s}[n]] = \mathbb{E}[\mathbf{s}^*[n]\alpha_{\text{MRT}}^2 (\tilde{\mathbf{H}} \tilde{\mathbf{H}}^*) \mathbf{s}[n]] = 1, \quad (3.3)$$

which can be rearranged to for α_{MRT} to give

$$\alpha_{\text{MRT}} = \frac{1}{\sqrt{\mathbb{E}[\mathbf{s}^*[n]\tilde{\mathbf{H}} \tilde{\mathbf{H}}^* \mathbf{s}[n]]}}. \quad (3.4)$$

Using the fact the \tilde{h}_{km} are uncorrelated to s_k , it is possible to separate the product of the expectations. Dealing with $\mathbb{E}[(\tilde{\mathbf{H}} \tilde{\mathbf{H}}^*)]$ first, as the \tilde{h}_{km} are from an i.i.d. Rayleigh fading channel, they are uncorrelated with each other, and the expectation of the product of \tilde{h}_{km} and \tilde{h}_{mk}^* are only non-zero when $m = k$. Therefore $\mathbb{E}[(\tilde{\mathbf{H}} \tilde{\mathbf{H}}^*)]$ reduces to

$$\mathbb{E}[(\tilde{\mathbf{H}} \tilde{\mathbf{H}}^*)] = \mathbf{I}_k \mathbb{E} \left[\sum_{m=1}^M \tilde{h}_{kk} \tilde{h}_{kk}^* \right] = \mathbf{I}_k M \mathbb{E}[\tilde{h}_{kk} \tilde{h}_{kk}^*]. \quad (3.5)$$

Using the definition of δ_k from (2.10) then,

$$\mathbb{E}[(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^*)] = M \begin{bmatrix} \delta_1 & 0 & \dots & 0 \\ 0 & \delta_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \delta_K \end{bmatrix}. \quad (3.6)$$

To find $\mathbb{E}[\mathbf{s}^*[n]\tilde{\mathbf{H}}\tilde{\mathbf{H}}^*\mathbf{s}[n]]$ which is required for (3.4), using the definition of ξ_k from (2.20) and (3.6) and the fact \tilde{h}_{km} and s_k are uncorrelated gives,

$$\mathbb{E}[\mathbf{s}^*[n]\tilde{\mathbf{H}}\tilde{\mathbf{H}}^*\mathbf{s}[n]] = M \sum_{k=1}^K \xi_k \delta_k, \quad (3.7)$$

which when substituted in (3.4) gives

$$\alpha_{\text{MRT}} = \sqrt{\frac{1}{M \sum_{k=1}^K \xi_k \delta_k}}. \quad (3.8)$$

3.1.2 Normalisation Value for ZF

Similarly a value of the normalisation factor for ZF, α_{ZF} , as introduced in the definition of ZF (2.18), is derived in this section. Again, beginning with substituting (2.14) into (2.15) one can write

$$\sum_{m=1}^M \mathbb{E}[|\mathbf{u}_m[n]|^2] = \mathbb{E}[\mathbf{u}^*[n]\mathbf{u}[n]] = \mathbb{E}[\mathbf{s}[n]^* \mathbf{W}_{\text{ZF}}^* \mathbf{W}_{\text{ZF}} \mathbf{s}[n]] = 1. \quad (3.9)$$

Using the definition of \mathbf{W}_{ZF} given in (2.18) it can be shown that

$$\mathbf{W}_{\text{ZF}}^* \mathbf{W}_{\text{ZF}} = \alpha_{\text{ZF}}^2 (\tilde{\mathbf{H}}^* (\tilde{\mathbf{H}}\tilde{\mathbf{H}}^*)^{-1})^* (\tilde{\mathbf{H}}^* (\tilde{\mathbf{H}}\tilde{\mathbf{H}}^*)^{-1}) \quad (3.10)$$

By using the fact that for any matrices \mathbf{A} and \mathbf{C} [32]

$$(\mathbf{AC})^* = \mathbf{C}^* \mathbf{A}^*, \quad (3.11)$$

(3.10) can be rearranged to give

$$\mathbf{W}_{\text{ZF}}^* \mathbf{W}_{\text{ZF}} = \alpha_{\text{ZF}}^2 ((\tilde{\mathbf{H}}\tilde{\mathbf{H}}^*)^{-1})^* (\tilde{\mathbf{H}}\tilde{\mathbf{H}}^*) (\tilde{\mathbf{H}}\tilde{\mathbf{H}}^*)^{-1}, \quad (3.12)$$

which reduces to

$$\mathbf{W}_{ZF}^* \mathbf{W}_{ZF} = \alpha_{ZF}^2 ((\tilde{\mathbf{H}}\tilde{\mathbf{H}}^*)^{-1})^*. \quad (3.13)$$

Then using the fact that for any invertible matrix \mathbf{A} [32]

$$(\mathbf{A}^{-1})^* = (\mathbf{A}^*)^{-1} \quad (3.14)$$

and (3.11), then (3.13) can be simplified to

$$\mathbf{W}_{ZF}^* \mathbf{W}_{ZF} = \alpha_{ZF}^2 (\tilde{\mathbf{H}}\tilde{\mathbf{H}}^*)^{-1}. \quad (3.15)$$

Via simulation it was seen that

$$\mathbb{E}[(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^*)^{-1}] = \frac{1}{M-K} \begin{bmatrix} \frac{1}{\delta_1} & 0 & \dots & 0 \\ 0 & \frac{1}{\delta_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\delta_K} \end{bmatrix}, \quad K < M \quad (3.16)$$

Using (3.15) and (3.16), the definition of $\tilde{\zeta}_k$ from (2.20) and the fact \tilde{h}_{km} and s_k are uncorrelated gives

$$\mathbb{E}[\mathbf{s}[n]^* \mathbf{W}_{ZF}^* \mathbf{W}_{ZF} \mathbf{s}[n]] = \frac{\alpha_{ZF}^2 \sum_{k=1}^K \frac{\tilde{\zeta}_k}{\delta_k}}{M-K} = 1 \quad (3.17)$$

which when rearranged for α_{ZF} gives

$$\alpha_{ZF} = \sqrt{\frac{M-K}{\sum_{k=1}^K \frac{\tilde{\zeta}_k}{\delta_k}}} \quad (3.18)$$

3.2 Array Gain

3.2.1 Array Gain for MRT

The gain due to diversity defined by (2.25) from [20] is

$$g_k \equiv \frac{\mathbb{E}[s_k[n]^* r_k]}{\sqrt{\delta_k \tilde{\zeta}_k}}$$

Using the definition for the received signal in the absence of channel estimation error, r_k , from (2.22), and the definition of W_{MRT} from (2.17)

$$g_{k\text{MRT}} = \frac{\mathbb{E}[s_k^*[n] \tilde{\mathbf{H}} \alpha_{\text{MRT}} \tilde{\mathbf{H}}^* s_k[n]]}{\sqrt{\delta_k \tilde{\zeta}_k}} \quad (3.19)$$

Using (3.6), the fact the \tilde{h}_{km} and s_k are uncorrelated and the definition of $\tilde{\zeta}_k$ from (2.20) gives

$$g_{k\text{MRT}} = \alpha_{\text{MRT}} M \delta_k \quad (3.20)$$

Using the definition of α_{MRT} from (3.8) and simplifying for δ_k and $\tilde{\zeta}_k$ gives

$$g_{k\text{MRT}} = \sqrt{\frac{\delta_k M}{\sum_{k=1}^K (\tilde{\zeta}_k \delta_k)}} \quad (3.21)$$

Note that if all the δ_k are equal, (3.21) reduces to $g_k = \sqrt{M}$, which is the array gain in [20].

3.2.2 Array Gain for ZF

Again using the definition of g_k from (2.25) and the definition of r_k from (2.22), but instead substituting the definition of W_{ZF} from (2.18), gives

$$g_{k\text{ZF}} = \frac{\mathbb{E}[s_k^*[n] \tilde{\mathbf{H}} \alpha_{\text{ZF}} \tilde{\mathbf{H}}^* (\tilde{\mathbf{H}} \tilde{\mathbf{H}}^*)^{-1} s_k[n]]}{\sqrt{\delta_k \tilde{\zeta}_k}} \quad (3.22)$$

which using the definition of $\tilde{\zeta}_k$ from (2.20) reduces to

$$g_{k\text{ZF}} = \frac{\alpha_{\text{ZF}}}{\sqrt{\delta_k}} \quad (3.23)$$

Then using the definition of α_{ZF} from (3.18) gives

$$g_{k\text{ZF}} = \sqrt{\frac{M - K}{\delta_k \sum_{k=1}^K (\frac{\tilde{\zeta}_k}{\delta_k})}} \quad (3.24)$$

Also note that if all the δ_k are equal, (3.24) reduces to $g_k = \sqrt{M - K}$, which is the array gain in [20].

3.3 Array Gain Variance

This section uses the original values of array gain derived in Sec.3.2 and applies them to give values of gain variance.

3.3.1 MRT Array Gain Variance

To find the values of gain variance for MRT, $I_{k\text{MRT}}$, begin with the definition of instantaneous gain fluctuation given by (2.26), reproduced below

$$i_k[n] \equiv r_k[n] - \sqrt{\delta_k} g_k s_k[n]$$

Substituting the value of array gain $g_{k\text{MRT}}$ as a function of α_{MRT} given in (3.20), the definition of $r_k[n]$ given in (2.22), then the definition of W_{MRT} given in (2.17) and the definition of i_k given in (2.26) to give

$$i_{k\text{MRT}} = \alpha_{\text{MRT}} (\tilde{\mathbf{H}}\tilde{\mathbf{H}}^* \mathbf{s}[n])_k - \alpha_{\text{MRT}} \delta_k M s_k[n] \quad (3.25)$$

Note that by (3.5)

$$(\mathbb{E}[\tilde{\mathbf{H}}\tilde{\mathbf{H}}^*])_k = I_k M \delta_k, \quad (3.26)$$

and the off diagonal elements of $(\mathbb{E}[\tilde{\mathbf{H}}\tilde{\mathbf{H}}^*])_k$ are zero. This means i_k has a distribution with zero mean and variance given by

$$I_{k\text{MRT}} \equiv (\mathbb{E}[|i_{k\text{MRT}}|^2])_k = \text{Var}[(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^*)_{kk}] = \delta_k. \quad (3.27)$$

This is the same value of gain variance used for MRT in [20].

