Exploiting Spatial Degrees-of-Freedom for Energy-Efficient Next Generation Cellular Systems

Miao Yao

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Electrical Engineering

Jeffrey H. Reed, Co-Chair
Cameron D. Patterson, Co-chair
Yaling Yang
Harpreet S. Dhillon
Douglas R. Bish

March 31, 2017
Blacksburg, Virginia

Keywords: Green Communications, DAS, Massive MIMO, PAPR, DPD

Copyright 2017, Miao Yao

Exploiting Spatial Degrees-of-Freedom for Energy-Efficient Next Generation Cellular Systems

Miao Yao

(ABSTRACT)

This research addresses green communication issues, including energy efficiency, peak-to-average power ratio (PAPR) reduction and power amplifier (PA) linearization. Green communication is expected to be a primary goal in next generation cellular systems because it promises to reduce operating costs.

The first key issue is energy efficiency of distributed antenna systems (DASs). The power consumption of high power amplifiers (HPAs) used in wireless communication systems is determined by the transmit power and drain efficiency. For unequal power allocation of orthogonal frequency division multiplexing (OFDM), the drain efficiency of the PA is determined by the PAPR and hence by the power distribution. This research proposes a PAPR-aware energy-efficient resource allocation scheme for joint orthogonal frequency division multiple access (OFDMA)/space division multiple access (SDMA) downlink transmission from DASs. Grouping-based SDMA is applied to exploit the spatial diversity while avoiding performance degradation from correlated channels. The developed scheme considers the impact of both system data rate and effective power consumption on the PAPR during resource allocation. We also present a suboptimal joint subcarrier and power allocation algorithm to facilitate implementation of power-efficient multi-channel wireless communications. By solving Karush-Kuhn-Tucker conditions, a closed-form solution for the power allocation of each remote radio head is obtained.

The second key issue is related with PAPR reduction in the massive multiple-input multiple-output (MIMO) systems. The large number of PAs in next generation massive MIMO cellular communication system requires using inexpensive PAs at the base station to keep array cost reasonable. Large-scale multiuser (MU) MIMO systems can provide extra spatial degrees-of-freedom (DoFs) for PAPR reduction. This work applies both recurrent neural network (RNN)- and semidefinite relaxation (SDR)-based schemes for different purposes to reduce PAPR. The highly parallel structure of RNN is proposed in this work to address the issues of scalability and stringent requirements on computational times in PAPR-aware precoding problem. An SDR-based framework is proposed to reduce PAPR that accommodates channel uncertainties and intercell coordination. Both of the proposed structures reduce linearity requirements and enable the use of lower cost RF components for large-scale MU-MIMO-OFDM downlink.

The third key issue is digital predistortion (DPD) in the massive MIMO systems. The primary source of nonlinear distortion in wireless transmitters is the PA, which is commonly modeled using polynomials. Conventional DPD schemes use high-order polynomials to accurately approximate and compensate for the nonlinearity of the PA. This is impractical for scaling to tens or hundreds of PAs in massive MIMO systems. This work therefore proposes a scalable DPD method, achieved by exploiting massive DoFs of next generation front ends. We propose a novel indirect learning structure which adapts the channel and PA distortion iteratively by cascading adaptive zero-forcing precoding and DPD. Experimental results show that over 70% of computational complexity is saved for the proposed solution, it is shown that a 3rd order polynomial with the new solution achieves the same performance as the conventional DPD using 11th order polynomial for a 100×10 massive MIMO configuration.

Exploiting Spatial Degrees-of-Freedom for Energy-Efficient Next Generation Cellular Systems

Miao Yao

(GENERAL AUDIENCE ABSTRACT)

The global climate change has emerged as a critical issue over the last decades. The increasing popularity of wireless communication networks, has resulted in information and communication technology becoming a non-negligible contributor to the overall carbon footprint. The increasing number of base stations and remote radio heads leads to higher operating expenditure mainly because of the higher energy consumption. This growth can be attributed not only to the increase in the number of smart devices in emerging economies, but also to the growth of shared multimedia data and online games. The wireless industry needs significant improvements in the energy efficiency of base stations and other network infrastructure to compensate for the increased energy demands from the network growth. Therefore, designing energy-efficient communication systems has become a critical issue for 5G, which promises massive deployment of smart devices served new infrastructure elements.

In this dissertation, we primarily investigate the theoretical foundations and practical algorithms for the next generation wireless technologies, and discuss the impact of ongoing trends in cellular communications, such as shrinking cell sizes and multi-antenna system deployments, on energy-efficient 5G networks. The theoretical development and wireless algorithms are valuable for the deployment of next generation wireless network systems.

Dedication

To my family.

Acknowledgments

I am extremely grateful for being a Ph.D. student of Dr. Jeffrey Reed. I would like to express my sincere gratitude for Dr. Reed's supervision and guidance during the past five years. Dr. Reed offered me with the great opportunities of a variety of projects which have greatly expanded my professional expertise, read my papers carefully and provided me with insightful comments.

I sincerely thank my co-advisor Dr. Cameron Patterson for the insightful discussions and invaluable feedback on my hardware works. I also want to thank my Ph.D. committee members: Dr. Yaling Yang, Dr. Harpreet Dhillon, Dr. Douglas Bish and Dr. Diego Moran for their interests in my work, and the valuable comments I have been receiving from them.

I extend great gratitude to Dr. Vuk Marojevic, Dr. Taeyoung Yang and Randall Neally's guidance and help on the inspiration and insightful discussions on my academic research throughout my Ph.D study.

I sincerely thank all my colleagues in Wireless@Virginia Tech group, and all my friends at Blacksburg with whom I have experienced such a joyful period of my life. Last but not least, I would also like to thank my wife, my parents and my sister, their unconditional love and support has always helped me in my life.

Contents

List of Figures

List of Tables

Chapter 1

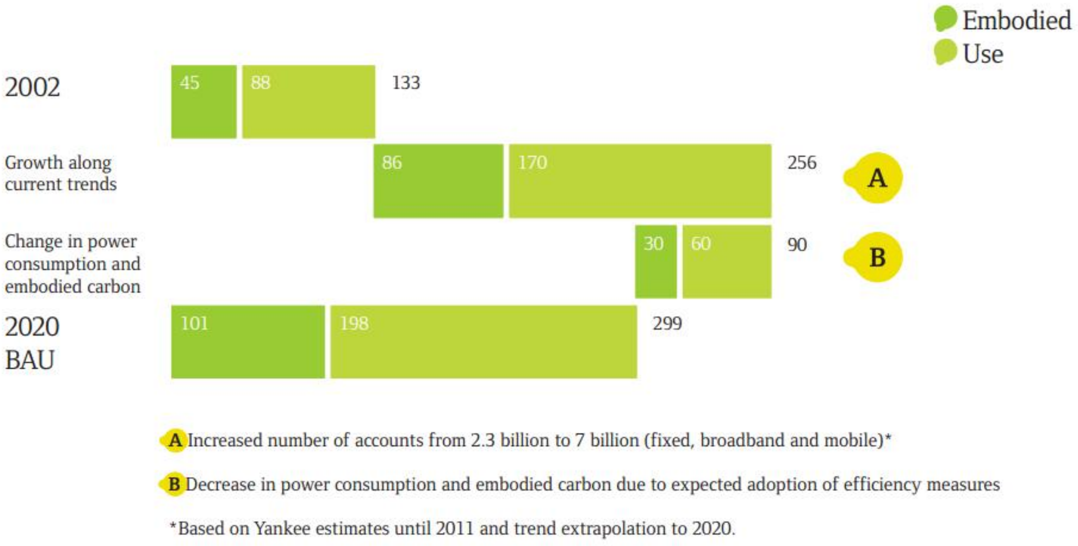
Introduction

1.1 Motivation

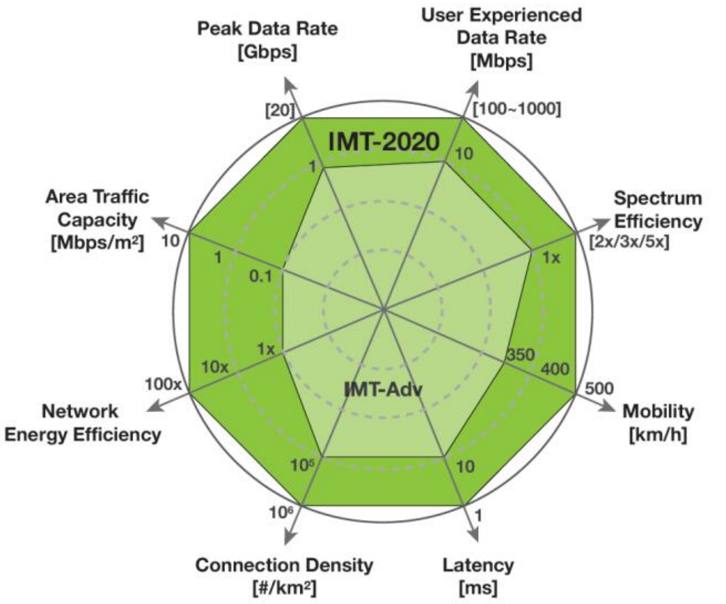
As global climate change has emerged as a critical issue over the last few decades, the problem of power consumption of wireless communication services has drawn enormous attention from both academia and industry. Increasing number of base stations (BSs) and remote radio heads (RRHs) leads to greater power consumption and hence higher operating expenditure (OPEX). According to a report proposed by Global eSustainbility Initiative (GeSI) [4]:

> ”The telecoms infrastructure footprint, including ongoing energy use and carbon embodied in the infrastructure, was 133 million metric tons of carbon dioxide equivalent (MtCO2e) in a business as usual (BAU) scenario in 2002. This is expected to more than double to 299 MtCO2e by 2020, a growth rate of 5% per year.”

The estimated carbon footprint of global telecommunication infrastructure with the estimation of 2020 is shown in Figure 1.1a. As the desire for wireless communication devices grows



(a) The global telecommunication infrastructure footprint [4].



(b) The ITU working party view of next generation cellular system [5]

Figure 1.1: The estimation and expectation of wireless communication industry by the year 2020

so, will the need for infrastructures that support these devices. This growth is not only due to the massive deployments of heterogeneous wireless networks in emerging economies, but also to the increase in sharing multimedia and online games. The wireless industry needs significant improvements in the energy efficiency of BSs and other network infrastructure equipment to compensate for the increased energy demands from overall network growth. Therefore, designing energy-efficient communication systems has become a critical issue especially in 5G new radio (NR). The International Mobile Telecommunications (IMT) view of next generation cellular system, shown in Figure 1.1b, indicates the expectation of 100× improvement in network energy efficiency by the year 2020. Concepts such as ”green communications” have recently emerged and hence sloganized to describe the trends of energy-efficient wireless communications.

Table 1.1: Base station power consumption at maximum load of a LTE system with 2×2 MIMO for different BS types [3]

			Macro	**RRH**	**Micro**	**Pico**	**Femto**
BS	Max Tx power	[dBm]	43.0	43.0	38.0	21.0	17.0
	Average P_{max}	[W]	20.0	20.0	6.3	0.13	0.05
	Feeder loss σ_{feed}	[dB]	3	0	0	0	0
PA	Back-off	[dB]	8.0	8.0	8.0	12.0	12.0
	Max PA out (peak)	[dBm]	54.0	51.0	46.0	33.0	29.0
	PA eff η_{PA}	[dB]	31.1	31.1	22.8	6.7	4.4
	Total PA	[dB]	**128.2**	**64.4**	**27.7**	**1.9**	**1.1**
RF	P_{TX}	[W]	6.8	6.8	3.4	0.4	0.2
	P_{RX}	[W]	6.1	6.1	3.1	0.4	0.3
	Total RF	[W]	**12.9**	**12.9**	**6.5**	**1.0**	**0.6**
BB	Radio (inner Rx/Tx)	[W]	10.8	10.8	9.1	1.2	1.0
	Turbo code (outer Rx/Tx)	[W]	8.8	8.8	8.1	1.4	1.2
	Processors η_{PA}	[W]	10.0	10.0	10.0	0.4	0.3
	Total BB	[W]	**29.6**	**29.6**	**27.3**	**3.0**	**2.5**
	Total per TRX chain	[W]	**225.0**	**125.8**	**72.3**	**7.3**	**5.2**
	Percentage of PA	[%]	**60.0**	**51.2**	**38.3**	**26.0**	**21.2**

As shown in Table 1.1, approximately 60% of the total power consumption of a BS is attributed to the power amplifier (PA) in a macro cell 2×2 MIMO LTE system. As a result, the PAs are the most critical components to consider when dealing with the energy efficiency of cellular networks. We especially focus on the peak-to-average power ratio (PAPR) of the signal and digital predistortion (DPD) of the PA in this research. The notations across this dissertation are given as follows: we denote vectors by boldface lowercase letters, e.g. $\mathbf{a}$, and matrices by boldface uppercase letters, e.g. $\mathbf{A}$, the ith component of a vector $\mathbf{a}$ is written as a_i. Given two matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{A} \succ \mathbf{B}$ ($\mathbf{A} \succeq \mathbf{B}$) means that $\mathbf{A} - \mathbf{B}$ is positive definite (semidefinite), $\mathbb{C}$ and $\mathbb{R}$ denote the complex and real sets, $\{\cdot\}^H$ the Hermitian transpose, $E\{\cdot\}$ the expectation, $\mathbf{I}_M$ the $M \times M$ identity matrix, $\|x\|_2^2$ the Euclidean norm, $\|x\|_\infty^2$ the infinity norm, $\mathrm{Tr}\{\cdot\}$ the trace of a matrix, $\Re\{\cdot\}$ the real part of a matrix, $\Im\{\cdot\}$ the image part of a matrix, $\mathrm{diag}\{\cdot\}$ the block diagonal matrix whose diagonal elements are matrices, $[x]^+ = \max(0, x)$ for $x \in \mathbb{R}$. Kronecker product is denoted by $\otimes$, element-wise multiplication is denoted by $\odot$.

1.2 Basics of PAPR issue in Wireless Communication System

Orthogonal frequency division multiplexing (OFDM) has been widely adopted as the air interface of various wireless communication systems such as 4G (LTE and its evolution) and IEEE 802.11. Its success is associated with many advantages including robustness against multipath fading, high spectral efficiency [6], and the support for orthogonal frequency-division multiple access (OFDMA). However, OFDM suffers from a high PAPR at the transmitter side as a result of the constructive addition of modulation symbols carried over narrowband subcarriers. The high peaks lead to signal excursions into nonlinear region of the PA and, thereby, to nonlinear signal distortion and spectral spreading [7]. High input back-off (IBO) is required at the PA to keep the peaks of an OFDM signal in the amplifier's linear region. In cellular networks, the high IBO requirement causes low energy efficiency at the BS and higher cost PAs, and increases the utility and cooling costs for the operator [8].

1.2.1 PAPR Definition

In order to accommodate the large variations in the instantaneous transmission power at each antenna, the PA built for OFDM signals must have a wide linear region. Since linear RF components are generally more costly and less power efficient compared with their nonlinear counterparts, practical implementations of OFDM usually employ sophisticated PAPR reduction schemes. The power efficiency of highly linear PAs are typically about 15%-25% [9]. Consider a complex baseband OFDM signal with N_c subcarriers, the PAPR of signal $\mathbf{x}_m$ at the mth antenna is defined as the ratio between the maximum power and the average power [7]:

$$\mathrm{PAPR}(\mathbf{x}_m) = \frac{\|\mathbf{x}_m\|_\infty^2}{\frac{1}{N_c}\|\mathbf{x}_m\|_2^2}. \tag{1.1}$$

Note that the PAPR of complex baseband signal $\mathbf{x}_m$ satisfies $1 \leq \mathrm{PAPR}(\mathbf{x}_m) \leq N_c$ due to the standard vector norm relationships.

1.2.2 Spatial and Frequency Resources and Degrees-of-Freedom

The interference can be interpreted as intersection of resources in the frequency, time and space domains, as shown in Figure 1.2. To recover the desired signal at the receiver, the radio resource management (RRM) has to distinguish the interference in at least one domain.