3.3.2 ZF Array Gain Variance

Similarly to find the values of $I_{k\text{ZF}}$ use the value of $g_{k\text{ZF}}$ as a function of α_{ZF} given in (3.23), the definition of $r_k[n]$ given in (2.22), the definition of W_{ZF} given in (2.18) and the definition of i_k given in (2.26) to give

$$i_{k\text{ZF}} = \alpha_{\text{ZF}} (\tilde{\mathbf{H}}\tilde{\mathbf{H}}^* (\tilde{\mathbf{H}}\tilde{\mathbf{H}}^*)^{-1} \mathbf{s}[n])_k - \alpha_{\text{MRT}} s_k[n] = 0, \quad \forall k \quad (3.28)$$

Trivially the value of $I_{k\text{ZF}}$ is then

$$I_{k\text{ZF}} \equiv \mathbb{E}[|i_{k\text{ZF}}|^2] = 0 \quad (3.29)$$

Again, this is the same value of gain variance for ZF used in [20].

3.4 Channel Estimation Error Variance

To find the variance of the elements of the channel estimation error vector E_k , begin with the definition of the channel estimation error vector (2.23), reproduced below

$$\mathbf{e}[n] \equiv \mathbf{F}\mathbf{W}\mathbf{s}[n].$$

Note that given the variance of the values of \mathbf{F} given in (2.13) and the variance of the elements of $\mathbf{u}[n] = \mathbf{W}\mathbf{s}[n]$ from (2.16) and the fact that the elements of \mathbf{F} and $\mathbf{u}[n]$ are uncorrelated, the value of E_k can be defined as

$$E_k = 1 - \delta_k \quad (3.30)$$

This is the same for MRT and ZF and is the same value as [20].

3.5 Mean SINR Values for Linear System

Having derived original values for g_k for both MRT (3.21) and ZF (3.24), this results in a expanded expressions for SINR_k for MRT and ZF. These expressions are then used to generate an expression for maximum-minimum optimisation by varying downlink power distribution ζ_k in Sec.3.6 for the linear case and a maximum-minimum optimisation process in Sec.3.7.

3.5.1 Mean SINR Value MRT, Linear System

The expression value for SINR_k for given values of g_k , I_k and E_k is given in (2.33) which is derived from [20] and is reproduced below

$$\text{SINR}_k = \frac{\eta\delta_k\zeta_k\beta_k g_k^2}{\eta\beta_k(I_k + E_k) + 1}.$$

For MRT, substituting (3.21) for g_k , (3.27) for I_k and (2.29) for E_k into (2.33) gives a predicted mean term SINR_k of

$$\text{SINR}_{k\text{MRT}} = \frac{M\eta\beta_k\delta_k^2\zeta_k}{(\eta\beta_k + 1)\sum_{k=1}^K\zeta_k\delta_k}. \quad (3.31)$$

3.5.2 Mean SINR Value for ZF, Linear System

For ZF, substituting (3.24) for $g_{k\text{ZF}}$, (3.29) for $I_{k\text{ZF}}$ and (2.29) for $E_{k\text{ZF}}$ into (2.33) gives

$$\text{SINR}_{k\text{ZF}} = \frac{(M - K)\eta\beta_k\zeta_k}{(\eta\beta_k(1 - \frac{B\rho\zeta_k\beta_k}{1+B\zeta_k\rho\beta_k}) + 1)\sum_{k=1}^K\frac{\zeta_k}{\delta_k}} \quad (3.32)$$

3.6 Linear Optimisation

In this section expressions for optimising maximum-minimum SINR by varying the downlink power allocation ζ_k for MRT and ZF are derived. In Sec.3.6.1 a general solution for optimising the form of equation is outlined, while in Sec.3.6.2 and Sec.3.6.3 explicit expression for optimising for MRT and ZF, respectively, are outlined.

3.6.1 General Expression

This section derives an expression to optimise maximum-minimum SINR in a general case which is similar to [19], however some intermediate results are different and the final solution takes a different form. Given the aim to optimise the minimum maximum SINR, it can be stated at the desired value of ζ_k

$$\text{SINR}_k = \text{SINR}, \quad \forall k. \quad (3.33)$$

That is all SINR_k are equal. Both equations for SINR_k , (3.31) for MRT and (3.32) for ZF, can be written in the form

$$\text{SINR}_k = a_k\zeta_k, \quad \forall k. \quad (3.34)$$

where the a_k have different values for MRT and ZF. It is then possible to create a solution to optimise the SINR. Given (3.33) and (3.34)

$$a_1\tilde{\zeta}_1 = a_2\tilde{\zeta}_2 = \dots = a_K\tilde{\zeta}_K \quad (3.35)$$

Selecting arbitrarily to rearrange (3.35) for $\tilde{\zeta}_1$ gives

$$\tilde{\zeta}_1 = \frac{a_k\tilde{\zeta}_k}{a_1}, \quad \forall k, \quad (3.36)$$

and setting arbitrarily to write $\tilde{\zeta}_1$ as a function of $\tilde{\zeta}_K$ gives

$$\tilde{\zeta}_1 = \frac{a_K\tilde{\zeta}_K}{a_1}. \quad (3.37)$$

Substituting (2.4), i.e., that the $\tilde{\zeta}_k$ sum to 1, into (3.37) gives,

$$\tilde{\zeta}_1 = \frac{a_K}{a_1} \left(1 - \sum_{k=1}^{K-1} \tilde{\zeta}_k \right). \quad (3.38)$$

Rearranging (3.36) for $\tilde{\zeta}_k$ and substituting into (3.38) gives

$$\tilde{\zeta}_1 = \frac{a_K}{a_1} \left(1 - a_1\tilde{\zeta}_1 \sum_{k=1}^{K-1} \left(\frac{1}{a_k} \right) \right). \quad (3.39)$$

Rearranging for $\tilde{\zeta}_1$ gives

$$\tilde{\zeta}_1 = \left(\frac{a_K}{a_1} \right) \left(\frac{1}{1 + \sum_{k=1}^{K-1} \left(\frac{1}{a_k} \right)} \right). \quad (3.40)$$

Once $\tilde{\zeta}_1$ is known, it is possible to rearrange (3.36) for $\tilde{\zeta}_k$, to give all other $\tilde{\zeta}_k$.

3.6.2 Linear Optimisation for MRT

Having found a general downlink optimisation for maximum-minimum SINR as a function of downlink power allocation, it is possible to apply the solution to optimise $\text{SINR}_{k\text{MRT}}$ as given in (3.31), reproduced below

$$\text{SINR}_{k\text{MRT}} = \frac{M\eta\beta_k\delta_k^2\tilde{\zeta}_k}{(\eta\beta_k + 1)\sum_{k=1}^K\tilde{\zeta}_k\delta_k}.$$

As stated in (3.33), the maximum-minimum occurs when all SINR_k are equal, that is

$$\text{SINR}_1 = \text{SINR}_2 = \dots = \text{SINR}_K. \quad (3.41)$$

Substituting (3.31) into (3.41) gives

$$\frac{M\eta\beta_1\delta_1^2\zeta_1}{(\eta\beta_1+1)\sum_{k=1}^K\zeta_k\delta_k} = \frac{M\eta\beta_2\delta_2^2\zeta_2}{(\eta\beta_2+1)\sum_{k=1}^K\zeta_k\delta_k} = \dots = \frac{M\eta\beta_K\delta_K^2\zeta_K}{(\eta\beta_K+1)\sum_{k=1}^K\zeta_k\delta_k}. \quad (3.42)$$

By removing common terms, the (3.42) can be reduced to

$$\frac{\beta_1\delta_1^2\zeta_1}{(\eta\beta_1+1)} = \frac{\beta_2\delta_2^2\zeta_2}{(\eta\beta_2+1)} = \dots = \frac{\beta_K\delta_K^2\zeta_K}{(\eta\beta_K+1)}. \quad (3.43)$$

This can be seen to be an equation in the form given in (3.35) where

$$a_{k\text{MRT}} = \frac{\beta_k\delta_k^2}{\eta\beta_k+1}, \quad \forall k, \quad (3.44)$$

which can be substituted into (3.40) to give

$$\zeta_{1\text{MRT}} = \left(\frac{\beta_K\delta_K^2(\eta\beta_1+1)}{\beta_1\delta_1^2(\eta\beta_K+1)} \right) \left(\frac{1}{1 + \sum_{k=1}^{K-1} \left(\frac{\eta\beta_k+1}{\beta_k\delta_k^2} \right)} \right). \quad (3.45)$$

Finally, $\zeta_{1\text{MRT}}$ can be combined with (3.36) to give all other $\zeta_{k\text{MRT}}$.

3.6.3 Linear Optimisation for ZF

Similarly using equation (3.32), reproduced below, a linear optimisation formula for ZF is derived

$$\text{SINR}_{k\text{ZF}} = \frac{(M-K)\eta\beta_k\zeta_k}{(\eta\beta_k(1 - \frac{B\rho\zeta_k\beta_k}{1+B\zeta_k\rho\beta_k}) + 1)\sum_{k=1}^K \frac{\zeta_k}{\delta_k}}.$$

Using (3.35) where after simplification by removing common terms gives

$$a_{k\text{ZF}} = \frac{\beta_k}{\eta\beta_k(1 - \delta_k) + 1}, \quad \forall k \quad (3.46)$$

as the form of the equation is the same, it can be substituted into (3.40) with $a_{k\text{ZF}}$ to give

$$\zeta_{1ZF} = \left(\frac{\beta_K(\eta\beta_1(1-\delta_1)+1)}{\beta_1(\eta\beta_K(1-\delta_K)+1)} \right) \left(\frac{1}{1 + \sum_{k=1}^{K-1} \left(\frac{\eta\beta_k(1-\delta_k)+1}{\beta_k} \right)} \right). \quad (3.47)$$

Again by rearranging (3.36), it is possible to derive all other ζ_{kZF} from ζ_{1ZF} .

3.7 Optimisation Process with Nonlinear Amplifiers

Having outlined linear optimisation expressions in Sec.3.6, this section will apply those expressions as part of a process to approximate the optimal downlink power distribution, that is, the values of ζ_k for maximum-minimum SINR in a nonlinear system. This process is applied to a MATLAB simulation of a specific cell system in Chap.4 for MRT and ZF. Simulating a cell system involved generating many instances of large scale fading and for each instance of large scale fading generating many instances of small scale fading.

The general equation for SINR_k , for a nonlinear system was given in (2.44) and is reproduced below

$$\text{SINR}_k = \frac{\delta_k \zeta_k \eta \beta_k |g_k + c_k|^2}{\eta \beta_k ((I_k + E_k) |1 + \rho_k|^2 + D_k) + 1}$$

This section will outline a process to maximise the minimum value SINR_k of (2.44) by varying ζ_k . The value of the average downlink SNR, η , is fixed for the system. While, for a given instance of large scale fading, several parameters are known *a priori* without the need for small scale fading generation. These are g_k , I_k , E_k , δ_k and β_k . These are the parameters required to describe a linear system SINR_k as given in (2.33). The linear optimisation formula for MRT is given in Sec.3.6.2, while for ZF the linear optimisation formula is given in Sec.3.6.3. The values of ζ_k used to optimise a linear system are used as a first approximation to the optimum values of ζ_k in the nonlinear system.