The energy efficiency of MIMO-OFDM systems benefits not only from the multiuser (MU)

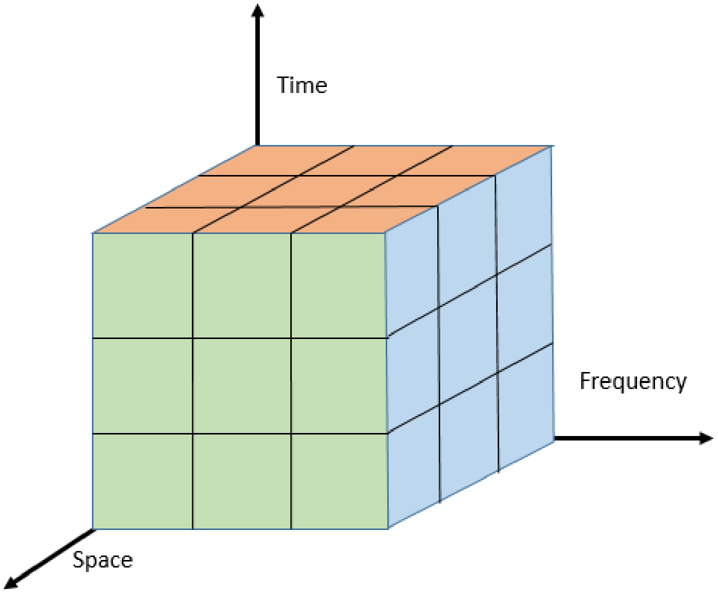


Figure 1.2: The RRM has to distinguish at least one of the three fundamental domains of time, frequency and space of the signal.

diversity, but also from the combined space, time and frequency domain flexibility for RRM. MIMO-OFDM networks offer more flexibility and higher granularity for RRM than single carrier/single antenna networks. The channel variations in the time, frequency and space domains provide inherent degrees-of-freedom (DoFs) for adaptive transmission. In the space domain, allowing intra-cell bandwidth reuse by means of spatial multiplexing is an advantage of multi-antenna systems. It can greatly enhance the spectral efficiency by sharing frequency and time resources among spatially separable users. This technique can be leveraged for significantly increasing the energy efficiency. The fundamental principle is that with a larger number of antennas, aggressive spatial RRM can focus on the instantaneous waveforms, which are collectively emitted by all antennas to reduce the input back-off margin of PAs or DPD complexity. For example, the extra DoFs can be used for signal shaping in 100×10 massive MIMO configuration to increase the PA efficiency as well as eliminate the in-/out-band distortions and spectral regrowth. The available DoF for realizing energy-efficient spatial RRM can be quantized by the precoding matrix which aims to suppress the MU interference (MUI). Our approach uses these extra DoFs either to leverage spatial resource allocation (instantaneous precoding) by jointly considering MUI and PA efficiency or to implement a scalable DPD for massive MIMO systems.

In the frequency domain, traditional RRM allocates different power levels to different subcarriers as a function of the channel characteristics to improve the overall MU capacity. This is called MU water filling. Energy efficiency metrics have been proposed to evaluate the performance of the wireless network RRM in the frequency domain [10]. The energy efficiency at the BS reflects the achievable data rate scaled by the system power consumption and can be represented as

$$\epsilon_{EE} = \frac{R(\boldsymbol{p})}{\sum_m \frac{1}{\eta_m} \sum_n p_{m,n} + P_r}, \tag{1.2}$$

where $R(\boldsymbol{p})$ represents the sum data rate and $p_{m,n}$ the power allocation matrix. The first part in the denominator represents the power consumption of the PAs, where η_m captures the PA efficiency of the PA at the mth antenna. The power consumption of the remaining part of the transmission system P_r in (1.2) is associated with the power supply, analog and digital signal processing, and air conditioning, among others. According to our measurements shown in Figure 1.3, the PA efficiency increases with the transmitted power. The PA efficiency improves as we drive the PA close to its saturation point. Beyond that, nonlinear effects occur and degrade the performance.

1.2.3 PAPR-aware Wireless Communication Systems Exploiting DoFs

Distributed antenna system (DAS) is considered as a promising technology for future wireless systems [1, 11, 12]. In a traditional DAS architecture, the RRHs are geographically

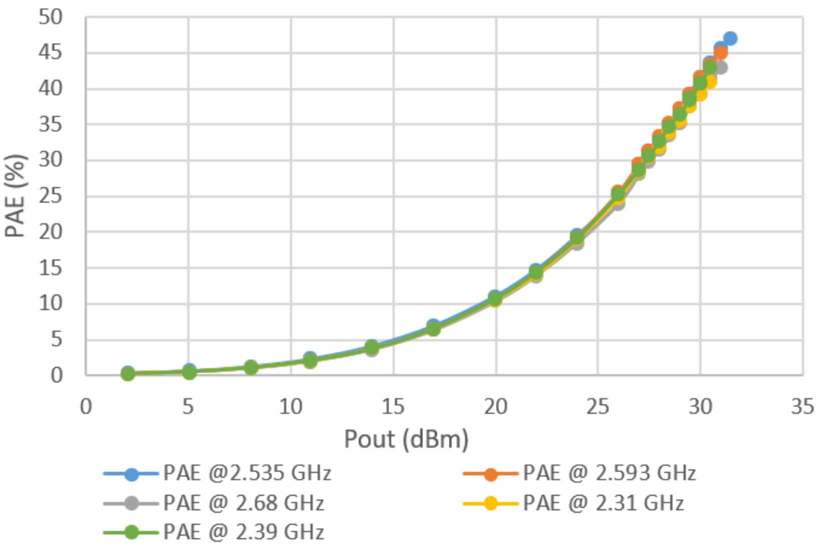


Figure 1.3: PA efficiency (PAE) over Pout for commercial LTE power amplifiers obtained from our real measurements for 20 MHz LTE system bandwidth.

distributed and connected to the central unit of the system via optical fiber or cable [11]. This research studies the effect of joint channel capacity and PAPR optimization in OFDMA based DAS. The multi-antenna subcarrier allocation provides numerous DoFs for jointly optimizing channel capacity and PAPR at each RRH. For this reason, the technique proposed in this research provides capacity improvement using DAS over a single antenna OFDMA system.

Massive MIMO systems have the potential to reduce the instantaneous PAPR by operating on the precoding matrix during the process of suppressing multiuser interference (MUI) [2]. The main idea behind the downlink transmission scheme in this research is to jointly perform MU precoding and PAPR reduction with either a recurrent neural network (RNN) or semidefinite relaxation (SDR) framework, by exploiting the numerous DoFs available in massive MIMO systems.

1.3 Basics of Semidefinite Relaxation

The SDR is a powerful and computationally efficient approximation technique for a variety of difficult optimization problems. In particular, it can be applied to many nonconvex quadratically constrained quadratic programs (QCQPs) as shown below

$$\min_{\mathbf{x}} \quad \mathbf{x}^H \mathbf{C} \mathbf{x} \tag{1.3}$$

$$\text{subject to} \quad \mathbf{x}^H \mathbf{A}_i \mathbf{x} \geq b_i \tag{1.4}$$

A necessary step in deriving an SDR of this problem is to observe that

$$\mathbf{x}^H \mathbf{C} \mathbf{x} = \mathrm{Tr}(\mathbf{C} \mathbf{x} \mathbf{x}^H) \tag{1.5}$$

$$\mathbf{x}^H \mathbf{A}_i \mathbf{x} = \mathrm{Tr}(\mathbf{A}_i \mathbf{x} \mathbf{x}^H) \tag{1.6}$$

By introducing a new variable $\mathbf{X} = \mathbf{x}\mathbf{x}^H$, we have the following problem

$$\min_{\mathbf{X}} \quad \mathrm{Tr}(\mathbf{C}\mathbf{X}) \tag{1.7}$$

$$\text{subject to} \quad \mathrm{Tr}(\mathbf{A}_i \mathbf{X}) \geq b_i \tag{1.8}$$

$$\mathbf{X} \succ 0 \tag{1.9}$$

$$\mathrm{rank}(\mathbf{X}) = 1 \tag{1.10}$$

The rank constraint $\mathrm{rank}(\mathbf{X}) = 1$ is nonconvex. By dropping the rank constraint, we can obtain the relaxed version of the QCQP. The relaxed problem is convex and can be efficiently solved using the interior-point optimization technique.

1.4 Basics of Digital Predistortion

1.4.1 Memoryless Digital Predistortion

The memoryless PA can be described as a nonlinear function of the input and output signal in the baseband. This memoryless nonlinearity can be described by a polynomial series

$$z(n) = \sum_{k=1}^{K} b_k y(n)|y(n)|^{2k} \tag{1.11}$$

where $y(n)$ is the PA input, and $z(n)$ the PA output.

1.4.2 Digital Predistortion with Memory Effect

The PA power response is time-variant when dealing with wideband signal inputs such as 10 or 20 MHz LTE. Therefore, the memory-based Volterra model is widely applied to represent the behavior of PA in wideband scenarios. Literature shows that a wide range of PA nonlinearities can be approximated with considerable precision by the Volterra series-based filter of sufficient order and memory depth [13]. The memory structure is necessary for the digital predistorter to deal with long-term memory introduced by the RF circuity.

The Hammerstein model is a memoryless nonlinearity followed by a linear time-invariant system [14]. It is described by

$$v(n) = \sum_{k=1}^{K} b_k y(n)|y(n)|^{2k} \tag{1.12}$$

$$z(n) = \sum_{l=0}^{L-1} c_l v(n-1) \tag{1.13}$$

where b_k are the coefficients for the memoryless nonlinearity and c_l are the impulse response values of the linear time-invariant system. Substitution of (1.12) into (1.13) leads to

$$z(n) = \sum_{k=1}^{K} c_l \sum_{l=0}^{L-1} b_k y(n-1)|y(n-1)|^{2k} \tag{1.14}$$

The memory polynomial model uses the diagonal kernels of the Volterra series and can be viewed as a generalization of the Hammerstein model as shown in indirect learning architecture of DPD in Figure 1.4.

$$z(n) = \sum_{k=1}^{K} \sum_{l=0}^{L-1} \omega_{kl} y(n-1)|y(n-1)|^{2k} \tag{1.15}$$

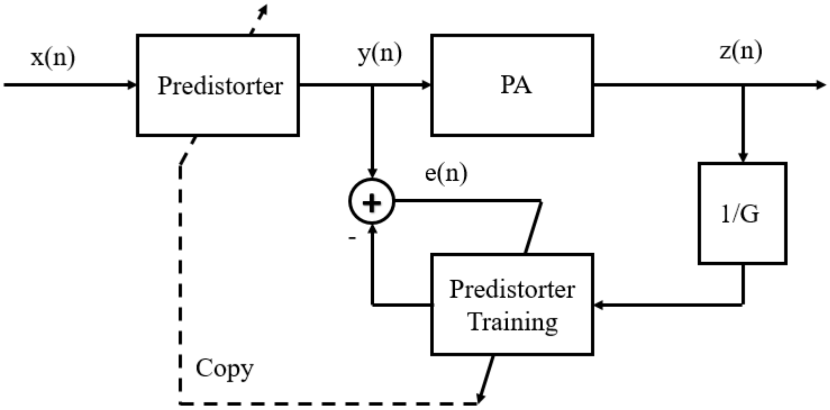


Figure 1.4: The indirect learning architecture for the PA predistorter.

1.5 Related Works

This research impacts the energy efficiency of wireless communication systems with excess spatial DoFs. This section reviews related works that form the foundation of our proposed approaches to address the PAPR reduction and DPD problems.

A variety of research has been done to address the PAPR issue of OFDM. The proposed solutions sacrifice signal distortion or redundancy to improve efficiency [15–21]. One of the simplest and most extensively used schemes is clipping and filtering [15, 16] which limits the PAPR below a threshold level, but causes both subcarrier inter-modulation and out-of-band radiation. The coding scheme proposed in [17] transforms the subcarriers into codewords with low PAPR, but reduces spectrum efficiency, especially when the number of subcarriers is large. Tone reservation [18] and constellation error shaping [19] schemes avoid sending data on a small subset of subcarriers and extends outer constellation points to minimize the PAPR of the OFDM symbols; however, this approach degrades spectrum efficiency. The selective mapping (SLM) scheme [22] is realized by multiplying the input symbol sequence to select alternative input sequences; but this requires side information to recover the signal at the receiver. The partial transmit sequence (PTS) scheme [20, 21] partitions the input symbol sequence into a variety of disjoint symbol subsequences and also needs additional side information. Unlike to the existing PAPR reduction approaches, the proposed PAPR-aware approaches do not sacrifice either the transmission rate or the spectral purity.

Multiuser resource allocation for OFDMA systems has drawn the attention of many researchers. The authors of [23–26] introduce algorithms that allocate power to each subcarrier as a function of the channel state information (CSI) to provide optimal resource allocation. Growing economic pressures and environmental awareness have shifted the attention of designers towards energy efficiency. The fundamental tradeoffs between spectral efficiency and energy efficiency in resource allocation are discussed in [27, 28]. A closed-form power allocation solution is derived for optimal energy efficiency of a point-to-point single carrier system with a minimum rate constraint in [29]. Both uplink and downlink energy efficiency in OFDMA networks are considered with QoS constraint in [10]. In [30–32], the system energy efficiency for OFDM-based cognitive radio systems is maximized by considering the transmission power budget of the system, the interference threshold of the primary users and the traffic demands of the secondary users. It has been proven that a unique global maximum of energy efficiency exists and that it can be reached by proper radio resource management [33, 34].

On the other hand, there has been recent research to enhance the system energy efficiency by using multiple collocated or distributed antennas [1, 12, 35]. He et al. [1] proposes a suboptimal resource allocation algorithm to maximize the energy efficiency while considering the proportional fairness for different mobile stations (MSs) and discusses the tradeoff between energy efficiency and spectral efficiency for DAS [12]. Ng et al. [35] propose an energy-efficient resource allocation algorithm with a large number of collocated antennas. However, these methods neglect the effect of power allocation on the PAPR [36] but, rather, assume constant and equally efficient PAs. In other words, the interrelation among PAPR, drain efficiency and subcarrier resource allocation is neglected. The PAPR-aware resource allocation algorithm for energy-efficient DASs which proposed in this work is based on a newly derived closed-form theoretical PAPR model to optimize the energy efficiency.

Unlike the PAPR-aware energy efficiency optimization approach in DAS, the massive MIMO systems have additional potential to reduce the instant PAPR by operating on the precoding method during the process of suppressing MUI [2]. Therefore, the main idea of the downlink transmission scheme in this work is to jointly perform MUI suppression and PAPR reduction with RNNs or SDR, by exploiting the extra DoF in massive MIMO systems. Artificial neural networks (ANNs) have been applied in PAPR reduction during the past few years. An active constellation expansion (ACE) PAPR reduction scheme based on multilayer perceptron (MLP) neural network is proposed in [37] to reduce the envelope fluctuation in OFDM systems. A similar neural network architecture was proposed in [38] where the MLP was implemented in the receiver side to reduce the transmitter complexity. The authors in [39] introduced a PAPR reduction scheme based on clipping and filtering and neural network mapper to reduce the iterative processing instead of traditional iterative clipping and filtering (ICF) method. Depending on the individual architecture, ANN-based models can be divided into three catalogs, namely: (1) MLP [40]; (2) radial basis function artificial neural network (RBFNN) [41]; and (3) RNN [42]. Among the different ANN types and architectures, the RNN is perhaps the most relevant. Tank and Hopfield [43] applied the RNN into linear programming problems in 1986, which inspired various research on constrained optimization solutions based on neural networks. The RNN exploits the similarity between precoding-PA architecture in MU-MIMO system and perceptron-activation function architecture in RNNs. A highly parallel and hardware-friendly structure is proposed.

A semidefinite relaxation (SDR)-based PAPR-aware precoding for massive MIMO-OFDM system is also proposed in this work. The optimization problem is formulated as a per OFDM symbol (instantaneous) transmit power minimization subject to PAPRs and MUI below certain thresholds. It provides a mechanism for trading MUI, PAPR, and transmit power. This problem is a nonconvex quadratically constrained quadratic programming (QCQP) problem. We employ the SDR and then the rank reduction method to obtain an estimate of the QCQP. We leverage the SDR approach in [44–52] and apply a randomization scheme [49–52] to achieve PAPR-aware precoding solution. As a powerful and computationally efficient optimization technique, SDR has recently been applied to solve a variety of nonconvex or NP-hard problems in communications. The beamforming problem with the minimization of the transmit power while keeping the interference generated by other coexisting systems under certain tolerable level is approximated by SDR in [44]. Maximum-likelihood detection in MIMO M-ary phase-shift keying (M-PSK) communication systems, which is an NP-hard least squares search problem, was also approximated by SDR [49]. The application of SDR in MIMO detection was later extended to 16-quadrature amplitude modulation (16-QAM) [45] and to higher-order QAM alphabets for general QAM constellations [46]. In addition, the SDR has been applied to PAPR reduction of the single antenna system in [47,48]. The basis of our solution is expanding the feasible solution sets which yields an attractive polynomial-time approximation. Improvements brought by the traditional PAPR reduction techniques trade effective transmission rate or spectral purity to improve power efficiency. This work applies a PAPR-aware precoding scheme which takes advantage of excessive DoFs in massive MIMO to reduce PAPR while preserving transmission rate and out-of-band emissions.

Robust precoding in the multiple antenna systems has been an intensive research topic over several decades because of the imprecise knowledge of the CSI degrades performance of precoding. Traditional precoding techniques do not provide sufficient robustness and thus various robust precoding techniques have been developed [53–58]. The eigenspace-based precoder is obtained by projecting the weight vector of the minimum-variance precoder onto the signal subspace of the measurement covariance matrix [53, 54]. The worst-case approaches assume that the CSI errors are norm bounded [55, 56]. The less conservative probability-based approach aims at providing robustness against CSI uncertainties with a certain preselected probability [57]. Related literature in single antenna and multiple antenna system are summarized in Table 1.2.