The nonlinear optimisation process assumes that the nonlinear contributions to SINR introduced in Chap.2, that is, c_k the reduction in array gain due to saturation, ρ_k the reduction in gain variance and channel uncertainty error and D_k uncorrelated distortion are not known *a priori* for a given downlink power distribution and must be calculated empirically. It is observed for Sec.4.1, in the cell system described in

Chap.4 that the power amplifier efficiency, μ and maximum output amplitude A_{max} are not functions of ζ_k or δ_k and hence β_k , to within the level of accuracy used in this thesis. This means the radiated power P which helps determine the SNR, η , is not a function of ζ_k . Using this fact it was decided to separate the optimisation process into two distinct parts. Firstly, optimising SINR_k for a fixed β_k and ζ_k by varying the efficiency of the PAs, μ , for a fixed consumed power. Having determined the value of μ to use, the values of ζ_k are then found. This is done by iteratively by making successive approximations to the values of c_k , ρ_k and D_k and using them to find approximations to the optimal values of ζ_k .

A detailed explanation of the steps in the process to estimate the power distribution to maximise the maximum-minimum SINR is as follows.

1. Determine the median value of the power scaling factor for the particular large scale fading distribution used as described in Sec.2.1. Assign users this value of $\beta_k, \forall k$.
2. Assign equal power to all the users in the cell which by (2.4) means $\zeta_k = \frac{1}{M}, \forall k$.
3. Generate an instance of small scale fading \mathbf{H} from Sec.2.2, a pilot uplink signal at the base station receivers using (2.6) and an approximation to the small scale fading $\tilde{\mathbf{H}}$ using (2.7).
4. Generate unique sets of QPSK message symbols, s_k , to send to all the users with magnitude ζ_k as described by (2.20).
5. Generate MRT precoding matrix using $\tilde{\mathbf{H}}$ and (2.17) or ZF using (2.18).
6. Apply precoding matrix to message symbols and generate input to filters $\mathbf{u}[n]$ using (2.14).
7. Upsample $\mathbf{u}[n]$ and pass through transmit filters as outlined in Sec.2.6.
8. Create a set of input 1dB backoffs IBO_{dB} to generate a set of power scalars, b , using (2.52) for the upsampled filtered signal.
9. Pass the scaled signal through the nonlinear PA as described in (2.54). Note that A_{max} is unknown at this time.

10. Expand the running averages of the transmit power of the PA output using (2.55) and efficiency using (2.56).
11. Return to Step 3 if more instances of small scale fading are required.
12. Now have values of efficiency, μ , for different backoff values, which can be applied to (2.58) to find A_{\max} for different backoff values.
13. Can apply known values of μ and A_{\max} and repeat Step 3 to Step 7 to generate PA outputs for every instance of small scale fading which then passes through the channel to give a received signal to the users given by (2.59).
14. The received signal can be used to generate an average SINR for a given backoff using (2.30) across many instances of small scale fading. The value of input backoff which gives the peak SINR is used for the rest of the process.
15. Generate an instance of large scale fading coefficients β_k from the distribution described in Sec.2.1.
16. Use the expressions found for linear optimisation in either Sec.3.6.2 for MRT or Sec.3.6.3 for ZF to find optimal ξ_k values in a linear system for the given set of β_k .
17. Run Step 13 across many instances of small scale fading to find estimations of the values of c_k using (2.39), ρ_k using (2.40) and D_k (2.42) at the level of backoff found in Step 14 at the values of ξ_k found in Step 16.
18. Use the values of c_k , ρ_k and D_k found in Step 17 and use the nonlinear optimisation estimate outlined in Sec.3.7.1 for MRT or Sec.3.7.2 for ZF to derive new estimates of the optimal values for ξ_k . These ξ_k comprise the values of the first nonlinear optimisation iteration.
19. Use the values of ξ_k found in Step 18 and run Step 13 for many instances of small scale fading to find new values of c_k , ρ_k and D_k .
20. Use the values of c_k , ρ_k and D_k found in Step 19 and use the nonlinear optimisation estimate outlined in Sec.3.7.1 for MRT or Sec.3.7.2 for ZF to derive

new estimates of the optimal values for ζ_k . These ζ_k comprise the values of the second nonlinear optimisation iteration.

21. If additional instances of large scale fading are required return to Step 15 or else stop.

The process outlined in this section is used to produce CDFs of the minimum SINR values in Chap.4. In section Sec.3.7.1 an expression for a_k for MRT given which is later used for nonlinear maximum-minimum SINR optimisation. While Sec.3.7.2 does the same for ZF.

3.7.1 Nonlinear Optimisation MRT

It can be seen that the general equation for SINR_k in a nonlinear system given in (2.44), is an equation of the form outlined in Sec.3.6.1 with the same goal of optimising the maximum-minimum SINR. Substituting (2.44) into (3.34) and rearranging for $a_{k\text{NL}}$ gives

$$a_{k\text{NL}} = \frac{\eta \delta_k \beta_k |g_k + c_k|^2}{\eta \beta_k (|1 + \rho_k|^2 + D_k) + 1} \quad (3.48)$$

For MRT using (3.21) for g_k , (3.27) for I_k and (2.29) for E_k gives

$$a_{k\text{MRT}} = \frac{\eta \delta_k \beta_k \left| \sqrt{\frac{\delta_k M}{\sum_{k=1}^K (\zeta_k \delta_k)}} + c_k \right|^2}{\eta \beta_k (|1 + \rho_k|^2 + D_k) + 1} \quad (3.49)$$

3.7.2 Nonlinear Optimisation ZF

For ZF, using (3.24) for g_k , (3.29) for I_k and (2.29) for E_k gives

$$a_{k\text{ZF}} = \frac{\eta \delta_k \beta_k \left| \sqrt{\frac{M-K}{\delta_k \sum_{k=1}^K (\frac{\zeta_k}{\delta_k})}} + c_k \right|^2}{\eta \beta_k ((1 - \delta_k) |1 + \rho_k|^2 + D_k) + 1} \quad (3.50)$$

3.8 Conclusion

This chapter derived normalisation factors for MRT and ZF which the author believes are original. By applying these normalisation factors, explicit values for the

linear contributors to downlink SINR, introduced in Chap.2 for MRT and ZF, are derived, including values for the array gains which the author also believes are original. The values of the linear contributions to downlink SINR: array gain, array gain variance, and channel estimation error were integrated into the downlink SINR expression introduced in Chap.2. This was used to derive an expression to optimise the maximum-minimum downlink SINR by varying the downlink power distribution. A process is introduced to approximate the downlink power distribution to optimise the maximum-minimum SINR in the presence of PA nonlinearity when the base station power consumption is fixed. This process outlines the generation of empirical values for the nonlinear elements of downlink SINR as a function of downlink power distribution. Chap.4, applies the process in a simulated cell across many instances of large scale fading to show the benefits of applying the process to Massive MIMO systems using MRT and ZF.

Chapter 4

Results

This chapter will apply the process for optimising the maximum-minimum SINR in a nonlinear system outlined in Sec.3.7 to a particular cell experiencing large and small scale fading. Before applying the process it was necessary to derive various PA parameters introduced in Sec.2.8 via simulation. This is done in Sec.4.1. Next, these parameters are applied to the first steps of the nonlinear optimisation process, involving estimating the level of PA input backoff, done in Sec.4.2. Having found the input backoff level, simulations of a nonlinear downlink system are performed over many instances of small and large scale fading for ZF in Sec.4.3 and MRT in Sec.4.4. For each instance of large scale fading the remaining steps of the nonlinear optimisation process are performed. This makes it possible to compare the minimum SINR, i.e., the value of SINR intended to be optimised, for linear optimisation, first nonlinear optimisation iteration and second nonlinear optimisation iteration in all the instances of large scale fading. The relative benefits of the optimisation stages can then be seen on CDF plots for MRT and ZF.

4.1 Nonlinear Amplifier Derived Constants for MRT and ZF

To find the required input backoff, first it was necessary to create a MATLAB simulation of the nonlinear Rapp model amplifier with $p = 2$ defined in (2.54) and introduced in Sec.2.8. $p = 2$ is considered a reasonable value for a moderate cost PA [20, 22]. Values of p between 2 and 4 are considered representative [22], with a higher p value corresponding to more linear performance. The PA simulation was run for different input backoff values IBO_{dB} . These use constants given in Table 4.2.

Note that b was calculated using (2.52), μ was estimated using (2.56). While A_{\max} was estimated using (2.54) substituted into (2.55), combined with (2.58).

No dependence was observed for A_{\max} and μ on ξ_k , δ_k or whether MRT or ZF precoding was used to $2sf$. This was confirmed for δ_k by running simulations of cells where all users had maximum $\delta_k = 1, \forall k$ and minimum $\delta_k = 0.9091 \forall k$ corresponding to the maximum and minimum possible values of β_k . For ξ_k this was confirmed by running simulations where all the users had equal $\xi_k = 0.1, \forall k$ and $\xi_1 = 1$ while $\xi_k = 0, k = 2, \dots, K$. That is, even for the most extreme values of δ_k and ξ_k the approximations given in Table 4.1 hold. This make the application of a fixed input backoff across different instances of large scale fading a viable approach.

TABLE 4.1: Nonlinear Amplifier Derived Constants

IBO _{dB}	6	4	2	1	0	-1	-2	-3
$b(5sf)$	5.2057	3.2846	2.0724	1.6462	1.3076	1.0387	0.8251	0.6554
$\mu(2sf)$	0.37	0.44	0.51	0.54	0.57	0.60	0.63	0.65
$A_{\max}(2sf)$	0.24	0.20	0.17	0.16	0.15	0.14	0.135	0.13

4.2 Estimating Level of PA Input Backoff

In order to find an approximation to the optimal level of IBO_{dB} for MRT and ZF, a simulation was performed with a cell with β_k all equal to the median value for the cell and equal ξ_k . Using the values of b , μ and A_{\max} given in Table 4.1, and combined with the definition for SINR _{k} given in (2.30) with the definition of $y_k[n]$ for a nonlinear system given in (2.38), separate plots for MRT and ZF were generated. Both MRT and ZF used the constants given in Table 4.2.

4.2.1 MRT PA Input Backoff

Figure 4.1 shows the peak value of SINR in a nonlinear MRT system with equal β_k with a red "x". It can be seen that the peak SINR occurs at input backoff IBO_{dB} = -2dB or efficiency $\mu = 0.6297$ (4sf).