Table 1.2: Related literature in single antenna and multiple antenna system

Category	**Method**	**Descriptions**
PAPR reduction	Clipping and filtering [15, 16]	Limits the PAPR below a threshold level, but causes both the subcarrier inter-modulation and out-of-band radiation
	Coding [17]	Transforms the subcarriers into codewords with low PAPR
	Tone reservation [18]	Avoids to send data on a small subset of subcarriers
	Constellation error shaping [19]	Extends outer constellations to minimize the PAPR of the OFDM symbols
	Selective mapping [22]	Multiplies the input symbol sequence to select alternative input sequences which requires the side information
	Partial transmit sequence [20, 21]	Partitions the input symbol sequence into a variety of disjoint symbol subsequences which also needs additional side information
Resource allocation	Spectrum-efficient [23–26]	Allocate power to each subcarrier as a function of the CSI to provide optimal resource allocation to maximize spectrum efficiency
	Energy-efficient [10, 30–32]	System energy efficiency for OFDM-based system is maximized by considering the transmission power budget of the system

Continued on next page

Table 1.2 – *Continued from previous page*

Category	**Method**	**Descriptions**
Digital predistortion	Look-up table (LUT)-based [59, 60]	The difference between the ideal signal and the distorted output signal is saved in the LUT memory, low complexity and low precision
	Polynomial-based [13, 61]	The PA nonlinearities are approximated with polynomial series-based filter of sufficient order and memory depth, high computational complexity for high polynomial order
Semidefinite relaxation	MIMO detection [45, 46, 49]	The NP-hard least squares search problem was approximated by SDR to M-PSK system, and higher-order QAM alphabets
	Beamforming /precoding [44, 62]	Minimize the transmit power while takes the interference level into consideration is approximated by SDR
Robust precoding	Eigenspace based [53, 54]	Eigendecomposition is derived from signal plus interference subspace E
	Worse-case based [55, 56]	The channel is modeled as measured channel plus CSI uncertainty, δ is unknown and bounded as $\|\delta\| \leq \epsilon$
	Probability constraint based [57]	The channel is modeled as measured channel plus CSI uncertainty, δ obeys to Gaussian distribution
	Eigenvalue based [58]	Precoding vector lies in a known signal subspace and the rank of the signal correlation matrix is known

1.6 Dissertation Outline

1.6.1 Scope of Dissertation

This work focuses on developing multiple antenna energy-efficient green communication systems. While there is extensive research on multiple antenna systems, the current literature lacks the development of multiple antenna green networks by exploiting spatial DoFs. To achieve multiple antenna energy-efficient communication systems, we leverage the following observations:

- The multiple antenna systems have the potential to reduce the PAPR by exploiting the spatial DoFs;
- It is critical to extend the operation of the PA into weakly nonlinear region of the low-cost PAs in multiple antenna systems by exploiting the spatial DoFs and hence enable low-complexity DPD.

This work therefore proposes energy-efficient communication systems by applying DoF-aware PAPR reduction and DPD.

1.6.2 Dissertation Organization

This research leverages the excess DoFs of multiple antenna systems to achieve green communications. Chapter 2 presents the PAPR-aware resource allocation algorithm for energy-efficient OFDMA/SDMA DASs based on a newly derived closed-form theoretical PAPR model to optimize the energy efficiency. The multiple antenna subcarrier allocation technique exploits excess DoFs for jointly optimizing channel capacity and PAPR at each RRH.

For this reason, the technique that is proposed in this work is shown to provide greater capacity for DAS compared to single antenna OFDMA systems.

Chapter 3 presents a PAPR-aware downlink massive MIMO-OFDM system by exploiting the RNN structure. The analogy between a neural network and massive MIMO system is adopted to inspire a PAPR reduction technique that lends itself to a practical parallel implementation. The stability and convergence analysis of the proposed RNN is also derived. At the end of the chapter, we evaluate the proposed approach through simulation.

Chapter 4 presents a PAPR-aware downlink massive MIMO-OFDM system by exploiting the SDR structure. An optimization framework is formulated to minimize instantaneous transmission power with the assumption of perfect CSI to achieve a predefined PAPR and MUI at the transmitter and receiver, respectively. Both the bounded and statistical CSI errors are incorporated for the robust optimization. The co-channel interference at the cell edge from neighboring cells is suppressed by the centralized control unit performing intercell coordination. Evaluation of the proposed algorithm is provided at the end of this chapter.

Chapter 5 presents a scalable DPD structure to exploit the DoFs of a massive MIMO system. A novel indirect learning structure which learns both the characteristics of channel and PA distortion iteratively to realize a low complexity DPD. Both the simulation and experimental results are provided to validate the proposed scheme.

Chapter 6 concludes the dissertation by summarizing the works and achievements. Some future research issues are also suggested.

Chapter 2

PAPR-Aware Power Allocation for Energy-Efficient OFDMA/SDMA Distributed Antenna Systems

2.1 Introduction

In cellular networks, the high IBO requirement causes low energy efficiency at the BS and increases the utility and cooling costs for the operator [8]. It should be noted that approximately 60% of the total power consumption of a BS is attributed to the PA, 10% to analog and digital processing, and the remaining 30% to power supply conversion and air conditioning [63] which of course is there to help cool the PA. Therefore, PAs are the most critical components to consider when optimizing the energy efficiency of cellular networks.

In this chapter, we jointly consider the sum data rate and the PAPR characteristics with the objective of minimizing the overall energy consumption of the transmitter. In order to address this joint optimization problem, we derive the complementary cumulative distribution function (CCDF) for OFDM waveforms with unequal power allocation (UPA). Our result generalizes the closed-form solution for the PAPR distribution of OFDM that has been developed for equal power allocation [64]. The closed-form expression for the PAPR distribution of an OFDM allows analyzing the interdependency between the PA efficiency and resource allocation. We apply dual fractional programming to transform the original NP hard problem into a problem that can be solved iteratively with convex optimization solvers. Next, the convex optimization of power allocation in DAS is transformed to an iterative algorithm to avoid the quadratic computational complexity with respect to the number of RRHs to facilitate an embedded implementation.

The application of this solution reaches beyond pure OFDMA communication systems. By combining OFDMA with space division multiple access (SDMA) the system can exploit the rich spatial diversity of DAS for further improving the spectral efficiency. The bandwidth of SDMA is reused by letting multiple users which are spatially distinct share the same subcarriers. Various SDMA schemes have been proposed to increase the bandwidth efficiency based on zero forcing (ZF) or minimum mean square error (MMSE) [65] [66]. The authors in [67] and [68] use grouping method to prevent the performance loss due to spatially correlated users and exploit the multiuser diversity offered by SDMA. A joint OFDMA/SDMA method with greedy grouping is proposed in this work and its performance is compared with ZF and MMSE precoding.

2.2 System Model

In the context of a small cell, we assume that M distributed antennas/RRHs are connected to a processing center or cloud radio access network (C-RAN). As shown in Figure 2.1, K

users are served by the distributed antennas and the number of subcarriers for each antenna is N. The subcarriers are allocated on SDMA group basis, which means that each subcarrier is exclusive to one group. Furthermore, all the subcarriers allocated to different MSs are transmitted by each RRH and all the RRHs have the same information symbol group to be transmitted. Power is dynamically allocated to the same subcarrier of different RRH element according to the channel conditions. To jointly account for data rate and PAPR, we use energy efficiency as the objective function of our optimization framework. Energy efficiency has recently become an important consideration of the 3rd Generation Partnership

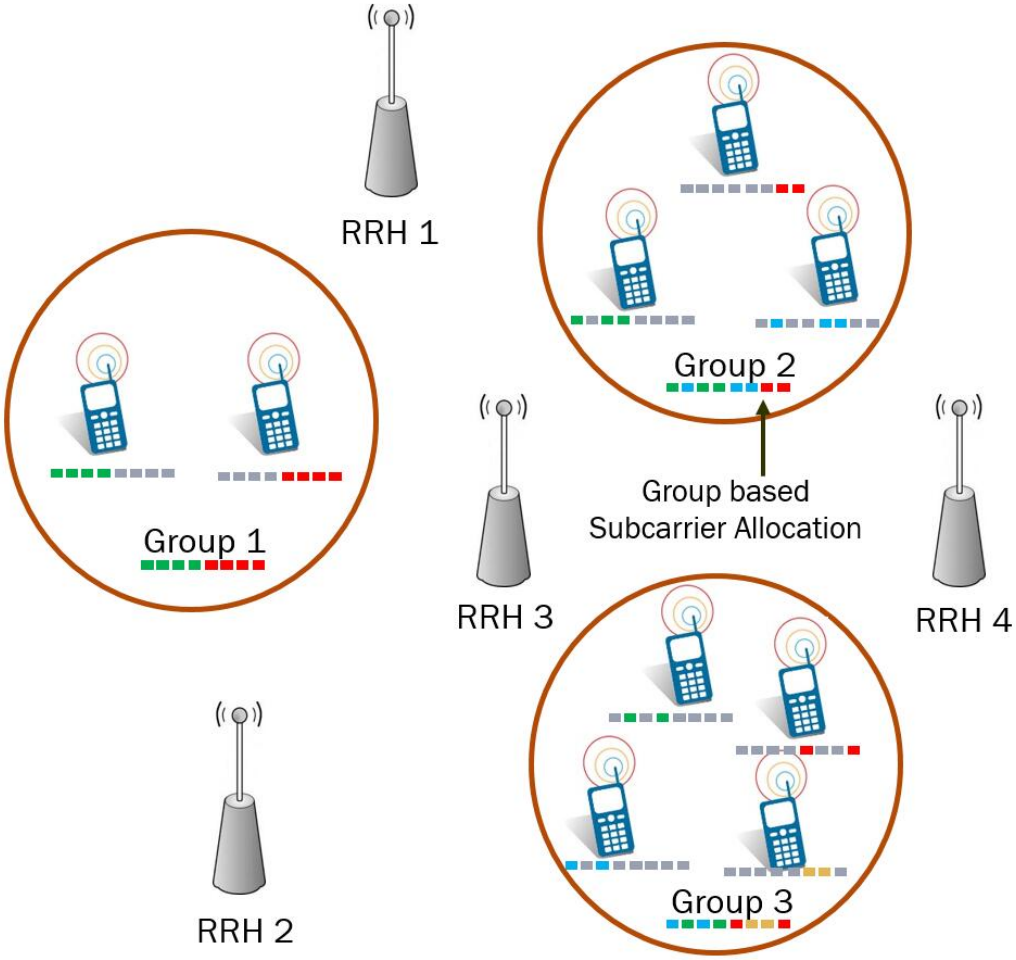


Figure 2.1: Structure of OFDMA/SDMA based DAS for small cell deployment with multiple distributed users.

Project (3GPP) [69]. The energy efficiency in Chapter 1 is rewritten as

$$\epsilon_{EE} = \frac{R(\boldsymbol{p})}{\sum_m \frac{1}{\eta_m} \sum_n p_{m,n} + P_r}, \tag{2.1}$$

where $R(\boldsymbol{p})$ represents the sum data rate and $\boldsymbol{p}$ the power allocation matrix, $\boldsymbol{p} = [p_{m,n}]_{M\times N}$. The first part of the denominator in (2.1) represents the power consumption of the PAs of DAS, where η_m captures the power efficiency of the HPA at the mth antenna. P_r encompasses the power consumption of the remaining part of the transmission system. This includes the power supply, baseband unit (BBU) for analog and digital signal processing and air conditioning.

Equation (2.1) is constraint by the power for each RRH $\sum_n p_{m,n} = P_{T,m}$, where $P_{T,m}$ is the transmission power limit for mth RRH. Normally, $P_{T,m}$ would be 46 and 23 dBm for Macro and Pico cell respectively in current LTE deployments [70]. It is impractical to assume that PA efficiencies are equal for all RRHs and independent of the power allocation scheme since the different RRHs may have their own transmit power limits. In contrast to the previous work, such as [1] [30], we assume that the PA efficiency depends on the PAPR and hence on the power allocation result.

The overall achievable data rate per unit bandwidth of the multi-user distributed antenna OFDMA system is simplified as [71]

$$R(\boldsymbol{p}) = \sum_{k=1}^{K} \sum_{n=1}^{N} \omega_{k,n} \log_2 \left(1 + \sum_{m=1}^{M} \frac{|h_{k,m,n}|^2 p_{m,n}}{\sigma_m^2} \right), \tag{2.2}$$

where $\omega_{k,n}$ is the channel assignment of the nth subcarrier to the kth user, either 0 or 1, $h_{k,m,n}$ denotes the channel state between the mth RRH and the kth user for the nth subcarrier, and σ_m^2 the noise spectral density at mth RRH. The next section presents the energy efficiency optimization problem which jointly considers the date rate $R(\boldsymbol{p})$ and the overall power consumption.

2.3 Energy Efficiency Optimization Problem

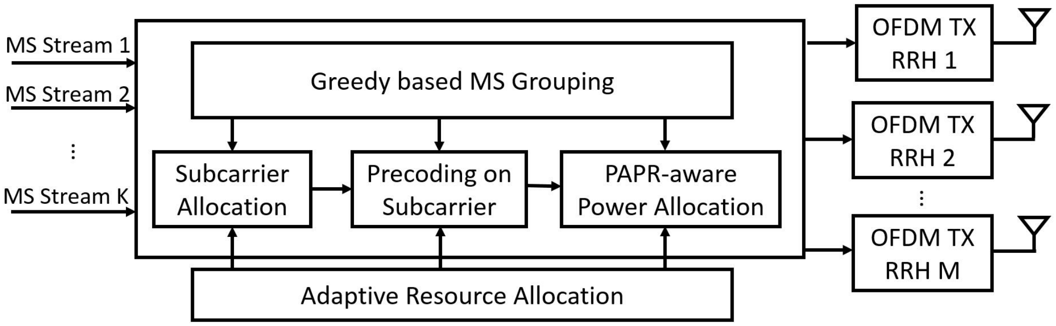


Figure 2.2: Structure of resource allocation for joint OFDMA/SDMA.

The energy efficiency ϵ_{EE} proposed in Section 2.2 is a function of both the subcarrier and power allocation. This presents an NP-hard optimization problem [71]. The overall optimization problem is thus divided into two steps to simplify the analysis and find a time-efficient solution. The first step deals with the joint OFDMA/SDMA, which allocates subcarriers to MSs while considering the spatial diversity. The second step allocates power to subcarriers of different RRHs. Suboptimal subcarrier allocation is carried out by considering only the sum capacity (without considering energy efficiency) in the first step. Based on the subcarrier allocation result, dual fractional programming is applied to solve the energy-efficient power allocation problem in the second step. Since channel state information at the transmitter (CSIT) is necessary for all the existing downlink resource allocation schemes, no additional effort for channel estimation is needed.

MSs can be multiplexed by employing the SDMA scheme, e.g., with a transmit ZF filter, while reusing the same resources in frequency and time. Therefore, a group-based joint OFDMA/SDMA multiple access scheme is adopted in this work and is illustrated in Figure

2.2. The MSs are divided into several SDMA groups, the MSs with spatially correlated channels are placed in the same group and multiplexed on different resources, e.g. on different subcarriers. The spectral efficiency can be improved by spatially reusing the uncorrelated channels while avoiding performance degradation introduced by the correlated channel through SDMA. The task of a grouping algorithm is to determine an efficient SDMA group with low complexity. The grouping metric in this work maps the characteristics of the spatial channels of the MSs to a scalar value that quantifies how efficiently these MSs can be separated in groups. The energy efficiency of joint SDMA/OFDMA can be written as

$$\epsilon_{EE}^{J} = \frac{\sum_{n=1}^{N} \sum_{i=1}^{|\mathcal{G}|} \sum_{k \in \mathcal{G}_i} \omega_{k,n} \log_2(1 + \sum_{m=1}^{M} \frac{|h_{k,m,n}|^2 p_{m,n}}{\sigma_m^2})}{\sum_m \frac{1}{\eta_m} \sum_n p_{m,n} + P_r}, \tag{2.3}$$

where $\mathcal{G}_i$ represents the set of MSs in ith SDMA group and $|\mathcal{G}|$ the total number of groups. The optimal solution for subcarrier and power allocation in terms of energy efficiency can be obtained by solving

$$\begin{aligned}
&\textbf{max} && \epsilon_{EE}^{J} && (2.4)\\
&\textbf{subject to} && C1: \quad \sum_{n} p_{m,n} = P_{T,m}, \forall m && (2.4a)\\
& && C2: \quad p_{m,n} \geq 0, \forall m, n, && (2.4b)\\
& && C3: \quad \sum_{k \in \mathcal{G}_i} \omega_{k,n} = 1, \forall n, i && (2.4c)\\
& && C4: \quad \sum_{k \in \mathcal{G}_i} \omega_{k,n} \in \{0, 1\}, \forall k, n, i && (2.4d)
\end{aligned}$$

2.3.1 Multiple Access Scheme

To determine the power allocation on subcarrier basis, it is important to first know the channel state information related to these subcarriers. However, the channel state information of individual subcarriers cannot be obtained without first carrying out subcarrier allocation for all MSs. A joint multiple access scheme is developed in this section. A greedy SDMA grouping algorithm is applied to overcome the deteriorated performance when the correlation between different MSs are high. The frequency resources are allowed to be shared between spatially separable MSs. Since different subcarriers are allocated to different MSs within the SDMA group, we have $\sum_{k \in \mathcal{G}_i} \omega_{k,n} = 1$ for all n, i and $\omega_{k,n} \in \{0, 1\}$ for all k and n.