TABLE 4.2: Nonlinear SINR constants, fixed β_k

Description	Symbol	Value
Number of Users	K	10
Number of Base Station Antennas	M	100
Pilot Length (Symbols)	B	10
Pilot Uplink SNR	ρ	1
Number of Small Scale Fades	N_H	500
Message Length (Symbols)	N_s	500
Large Scale Fading Coefficient	β_k	3.7, $\forall k$
Downlink Power Distribution	ζ_k	$\frac{1}{M} = 0.1, \forall k$
Filter roll-off	σ	0.22
Filter Samples per Symbol	κ	7
Filter span (Symbols)	τ	10
Data Downlink SNR	η	μ

4.2.2 ZF PA Input Backoff

Similarly, Figure 4.2 shows the peak value of SINR for a nonlinear ZF system with equal β_k with a red "x". It can be seen that the peak SINR also occurs at $\text{IBO}_{\text{dB}} = -2\text{dB}$ or $\mu = 0.6297$ (4sf).

4.3 Nonlinear Optimisation Results, ZF

The nonlinear optimisation process described in Sect.3.7 was applied to a cell experiencing large and small scale fading generated using the constants given in Table 4.2 with exceptions being in β_k and ζ_k not being constants and the changes and additions given in Table 4.3.

TABLE 4.3: Nonlinear optimisation constants, multiple large scale fades

Description	Symbol	Value
Number of Small Scale Fades per Large Scale Fade	N_H	100
Number of Large Scale Fades	N_β	380
Cell Radius	R	100
Cell Exclusion Zone	r	1
Reference Distance	d_0	100
Path Loss Exponent	α_s	3.8

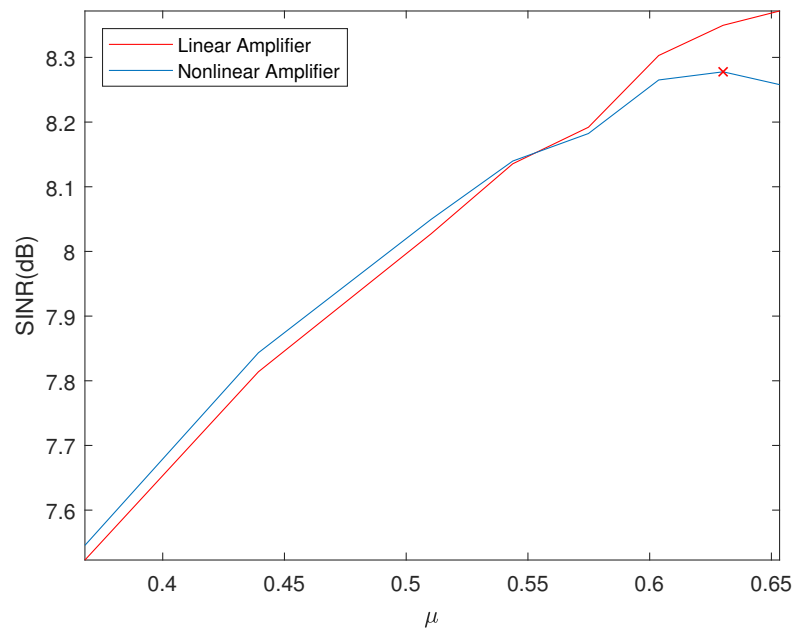


FIGURE 4.1: Average SINR of received signal, MRT

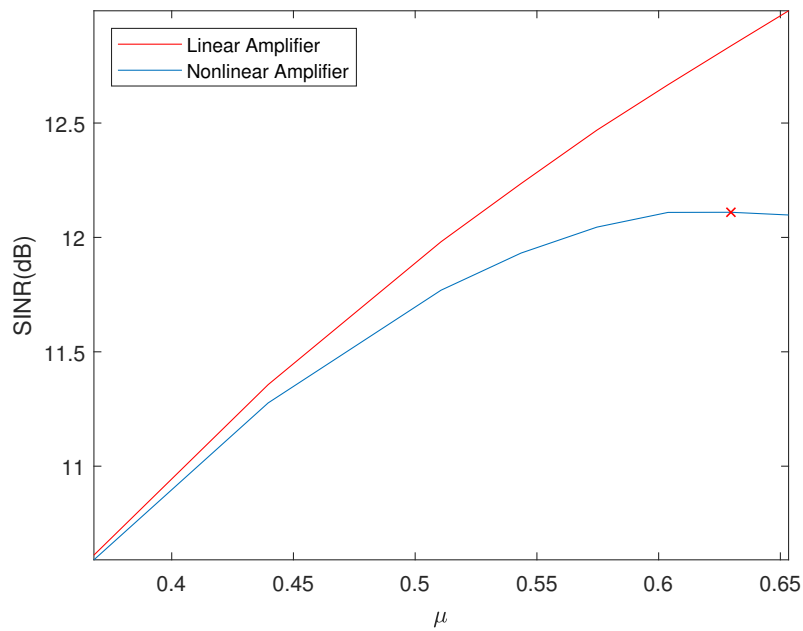


FIGURE 4.2: Average SINR of received signal, ZF

The results of this for ZF are displayed in Figure 4.3 across many large scale fades to give a cdf. It can be seen that the $\min(\text{SINR}_k)$ is greatly affected by nonlinearity and the first iteration of nonlinear optimisation provides substantial benefit. The benefit of a second iteration, however, is significantly less and can not be seen on the

scale of the plot.

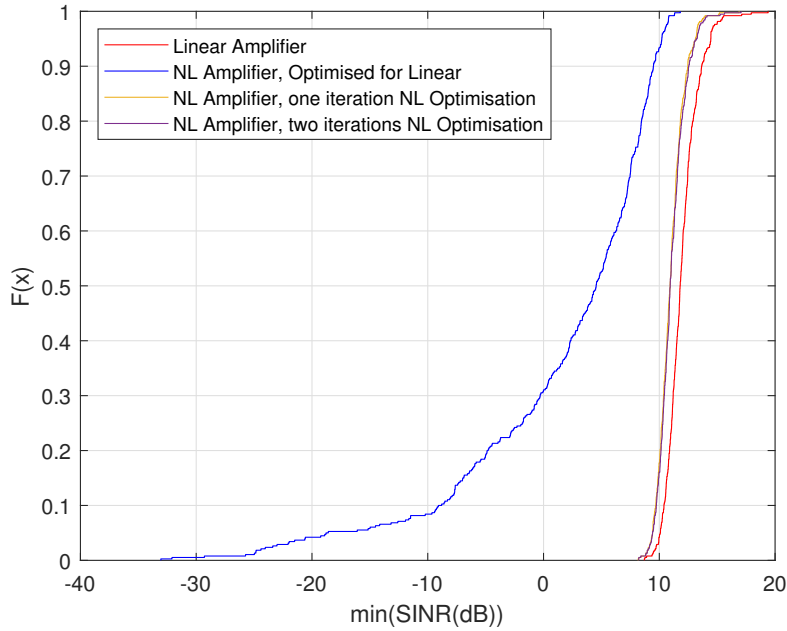


FIGURE 4.3: CDF of SINR for ZF across multiple large scale fades

In a version on the same plot with a smaller scale, Figure 4.4, a small benefit can be seen a second iteration, however it is a small fraction of a dB, with the vast amount of benefit from taking into account nonlinearity obtained on the first iteration.

The process defined in Sec.3.7 produces substantial benefits for ZF. The reason for this can be seen in Figure 4.5, Figure 4.6 and Figure 4.7, which plot the values of c_k , the distortion correlated to the wanted signal, D_k , the uncorrelated distortion term and ρ_k , the distortion correlated to the interference and channel estimation error, respectively as a function of the downlink allocation ζ_k . It can be seen that c_k is negatively correlated with ζ_k , D_k is positively correlated and ρ_k is effectively zero. When substituted into (2.44), there is a clear relation between ζ_k and the loss of performance due to nonlinearity, because an increase in ζ_k results in increased effects of nonlinearity from both a decrease in the numerator due to decreased c_k and an increase in the denominator due to a increase in D_k . Both effects work together to reduce SINR_k , while ρ_k has no effect. This leads to the clear reduction in $\min(\text{SINR})$ seen in Figure 4.2 and makes optimising SINR for nonlinear effects very beneficial for at least one iteration.

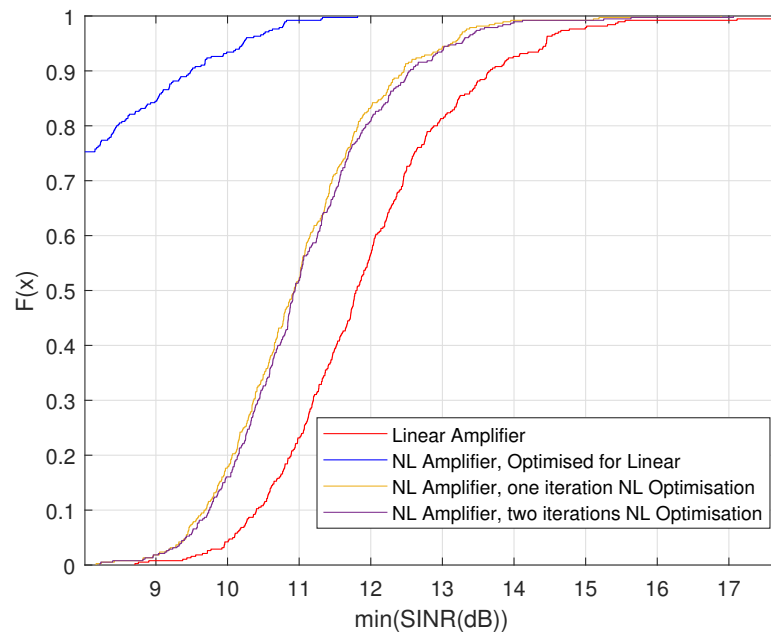


FIGURE 4.4: CDF of SINR for ZF across multiple large scale fades, smaller axis range