Assume that a MS uses maximum ratio combining (MRC) to combine the signals it receives from the different RRHs, where the received signal are weighted to maximize the overall signal to noise ratio (SNR) [72]. Hence, the MRC channel for the kth MS and nth subcarrier becomes $H_{k,n} = \sum_m h_{k,m,n}^2 / \sigma_m^2$. In this step, the data rate R_k of the kth MS is calculated with the assumption of equal power distribution among subcarriers for each RRH. Without loss of generality, we assume that all MSs have equal fairness. Different to the exclusive subcarrier allocation algorithm presented in [1], a grouping based suboptimal subcarrier allocation algorithm which provides a better spatial diversity is given in Algorithm 1.

This algorithm, Algorithm 1, considers the channel between the MS and all RRHs and assigns a subcarrier to the MS that has the best combined channel condition in terms of SNR. An equivalent resource allocation criteria is used for the power allocation algorithm presented in Section 2.3.2.

Algorithm 1 Group Based Subcarrier Allocation (GBSA) for Joint OFDMA/SDMA Communication System

```
Input: channel state $\mathbf{h} = [h_{k,n,m}]_{K\times N\times M}$
Output: MS grouping result $\mathcal{G}_{\hat{i}}$ for $\hat{i} = 1, 2, \cdots, |\mathcal{G}|$
    OFDMA Subcarrier allocation result $\Omega_k^{(\hat{i})} = \{\omega_{k,n}\}$ for $k = 1, 2, ...K$
$\mathcal{G}_1 = \mathcal{G}_2 \cdots \mathcal{G}_{|\mathcal{G}|} = \emptyset$
$\hat{i} \leftarrow 1$
for $\forall k \in \mathcal{G}\backslash\{\mathcal{G}_1 \cup \mathcal{G}_2 \cdots \mathcal{G}_{|\mathcal{G}|}\}$ do
    $\mathcal{G}_{\hat{i}} \leftarrow \mathcal{G}_{\hat{i}} + \{k\}$
    if $\mathrm{SINR}(\mathcal{G}_{\hat{i}}) < \mathrm{SINR}_0$ then
        $\mathcal{G}_{\hat{i}} \leftarrow \mathcal{G}_{\hat{i}} - \{k\}$
        $\hat{i} \leftarrow \hat{i} + 1$
    end if
end for
for $i = 1$ to $|\mathcal{G}|$ do
    Initialization : set $R_k^{(i)} = 0$ and $\Omega_k^{(i)} = \emptyset$ for $k = 1, 2, ...K$, $H_{k,n} \leftarrow \sum_m h_{k,m,n}^2/\sigma^2$ and
    $A^{(i)} = 1, 2, ...N$
    for $k = 1$ to $K$ do
        find $n$ satisfying $|H_{k,n}| \geq |H_{k,j}|$, $j = 1, 2, ...N$ and $j \neq n$
        let $\Omega_k^{(i)} \leftarrow \Omega_k^{(i)} \cup \{\omega_{k,n} = 1\}$ and $A^{(i)} \leftarrow A^{(i)} - \{n\}$
        update $R_k^{(i)}$
    end for
    while $A^{(i)} \neq \emptyset$ do
        find $k$ satisfying $R_k^{(i)} \leq R_i^{(i)}$, $i = 1, 2, ...N$ and $i \neq k$
        find $|H_{k,n}| \geq |H_{k,l}|$ for all $l \in A$
        let $\Omega_k^{(i)} \leftarrow \Omega_k^{(i)} \cup \{\omega_{k,n} = 1\}$ and $A^{(i)} \leftarrow A^{(i)} - \{n\}$
    end while
end for
return $\Omega_k^{(i)}$
```

2.3.2 Subcarrier Power Allocation

The IBO and the PA efficiency are determined by the PAPR. As a result, the PA efficiency cannot be assumed to be constant as the PAPR is influenced by the power allocation result [73]. The authors in [1] [35] use fractional programing to solve the nonlinear energy efficiency optimization problem. Similar to the approach used in [1], we optimize the following objective function until it converges:

$$F(\alpha_i) = R(\boldsymbol{p}) - \alpha_i \left\{ \sum_m \frac{1}{\eta_m} \sum_n p_{m,n} + P_r \right\}, \tag{2.5}$$

where α_i is the energy efficiency in the ith iteration. Therefore, the ratio in the objective function (2.1) is transformed to a linear combination of power allocation and PA efficiency. However, according to our assumption, the efficiency of the power amplifier η_m is not constant and depends on the power allocation result. Thus, the objective function in (2.5) is still neither concave nor convex and it can be further transformed to the following equation according to the sum-of-ratios fractional programming [74]:

$$F(\alpha_i, \boldsymbol{\zeta}_i) = R(\boldsymbol{p}) - \alpha_i \left\{ \sum_m \left(\sum_n p_{m,n} - \zeta_{i,m}\eta_m \right) + P_r \right\}, \tag{2.6}$$

where $\zeta_{i,m}$ is defined as the power consumption of the mth RRH in the ith iteration, $\sum_n p_{m,n}/\eta_m$. Therefore, $\boldsymbol{\zeta}_i$ is a $1 \times M$ vector which represents the power consumption of each RRH for the ith iteration. Since there are two fractional programming instances involved in this optimization problem, the term "dual fractional programming" is coined in this work.

$$N_C(l) = \frac{\sqrt{2\pi}\Delta f}{N} \cdot \frac{\sqrt{\sum P_k \sum k^2 P_k - (\sum k P_k)^2}}{(\sum P_k)^{3/2}} l \exp\left\{ -\frac{l^2}{2\sum P_k} \right\}. \tag{2.7}$$

The algorithm flow of dual fractional programming is listed below:

Algorithm 2 Dual Fractional Programming Algorithm for Optimal Power Allocation

```
Output: power allocation result $\boldsymbol{p}$
    Initialization : $\alpha_0 = 0.01$ ,$\boldsymbol{\zeta}_0 = [0.05, 0.05, \cdots, 0.05]_{1\times M}$ and $\varphi = 10^{-6}$
while $F(\alpha_i, \boldsymbol{\zeta}_i) \geq \varphi$ do
    Find $\alpha_i$ and $\zeta_i$ with $\boldsymbol{p}$
    $\alpha_i \leftarrow \frac{R(\boldsymbol{p})}{\sum_m \sum_n p_{m,n}/\eta_m + P_r}$
    $\zeta_{i,m} \leftarrow (\sum_n p_{m,n})/\eta_m$
    Substitute $\alpha_i$ and $\zeta_i$ into equation (2.6)
    $F(\alpha_i, \boldsymbol{\zeta}_i) \leftarrow R(\boldsymbol{p}) - \alpha_i\{\sum_m \sum_n p_{m,n} - \zeta_{i,m}\eta_m) + P_r\}$
    find optimal $\boldsymbol{p}$ which maximize $F(\alpha_i, \boldsymbol{\zeta}_i)$
end while
return $\boldsymbol{p}$
```

In order to optimize the objective function (2.3), we need to find the optimal $\boldsymbol{p}$ for each iteration of Algorithm 2. Since the power efficiency η_m is a function of power allocation scheme, the closed-form expression of PAPR in terms of power distribution needs to be derived. The authors in [64] use the level crossing rate of the bandlimited Rayleigh process to approximate the peak distribution of the OFDM signal with equal power distribution. The level crossing rate of the OFDM signal with both equal and unequal power under more realistic assumptions (e.g., correlation between the real component and the derivative of the imaginary component) is derived in Appendix A.

According to (2.7), the distribution of PAPR in terms of the CCDF can be derived from the level crossing rate

$$\mathcal{F}(\gamma) = \begin{cases} \left(1 - \frac{\sqrt{\gamma} e^{-\frac{\gamma}{2\sum P_k}}}{\sqrt{\bar{\gamma}} e^{-\frac{\bar{\gamma}}{2\sum P_k}}}\right)^{\nu\sqrt{\bar{\gamma}} e^{-\frac{\bar{\gamma}}{2\sum P_k}}} , & \text{for } \gamma > \bar{\gamma} \\ 0, & \text{for } \gamma \leq \bar{\gamma}, \end{cases} \tag{2.8}$$

where the reference level $\bar{\gamma}$ is an empirical value which is usually obtained from simulation.