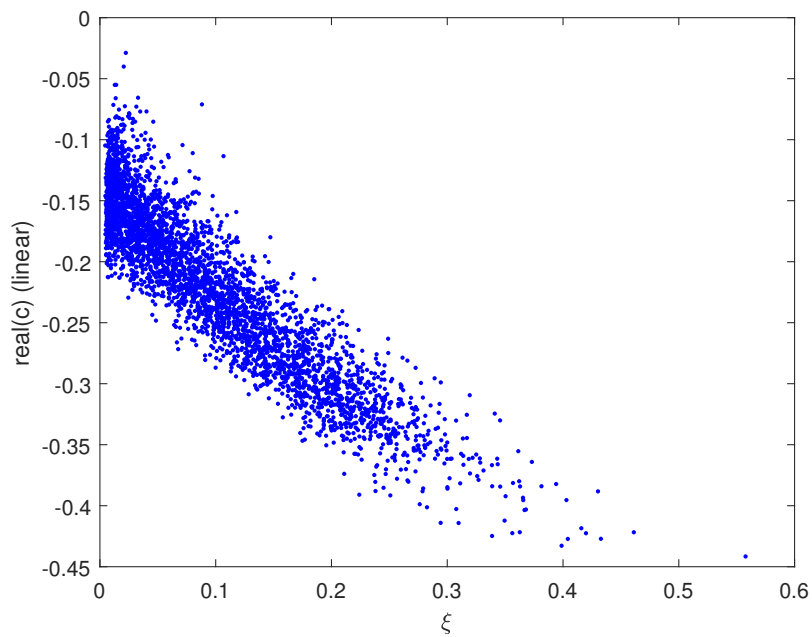
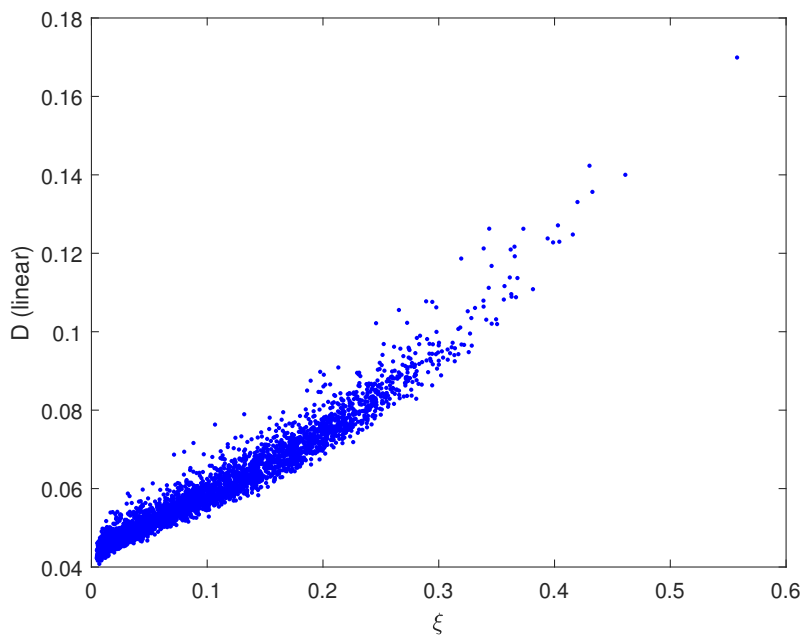
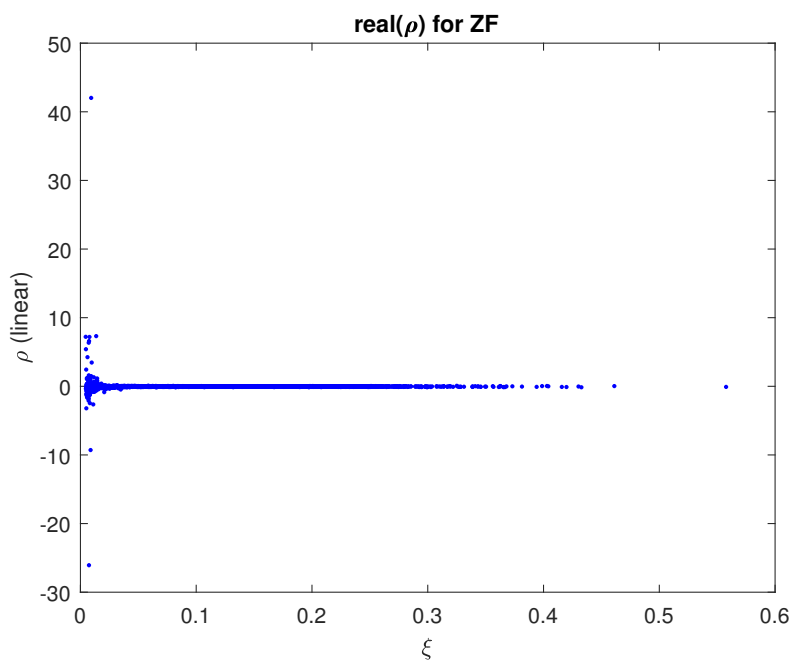


FIGURE 4.5: C as a function of ξ , ZF

FIGURE 4.6: D as a function of ξ , ZFFIGURE 4.7: ρ as a function of ξ , ZF

4.4 Nonlinear Optimisation Results, MRT

Using the same constants used in the ZF simulations given in Table 4.3, a cdf for MRT for many instances of large scale fading was generated in Figure 4.8. Interestingly, MRT is far less affected by nonlinearity than ZF. It also benefits far less from

being optimised for nonlinearity, with essentially no benefit to taking into account nonlinearity in optimising ζ_k for $\min(\text{SINR})$. It can be seen that the optimised for linear curve, one iteration nonlinear optimisation curve and the two iteration nonlinear optimisation curves essentially overlap in Figure 4.8.

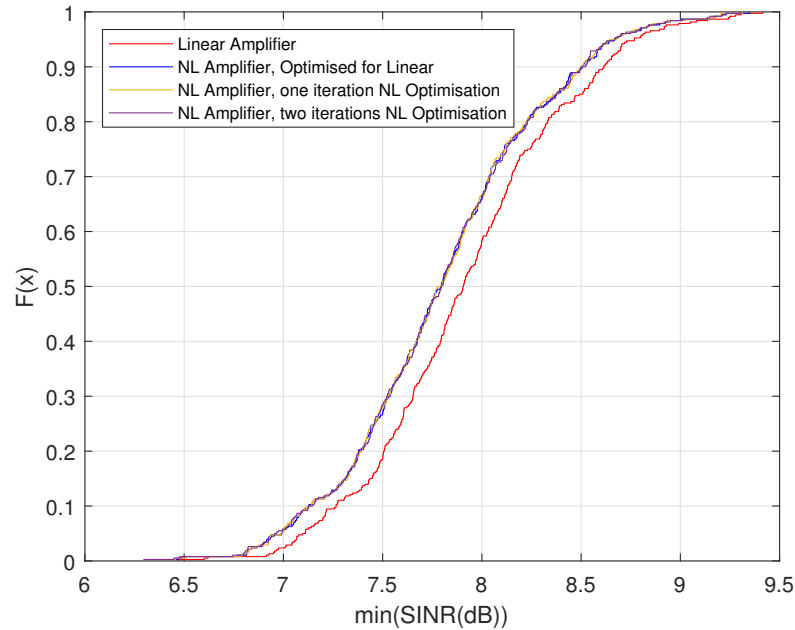


FIGURE 4.8: CDF of SINR for MRT across multiple large scale fades

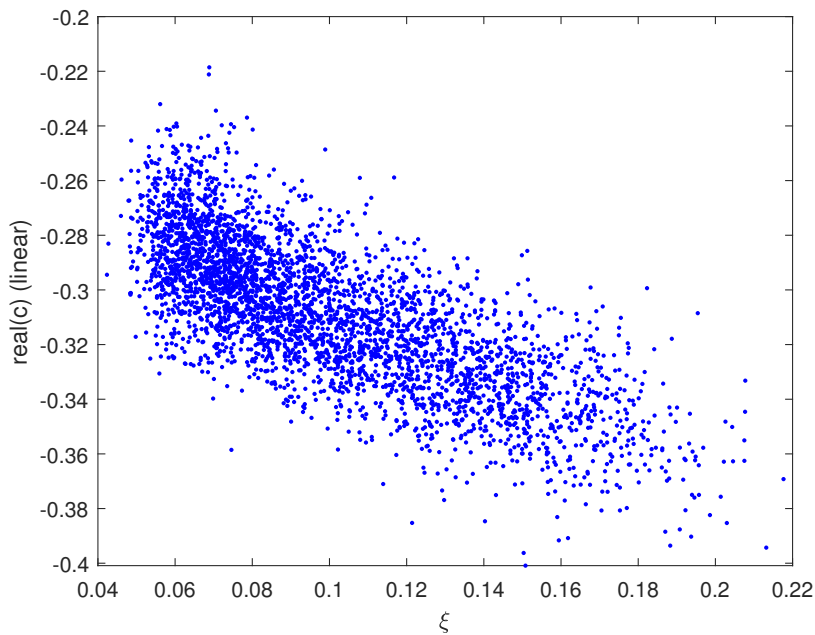
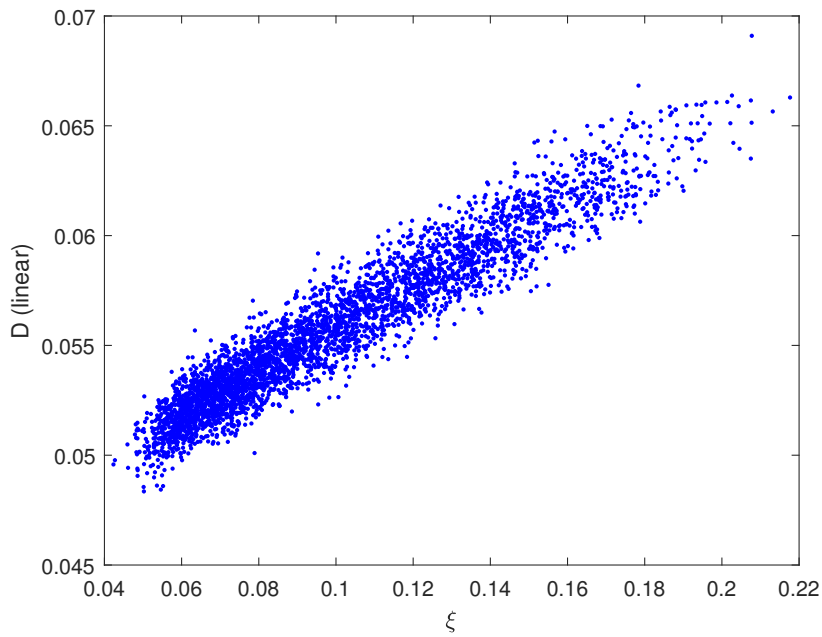
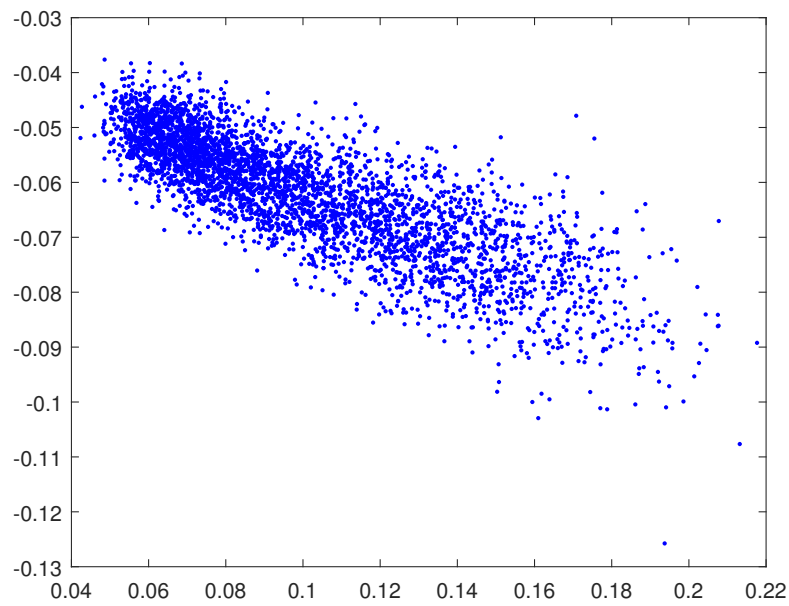


FIGURE 4.9: C as a function of ζ , MRT

FIGURE 4.10: D as a function of ξ , MRTFIGURE 4.11: ρ as a function of ξ , MRT

The reason that MRT is not affected by nonlinearity and does not benefit from optimisation to the same extent as ZF can be seen in Figure 4.9, Figure 4.10 and Figure 4.11. Again, similar to ZF, c_k is negatively correlated to ξ_k , while D_k is positively

correlated. A key difference is that ρ_k is negatively correlated to ξ_k , due to the significantly larger gain variance term, I_k . For MRT the reduction in interference due to saturation provides a substantial benefit that offsets the loss in SINR due to c_k and D_k . Note that when a similar simulation was done with a much lower downlink SNR $\eta = 0.1$, negative effects of nonlinearity were observed and could be partially countered by optimisation for nonlinearity. The CDF from that simulation is given in Figure 4.12. There is a visible improvement between optimisation for linear and one iteration for nonlinear optimisation.

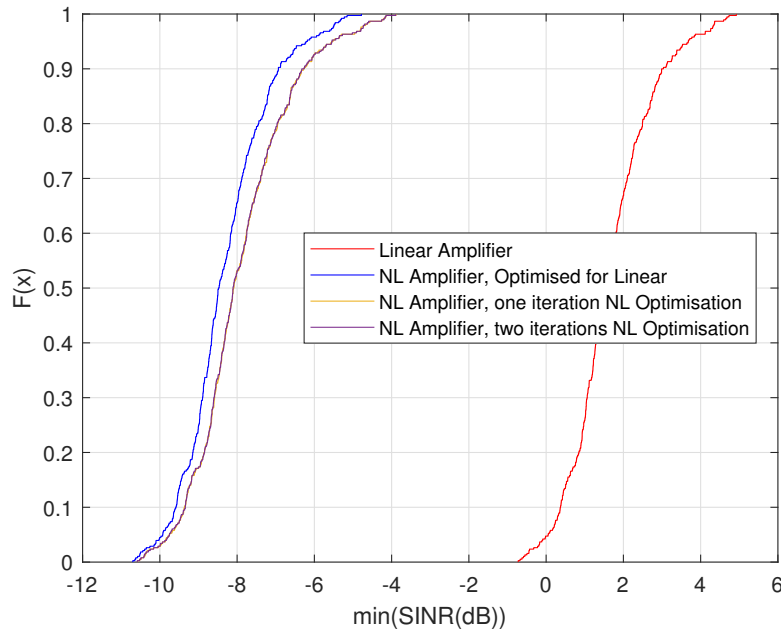


FIGURE 4.12: CDF of SINR for MRT across multiple large scale fade, $\eta = 0.1$

4.5 Conclusion

This chapter applied the nonlinear maximum-minimum SINR optimisation process by varying downlink power allocation as outlined in Sec.3.7 to a cell experiencing large and small scale fading. As an intermediate step, it estimated through simulation parameters for the PA model introduced in Sec.2.8. It found an approximate value for the input backoff for the nonlinear PAs to maximise SINR using MRT and

ZF precoding. By applying the nonlinear optimisation process across different instances of large scale fading, the effects of PA nonlinearity could be observed and the benefits of applying the optimisation process could be seen. It was noted that ZF precoding is substantially affected by PA nonlinearity and the average minimum SINR is greatly improved by the application of the optimisation process. MRT precoding was less affected by PA nonlinearity as so benefited less from the application of the optimisation process. Plots of array gain, array gain variance and uncorrelated distortion were also plotted as a function of downlink power distribution for ZF and MRT and the observed correlation was noted.

Chapter 5

Conclusion and Future Work

This thesis presented a detailed model of a Massive MIMO system. This system included large and small scale fading, an MMSE estimator matrix for small scale fading in a cell where users are experiencing different large scale fading, a parameter to control the downlink power distribution, the Rapp model for a nonlinear PA and a breakdown of the components of downlink SINR for user k for with a linear and nonlinear PA at the base station. The components of linear downlink SINR include array gain, array gain variance and channel approximation error. The addition of base station nonlinearity introduced additional factors in the downlink SINR due to saturation of the PA, including array gain loss, interference and channel approximation loss and uncorrelated distortion. Two common forms of base station precoding, MRT and ZF, were introduced. The ability to isolate individual contributors to SINR was essential to allow optimisation of the maximum-minimum average SINR by varying the downlink power allocation later in the thesis.

Having expanded the contributions to linear downlink SINR, we derived explicit values for these in both MRT and ZF systems. By applying the constraint of a constant average transmit power, which is independent of the channel estimation and power downlink distribution, normalisation factors were found for ZF and MRT, which to the knowledge of the author are original. By applying the normalisation factor, explicit values for the array gains, array gain variances and channel approximation errors were derived. As far as the author is aware, the values of the array gain for MRT and ZF are also original.

Once the values of the elements of linear downlink SINR were found for MRT and ZF, they were used to generate an expression to optimise the maximum-minimum

downlink average SINR by varying the downlink power allocation. Again as the array gains for MRT and ZF are believed to be original, the resulting derived optimisation closed form expressions are also believed to be original. Once the linear optimisation expression was found, a process was outlined to approximate the optimal downlink power distribution for SINR with maximum-minimum constraint in a nonlinear system. The derived explicit linear components were integrated into the nonlinear SINR optimisation process, while the nonlinear components were approximated via simulation. The efficacy of the nonlinear optimisation process was demonstrated for a particular cell using ZF precoding, showing the necessity of applying at least one iteration of the nonlinear process. While we demonstrated in the same cell conditions, that MRT was relatively immune to the effects of nonlinearity, and thus did not benefit significantly from application of the nonlinear optimisation process.

This thesis could lead to a range of future work. As noted in the Chap.4, simulations showed that several PA derived parameters were independent of the channel estimation and power downlink distribution, for the cell constants used in the simulations, to within the accuracy used in this thesis. It was shown in Chap.3 that the original normalisation factors ensures the average output power of the precoders, which is the input power to the PAs is independent of channel estimation error and power downlink distribution. Future work could attempt to prove the waveforms input to the PAs from different channel estimation errors and power downlink distributions are on average identical when the original normalisation factor is applied. If they are not identical, the author believes they are very similar and an approximate bound in their differences might be derived. The bound in differences would likely be related to the cell constants.

Additional future work could involve a full general analysis of nonlinear contributions to SINR as a function of the downlink power allocation. This thesis estimated through simulation the nonlinear elements of SINR on a particular cell experiencing large and small scale fading, using MRT and ZF precoding and fixed set of parameters. It was observed by the author that different cell parameters resulted in different relationships between downlink power distribution and nonlinear elements of SINR. For instance, increasing the number of users reduced the level of

correlation between the allocated downlink power and the nonlinear elements. The values of the nonlinear elements, could then be used to provide solutions, either exact or approximate, to optimise the SINR in arbitrary cells. Having solutions for an arbitrary cell would remove the need for a nonlinear optimisation process based on empirical results as used in this thesis.

Also future work could involve joint optimisation of input backoff and downlink power distribution. In this thesis, they are optimised separately with input backoff not adjusted for the specific instance of large scale fading. For cells with small numbers of users, where mean and median power scaling factors would vary substantially between instances of large scale fading, the author believes adjusting the level of input backoff for each instance could yield substantial benefits.

The results obtained in this thesis could also be applied to the highly active research areas of multi-cell or alternatively cell-free Massive MIMO.

Appendix A

Normalisation Factors Applied to Generate the Precoding Matrix Frobenius Norm

A.1 Effect of Changed Normalisation Factor on Frobenius Norm for MRT

A common value for the Frobenius norm of a precoding matrix is $\mathbb{E}[\|\mathbf{W}\|_F^2] = K$ [20, 24, 33]. K is typically the expectation of the Frobenius norm of the precoding matrix, because it does not affect the average total power when the total input power is 1, i.e., the input vector acted on by the precoding matrix will have the same average power as the output vector [24]. However that is only true if the elements of the input vector or the elements of the precoding matrix are statistically identical which they are not in this thesis. The use of α_{MRT} derived in 3.1.1 gives a different value for the Frobenius norm.

Using the definition of \mathbf{W}_{MRT} from (2.17)

$$\mathbb{E}[\|\mathbf{W}_{\text{MRT}}\|_F^2] = \mathbb{E}[\|\alpha_{\text{MRT}}\tilde{\mathbf{H}}^*\|_F^2] \quad (\text{A.1})$$

and using the fact that for any matrix \mathbf{A} [32]

$$\|\mathbf{A}\|_F^2 = \text{Tr}(\mathbf{A}^* \mathbf{A}) \quad (\text{A.2})$$

where Tr is the trace operation, then

$$\mathbb{E}[\|\mathbf{W}_{\text{MRT}}\|_{\text{F}}^2] = \alpha_{\text{MRT}}^2 \mathbb{E}[Tr(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^*)] \quad (\text{A.3})$$

using (3.6), (A.3) can be written

$$\mathbb{E}[\|\mathbf{W}_{\text{MRT}}\|_{\text{F}}^2] = \alpha_{\text{MRT}}^2 M \sum_{k=1}^K \delta_k \quad (\text{A.4})$$

substituting the derived value for α_{MRT} (3.8) into (A.4) gives

$$\mathbb{E}[\|\mathbf{W}_{\text{MRT}}\|_{\text{F}}^2] = \frac{\sum_{k=1}^K \delta_k}{\sum_{k=1}^K \zeta_k \delta_k} \quad (\text{A.5})$$

Note that $\mathbb{E}[\|\mathbf{W}_{\text{MRT}}\|_{\text{F}}^2]$ reduces to K when all ζ_k or all the δ_k are the same and (2.4) is satisfied.

A.2 Effect of Changed Normalisation Factor on Frobenius Norm for ZF

Similarly a new Frobenius norm can be derived \mathbf{W}_{ZF} using (2.18) giving

$$\mathbb{E}[\|\mathbf{W}_{\text{ZF}}\|_{\text{F}}^2] = \alpha_{\text{ZF}}^2 \mathbb{E}[\|\tilde{\mathbf{H}}^*(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^*)^{-1}\|_{\text{F}}^2] \quad (\text{A.6})$$

Using the simplification for $\mathbf{W}_{\text{ZF}}^* \mathbf{W}_{\text{ZF}}$ used in (3.15)

$$\mathbb{E}[\|\mathbf{W}_{\text{ZF}}\|_{\text{F}}^2] = \alpha_{\text{ZF}}^2 \mathbb{E}[Tr((\tilde{\mathbf{H}}\tilde{\mathbf{H}}^*)^{-1})] \quad (\text{A.7})$$

using (3.16), (A.7) can be written

$$\mathbb{E}[\|\mathbf{W}_{\text{ZF}}\|_{\text{F}}^2] = \frac{\alpha_{\text{ZF}}^2 \sum_{k=1}^K \frac{1}{\delta_k}}{M - K} \quad (\text{A.8})$$

Substituting the value for α_{ZF} given by (3.18)

$$\mathbb{E}[\|\mathbf{W}_{\text{ZF}}\|_{\text{F}}^2] = \frac{\sum_{k=1}^K \frac{1}{\delta_k}}{\sum_{k=1}^K \frac{\zeta_k}{\delta_k}} \quad (\text{A.9})$$

Note that $\mathbb{E}[\|\mathbf{W}_{ZF}\|_{\mathbb{F}}^2]$ also reduces to K when all ξ_k or all the δ_k are the same and (2.4) is satisfied.

Appendix B

Fixed Pilot Power Optimisation

Using the values for the downlink SINR_k for MRT and ZF given in (3.31) and (3.32) it is possible to derive equations to optimise the SINR when an additional constraint is added for total pilot uplink power. In the other sections of this thesis it is assumed that all users broadcast at their maximum power during the pilot uplink, however it may be desirable for some users to broadcast at less than their maximum power for system energy efficiency or to limit the required dynamic range of the base station receivers. Due to the complexity of the equations, we will only deal with the $K = 2$ case. We will apply two power constraints simultaneously, that is both the uplink and downlink have fixed total power. Both MRT and ZF are optimised for $\min(\text{SINR}_1, \text{SINR}_2)$. This appendix assumes a perfectly linear downlink.

B.1 Optimising SINR with Configurable Pilot Uplink Power, MRT

This section will introduce a new variable ζ_k , which is the uplink pilot power allocation. This is a power scaling factor which can be applied to (2.11) for δ_k to give

$$\delta_k = \frac{B\rho\zeta_k\beta_k}{1 + B\zeta_k\rho\beta_k}. \quad (\text{B.1})$$

It is assumed without loss of generality that the total pilot uplink power is 1. That is

$$\zeta_2 = 1 - \zeta_1, \quad (\text{B.2})$$

which makes δ_1 and δ_2 configurable according to (B.1).

Starting from the uplink SINR_k equation for MRT, (3.31) reproduced below,

$$\text{SINR}_{k\text{MRT}} = \frac{M\eta\beta_k\delta_k^2\tilde{\zeta}_k}{(\eta\beta_k + 1)\sum_{k=1}^K\tilde{\zeta}_k\delta_k}$$

and substituting the value of δ_k given by (B.1), gives

$$\text{SINR}_{k\text{MRT}} = \frac{M\eta\beta_k\left(\frac{B\rho\tilde{\zeta}_k\beta_k}{1+B\tilde{\zeta}_k\rho\beta_k}\right)^2\tilde{\zeta}_k}{(\eta\beta_k + 1)\sum_{k=1}^K\tilde{\zeta}_k\left(\frac{B\rho\tilde{\zeta}_k\beta_k}{1+B\tilde{\zeta}_k\rho\beta_k}\right)}. \quad (\text{B.3})$$

Note that by using the condition that when maximising $\min(\text{SINR}_1, \text{SINR}_2)$,

$$\text{SINR}_1 = \text{SINR}_2, \quad (\text{B.4})$$

and rearranging (2.4) when $K = 2$ gives

$$\tilde{\zeta}_2 = 1 - \tilde{\zeta}_1. \quad (\text{B.5})$$

Substituting (B.3) and (B.5) into (B.4) then solving (B.4) for $\tilde{\zeta}_1$, gives an equation for the values of $\tilde{\zeta}_{1\text{MRT}}$ for which $\text{SINR}_{1\text{MRT}} = \text{SINR}_{2\text{MRT}}$,

$$\tilde{\zeta}_{1\text{MRT}} = \frac{\tilde{\zeta}_{1\text{numMRT}}}{\tilde{\zeta}_{1\text{denMRT}}}, \quad (\text{B.6})$$

where the numerator is

$$\begin{aligned} \tilde{\zeta}_{1\text{numMRT}} = & \eta B^2 \rho^2 \beta_1^3 \beta_2^3 \zeta_1^4 - 2\eta B^2 \rho^2 \beta_1^3 \beta_2^3 \zeta_1^3 + \eta B^2 \rho^2 \beta_1^3 \beta_2^3 \zeta_1^2 + B^2 \rho^2 \beta_1^2 \beta_2^3 \zeta_1^4 - 2B^2 \rho^2 \beta_1^2 \beta_2^3 \zeta_1^3 \\ & + B^2 \rho^2 \beta_1^2 \beta_2^3 \zeta_1^2 + 2\eta B \rho \beta_1^2 \beta_2^3 \zeta_1^3 - 4\eta B \rho \beta_1^2 \beta_2^3 \zeta_1^2 + 2\eta B \rho \beta_1^2 \beta_2^3 \zeta_1 + 2B \rho \beta_1 \beta_2^3 \zeta_1^3 \\ & - 4B \rho \beta_1 \beta_2^3 \zeta_1^2 + 2B \rho \beta_1 \beta_2^3 \zeta_1 + \eta \beta_1 \beta_2^3 \zeta_1^2 - 2\eta \beta_1 \beta_2^3 \zeta_1 + \eta \beta_1 \beta_2^3 + \beta_2^3 \zeta_1^2 - 2\beta_2^3 \zeta_1 \\ & + \beta_2^3, \end{aligned} \quad (\text{B.7})$$

and the denominator is

$$\begin{aligned} \tilde{\zeta}_{1\text{denMRT}} = & 2\eta B^2 \rho^2 \beta_1^3 \beta_2^3 \zeta_1^4 - 4\eta B^2 \rho^2 \beta_1^3 \beta_2^3 \zeta_1^3 + 2\eta B^2 \rho^2 \beta_1^3 \beta_2^3 \zeta_1^2 + B^2 \rho^2 \beta_1^3 \beta_2^3 \zeta_1^4 \\ & - 2B^2 \rho^2 \beta_1^3 \beta_2^3 \zeta_1^3 + B^2 \rho^2 \beta_1^3 \beta_2^3 \zeta_1^2 + B^2 \rho^2 \beta_1^2 \beta_2^3 \zeta_1^4 - 2B^2 \rho^2 \beta_1^2 \beta_2^3 \zeta_1^3 + B^2 \rho^2 \beta_1^2 \beta_2^3 \zeta_1^2 \\ & - 2\eta B \rho \beta_1^3 \beta_2^3 \zeta_1^3 + 2\eta B \rho \beta_1^3 \beta_2^3 \zeta_1^2 - 2B \rho \beta_1^3 \beta_2^3 \zeta_1^2 + 2B \rho \beta_1^3 \beta_2^3 \zeta_1 + 2\eta B \rho \beta_1^2 \beta_2^3 \zeta_1^3 \\ & - 4\eta B \rho \beta_1^2 \beta_2^3 \zeta_1^2 + 2\eta B \rho \beta_1^2 \beta_2^3 \zeta_1 + 2B \rho \beta_1 \beta_2^3 \zeta_1^3 - 4B \rho \beta_1 \beta_2^3 \zeta_1^2 + 2B \rho \beta_1 \beta_2^3 \zeta_1 \\ & + \eta \beta_1^3 \beta_2^3 \zeta_1^2 + \beta_1^3 \zeta_1^2 + \eta \beta_1 \beta_2^3 \zeta_1^2 - 2\eta \beta_1 \beta_2^3 \zeta_1 + \eta \beta_1 \beta_2^3 + \beta_2^3 \zeta_1^2 - 2\beta_2^3 \zeta_1 + \beta_2^3. \end{aligned} \quad (\text{B.8})$$

Taking (B.6) and substituting into (B.3) then differentiating with respect to $\tilde{\zeta}_1$ and setting the derivative to zero to find the value of $\tilde{\zeta}_{1\text{optMRT}}$, the value of $\tilde{\zeta}_1$ which maximise SINR gives

$$\zeta_{1\text{optMRT}} = \frac{\beta_2^2 - \beta_1\beta_2\sqrt{(\rho\beta_1 + 1)\rho\beta_2 + 1} + \rho\beta_1\beta_2^2}{\rho\beta_1^2\beta_2 + \beta_1^2 - \rho\beta_1\beta_2^2 - \beta_2^2}, \beta_1 \neq \beta_2 \quad (\text{B.9})$$

B.2 Optimising SINR with Configurable Pilot Uplink Power, ZF

Similarly for ZF starting from the equation for SINR_k for ZF (3.32), and substituting (B.1) gives

$$\text{SINR}_{k\text{ZF}} = \frac{(M - K)\eta\beta_k\tilde{\zeta}_k}{(\eta\beta_k(1 - \frac{B\rho\tilde{\zeta}_k\beta_k}{1+B\tilde{\zeta}_k\rho\beta_k}) + 1)\sum_{k=1}^K \frac{\tilde{\zeta}_k}{\frac{B\rho\tilde{\zeta}_k\beta_k}{1+B\tilde{\zeta}_k\rho\beta_k}}} \quad (\text{B.10})$$

By substituting (B.10) and (B.5) into (B.4) then solving (B.4) for $\tilde{\zeta}_1$, gives an equation for the values of $\tilde{\zeta}_1$ for which $\text{SINR}_1 = \text{SINR}_2$.

$$\tilde{\zeta}_{1\text{ZF}} = \frac{\tilde{\zeta}_{1\text{numZF}}}{\tilde{\zeta}_{1\text{denZF}}} \quad (\text{B.11})$$

where the numerator is

$$\begin{aligned} \tilde{\zeta}_{1\text{numZF}} = & \beta_2 + \eta\beta_1\beta_2 + B\rho\beta_2^2 - B\rho\beta_2^2\tilde{\zeta}_1 + B\rho\beta_1\beta_2\tilde{\zeta}_1 \\ & + B^2\rho^2\beta_1\beta_2^2\tilde{\zeta}_1 + B\eta\rho\beta_1\beta_2^2 - B^2\rho^2\beta_1\beta_2^2\tilde{\zeta}_1^2 - B\eta\rho\beta_1\beta_2^2\tilde{\zeta}_1 \end{aligned} \quad (\text{B.12})$$

and the denominator is

$$\begin{aligned} \tilde{\zeta}_{1\text{denZF}} = & -B^2\rho^2\beta_1^2\beta_2\tilde{\zeta}_1^2 + B^2\rho^2\beta_1^2\beta_2\tilde{\zeta}_1 - B^2\rho^2\beta_1\beta_2^2\tilde{\zeta}_1^2 + B^2\rho^2\beta_1\beta_2^2\tilde{\zeta}_1 + \eta B\rho\beta_1^2\beta_2\tilde{\zeta}_1 + B\rho\beta_1^2\tilde{\zeta}_1 \\ & - \eta B\rho\beta_1\beta_2^2\tilde{\zeta}_1 + \eta B\rho\beta_1\beta_2^2 + B\rho\beta_1\beta_2 - B\rho\beta_2^2\tilde{\zeta}_1 + B\rho\beta_2^2 + 2\eta\beta_1\beta_2 + \beta_1 + \beta_2 \end{aligned} \quad (\text{B.13})$$

Taking (B.6) and substituting into (B.10) then differentiating with respect to $\tilde{\zeta}_1$ and setting the derivative to zero to find the value of $\tilde{\zeta}_{1\text{optZF}}$, the value of $\tilde{\zeta}_1$ which maximise SINR gives

$$\tilde{\zeta}_{1\text{optZF}} = \frac{\beta_2^2 - \beta_1\beta_2\sqrt{(\rho\beta_1 + 1)\rho\beta_2 + 1} + \rho\beta_1\beta_2^2}{\rho\beta_1^2\beta_2 + \beta_1^2 - \rho\beta_1\beta_2^2 - \beta_2^2}, \beta_1 \neq \beta_2 \quad (\text{B.14})$$

B.3 Conclusion

Note that

$$\zeta_{1\text{optZF}} = \zeta_{1\text{optMRT}} = \zeta_{1\text{opt}} = \frac{\beta_2^2 - \beta_1\beta_2\sqrt{(\rho\beta_1 + 1)\rho\beta_2 + 1} + \rho\beta_1\beta_2^2}{\rho\beta_1^2\beta_2 + \beta_1^2 - \rho\beta_1\beta_2^2 - \beta_2^2}, \beta_1 \neq \beta_2 \quad (\text{B.15})$$

This is not unexpected given the same method of channel approximation is used for both MRT and ZF. Also note that while $\zeta_{1\text{opt}}$ is undefined with direct substitution of $\beta_1 = \beta_2$ into (B.15), $\lim_{\beta_2 \rightarrow \beta_1} \zeta_{1\text{opt}} = 0.5$. This is as expected as to optimise $\min(\text{SINR}_1, \text{SINR}_2)$ with equal β_k , equal power distribution would be expected i.e., $\zeta_1 = \zeta_2 = 0.5$.

Once $\zeta_{1\text{opt}}$ and $\zeta_{2\text{opt}} = 1 - \zeta_{1\text{opt}}$ are found to optimise they can be substituted into (B.1) to give the desired values of δ_k to optimise the maximum minimum SINR for a fixed pilot uplink power. Having found the desired δ_k , they can be substituted into (3.45) for give $\zeta_{1\text{optMRT}}$ or substituted into (3.47) to give $\zeta_{1\text{optZF}}$. The values of $\zeta_{2\text{MRT}}$ or $\zeta_{2\text{ZF}}$ can then be found by (B.5).

Bibliography

- [1] E. Björnson, E. G. Larsson, and M. Debbah. “Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?” In: *IEEE Transactions on Communications* 15.2 (Feb. 2016), pp. 1293–1308.
- [2] E. Björnson, E. G. Larsson, and T. L. Marzetta. “Massive MIMO: Ten myths and one critical question”. In: *IEEE Communications Magazine* 54.2 (Feb. 2016), pp. 114–123.
- [3] E. Björnson, M. Matthaiou, and M. Debbah. “Massive MIMO with non-ideal arbitrary arrays: Hardware scaling laws and circuit-aware design”. In: *IEEE Communications Magazine* 54.8 (Aug. 2015), pp. 4353–4368.
- [4] E. Björnson et al. “Hardware impairments in large-scale MISO systems: Energy efficiency, estimation, and capacity limits”. In: *2013 18th International Conference on Digital Signal Processing (DSP)* (July 2013), pp. 1–6.
- [5] E. Björnson et al. “Massive MIMO is a reality - what is next? Five promising research directions for antenna array”. In: *Digital Signal Processing* 94 (June 2019), pp. 3–20.
- [6] E. Björnson et al. “Massive MIMO Systems with Non-Ideal Hardware: Energy Efficiency, Estimation, and Capacity Limit”. In: *IEEE Transactions on Information Theory* 60.11 (Nov. 2014), pp. 7112–7139.
- [7] H. V. Cheng, E. Björnson, and E. G. Larsson. “Uplink pilot and data power control for single cell massive MIMO systems with MRC”. In: *International Symposium on Wireless Communication Systems* (2015), pp. 396–400.
- [8] H. V. Cheng, D. Persson, and E. G. Larsson. “Optimal MIMO precoding under a constraint on the amplifier power consumption”. In: *IEEE Transactions on Communications* 67.1 (Jan. 2019), pp. 218–229.

-
- [9] Y. Dong, Y. Huang, and L. Qiu. "Energy-Efficient sparse beamforming for multiuser MIMO Systems with nonideal power amplifiers". In: *IEEE Transactions on Vehicular Technology* 66 (), pp. 134–145.
- [10] M. Fozooni et al. "Performance limits of MIMO systems with nonlinear power amplifiers". In: *2015 IEEE Global Communications Conference (GLOBECOM)* (Dec. 2015), pp. 1–7.
- [11] X. Gao et al. "Linear pre-coding performance in measured very-large MIMO channels". In: *2011 IEEE Vehicular Technology Conference (VTC Fall)* (2011), pp. 1–5.
- [12] S. Gunnarsson et al. "Channel hardening in massive MIMO-A measurement based analysis". In: *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)* (June 2018), pp. 1–5.
- [13] U. Gustavsson et al. "On the impact of hardware impairments on massive MIMO". In: *2014 IEEE Globecom Workshops* (Dec. 2014), pp. 294–300.
- [14] S. Haykin. *Communication Systems*. John Wiley & Sons, Inc, 2001.
- [15] R. W. Heath Jr. et al. "An overview of signal processing techniques for millimeter wave MIMO systems". In: *IEEE Journal on Selected Topics in Signal Processing* 10.3 (Apr. 2016), pp. 436–453.
- [16] E. Larsson et al. "Massive MIMO for next generation wireless systems". In: *IEEE Communications Magazine* 52.2 (Feb. 2014), pp. 186–195.
- [17] S. Li et al. "Massive MIMO for Ray-Based Channels". In: *2019 IEEE International Conference on Communications (ICC)* (May 2019), pp. 1–7.
- [18] T. L. Marzetta. "How much training is required for multiuser MIMO?" In: *2006 Fortieth Asilomar Conference on Signals, Systems and Computers* (Oct. 2006), pp. 359–363.
- [19] T. L. Marzetta et al. *Fundamentals of massive MIMO*. Cambridge University Press, 2016.
- [20] C. Mollén, E. G. Larsson, and T. Eriksson. "Waveforms for the massive MIMO downlink: Amplifier efficiency, distortion and performance". In: *IEEE Transactions on Communications* 64.12 (Dec. 2016), pp. 5050–5063.

-
- [21] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta. "Energy and spectral efficiency of very large multiuser MIMO systems". In: *IEEE Transactions on Communications* 61.4 (Apr. 2013), pp. 1436–1449.
- [22] H. Ochiai. "An analysis of band-limited communication systems from amplifier efficiency and distortion perspective". In: *IEEE Journal on Selected Areas in Communications* 61.4 (Apr. 2013), pp. 1460–1472.
- [23] J. C. Pedro and S. A. Maas. "A comparative overview of microwave and wireless power-amplifier behavioral modeling approaches". In: *IEEE Transactions on Microwave Theory and Techniques* 53.4 (Apr. 2005), pp. 1150–1163.
- [24] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst. "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization". In: *IEEE Transactions on Communications* 53.1 (Jan. 2005), pp. 195–202.
- [25] J. Qi and S. Aissa. "Analysis and compensation of power amplifier nonlinearity in MIMO transmit diversity systems". In: *IEEE Transactions on Vehicular Technology* 59.6 (July 2010), pp. 2921–2931.
- [26] J. Qi and S. Aissa. "On the Power Amplifier Nonlinearity in MIMO Transmit Beamforming Systems". In: *IEEE Transactions on Communications* 60.3 (Mar. 2012), pp. 876–887.
- [27] F. Rusek et al. "Scaling up MIMO: Opportunities and challenges with very large arrays". In: *IEEE Signal Processing Magazine* 30.1 (Jan. 2013), pp. 40–60.
- [28] S. Sun et al. "MIMO for millimeter-wave wireless communications: beamforming, spatial multiplexing, or both?" In: *IEEE Communications Magazine* 52.12 (Dec. 2014), pp. 110–121.
- [29] H. Tataria et al. "Uplink analysis of large MU-MIMO systems with space-constrained arrays in Ricean fading". In: *2017 IEEE International Conference on Communications (ICC)* (May 2017), pp. 1–7.
- [30] T. Van Chien, E. Björnson, and E. G. Larsson. "Joint Power Allocation and User Association Optimization for Massive MIMO Systems". In: *IEEE Transactions on Wireless Communications* 52.9 (Sept. 2016), pp. 6384–6399.

-
- [31] D. Verenzuela, E. Björnson, and M. Matthaiou. "Hardware design and optimal ADC resolution for uplink massive MIMO systems". In: *2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)* (July 2016), pp. 1–5.
- [32] M. Vetterli, J. Kovačević, and V. K. Goyal. *Foundations of Signal Processing*. Cambridge University Press, 2014.
- [33] H. Yang and T. L. Marzetta. "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems". In: *IEEE Journal on Selected Areas in Communications* 13.2 (Feb. 2013), pp. 172–179.
- [34] X. Zhang et al. "On the MIMO capacity with residual transceiver hardware impairments". In: *2014 IEEE International Conference on Communications (ICC)* (June 2014), pp. 5299–5305.
- [35] L. Zhao et al. "Energy efficient power allocation algorithm for downlink massive MIMO with MRT precoding". In: *2013 IEEE 78th Vehicular Technology Conference (VTC Fall)* 52.9 (Sept. 2013), pp. 1–5.