Parameter ν in (2.8) can be represented as $\sqrt{2\pi}\sqrt{\sum P_k \sum k^2 P_k - (\sum k P_k)^2}/\{N(\sum P_k)^{3/2}\}$, which we denominate the UPA factor in this work. By applying the first order Taylor expansion, the cumulative density function (CDF) of the PAPR can be approximated as

$$\mathcal{F}_C(\gamma) \approx \nu\sqrt{\gamma}e^{-\frac{\gamma}{2\sum P_k}}. \tag{2.9}$$

Thus, the CDF of the PAPR is a linear function of the UPA factor. To simplify the optimization problem, the PA efficiency is approximated by (see Appendix B)

$$\eta_m = c_1 + c_2\nu^2, \tag{2.10}$$

where c_1 and c_2 are constants. Equation (2.10) is applied to best fit the curve of power efficiency and PAPR. As a result, the optimization of equation (2.6) can be represented as

$$\begin{aligned} \textbf{max} \quad & R(\boldsymbol{p}) - \alpha_i \left\{ \sum_m \left(\sum_n p_{m,n} - \zeta_{i,m}\eta_m \right) + P_r \right\} \\ \textbf{subject to} \quad & \sum_n p_{m,n} = P_{T,m}, \forall m \\ & p_{m,n} \geq 0, \forall m, n, \end{aligned} \tag{2.11}$$

where $P_{T,m}$ represents the power constraints of the mth antenna. Since the equality in the summation of power allocation result is always satisfied, the power efficiency of the mth antenna can be represented as

$$\eta_m = c_1 + c_2 \frac{2\pi\Delta f^2 \left(P_{T,m}\sum n^2 p_{m,n} - (\sum n p_{m,n})^2\right)}{P_{T,m}^3}. \tag{2.12}$$

It is easy to prove that the objective function in (2.11) is quasi-concave, and the efficient convex programming routines, such as the interior point method, cannot be applied to solve it. However, this optimization problem can be converted to a convex optimization problem for which a closed-form solution exists. The power efficiency of the mth antenna can be rewritten as

$$\eta_m = c_1 + c_2 \frac{2\pi \Delta f^2 \left(P_{T,m} \sum n^2 p_{m,n} - (E_{p,m})^2\right)}{P_{T,m}^3} \tag{2.13}$$

with an additional linear constraint added to the optimization problem of (2.11):

$$\sum_n n p_{m,n} \leq E_{p,m}. \tag{2.14}$$

$E_{p,m}$ is obtained from $E_{p,m} = \sum_n n p'_{m,n}$ and $p'_{m,n}$ is derived from the previous power allocation. The problem has been transformed to a convex optimization problem so that a closed-form solution can be derived.

2.4 Iterative Power Allocation

We showed that the suboptimal problem formulation that we derived in Section 2.3 is a convex optimization problem. There are various real-time algorithms available for solving it [75]. However, the number of variables in the optimization problem increases with the number of RRHs squared, meaning that the original optimization problem will be computational intensive especially when considering massive antenna deployments. The energy efficiency problem reduces to a simple water filling problem if there is only one antenna element in the system. Keeping this in mind, we introduce an iterative power allocation algorithm where the power is allocated to only one antenna in each iteration while the powers of the remaining antennas are assumed constant [76]. During the optimization process, the power is allocated to a single antenna while the powers of the remaining antennas are assumed to be constant as obtained from previous iterations.

After getting rid of the constant part of the objective function, the original optimization in the mth iteration can be reduced to a new optimization problem

$$\begin{aligned}\textbf{max} \quad & \sum_{k=1}^{K}\sum_{n=1}^{N}\omega_{k,n}\log_2(C+\frac{|h_{k,m,n}|^2 p_{m,n}}{\sigma_m^2})+t\sum_{n} n^2 p_{m,n} \\ \textbf{subject to} \quad & \sum_{n} n p_{m,n} \le E_{p,m} \\ & \sum_{n} p_{m,n} = P_{T,m}, \forall m \\ & p_{m,n} \ge 0, \forall m,n\end{aligned}$$

where C and t in the objective function can be represented as

$$C = 1 + \sum_{\hat{m}=1,\hat{m}\neq m}^{M} \frac{|h_{k,\hat{m},n}|^2 p_{\hat{m},n}}{\sigma_{\hat{m}}^2} \tag{2.15}$$

and

$$t = \frac{c_2 \alpha_i \zeta_{i,m} 2\pi \Delta f^2}{P_{T,m}^2}. \tag{2.16}$$

The Lagrangian function which applies the Lagrangian multiplier $\boldsymbol{\lambda}$ can be represented as [77]

$$\begin{aligned}L(\boldsymbol{p}_m,\boldsymbol{\lambda}) = & \sum_{k=1}^{K}\sum_{n=1}^{N}\omega_{k,n}\log_2\left(C+\frac{|h_{k,m,n}|^2 p_{m,n}}{\sigma_m^2}\right) \\ & + t\sum_{n=1}^{N} n^2 p_{m,n} - \lambda_1\left(\sum_{n=1}^{N} n p_{m,n} - E_{p,m}\right) \\ & - \lambda_2\left(\sum_{n=1}^{N} p_{m,n} - P_{T,m}\right),\end{aligned} \tag{2.17}$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$. By differentiating the Lagrangian function with respect to $p_{m,n}$ and equaling it to zero

$$\frac{\partial L(\boldsymbol{p}_m,\boldsymbol{\lambda})}{\partial p_{m,n}} = 0, \tag{2.18}$$

we get the closed-form result for the power allocation of a single antenna

$$p_{m,n}^{opt} = \left[\frac{1}{\delta_{m,n}} - \frac{\sigma_m^2 C}{|h_{k,m,n}|^2} \right]^+, \tag{2.19}$$

where the operation $[x]^+$ is defined as $[x]^+ := \max(x, 0)$, and $\delta_{m,n}$ can be derived as

$$\delta_{m,n} = (\lambda_1 n + \lambda_2 - t n^2) \ln 2. \tag{2.20}$$

Although the PAPR is considered during the resource allocation, the computing complexity of our DAS power allocation solution increases linearly with the number of distributed antennas M and is thus suitable for massive deployments of RRHs. The proposed power allocation algorithm for a single antenna as part of the iterative power allocation process can thus run on embedded platforms or software-defined radios with limited computational resources.

The algorithm of iterative power allocation is shown as Algorithm 3, where step 3 can be calculated with equation (2.19)

Algorithm 3 Iterative Power Allocation Algorithm for Energy Efficiency Optimization in DAS

```
Input: subcarrier allocation result $\Omega_k$ for $k = 1, 2, ...K$
Output: power allocation result $\boldsymbol{p}$
    Initialization : $p_{m,n} = 0$ for $\forall$ $m$ and $n$
1: while the desired accuracy is not satisfied do
2:     for $m = 1$ to $M$ do
3:         $\{p_{m,n}\}_{n=1\cdots N} = \text{argmax}_p F(\alpha_i, \boldsymbol{\zeta}_i)$
4:     end for
5: end while
6: return $\boldsymbol{p}$
```

We still need to prove that our iterative power allocation algorithm converges. Since the power allocation of the rest of the RRHs in each iteration is taken from the previous iteration, the energy efficiency objective is non-decreasing during each iteration. Furthermore, the objective function does not go to infinity and an upper limit can always be found. According to the monotone convergence theorem, the objective inside each iteration of Algorithm 3 is non-decreasing and higher bounded which is, thus, convergent.

The proposed iterative algorithm presents a problem structure which is based on the single antenna power allocation while assuming the remaining parts to be constant. This allows taking advantage of the channel state eigenvalue decomposition used by the water filling approaches. The single antenna water filling problem with closed-form result, which is more efficient than the general purpose convex optimization algorithms, is derived to facilitate the convergence speed of Algorithm 3.

2.5 Numerical Results

Table 2.1: Simulation parameters of the PAPR-aware DAS system

Cell radius R	2 km
Reference distance	35 m
Carrier center frequency	3.55 GHz
Number of subcarriers	128
Subcarrier spacing	15kHz
Number of RRH	5-40
Number of MS	50
MS placement	Uniform
Modulation	QPSK
Small scale fading distribution	Rayleigh fading
Noise power per subcarrier	-131 dBm
Remaining part power consumption P_r	30 dBm [78]
Oversampling factor L	4
PA efficiency @ 0dB PAPR	0.35

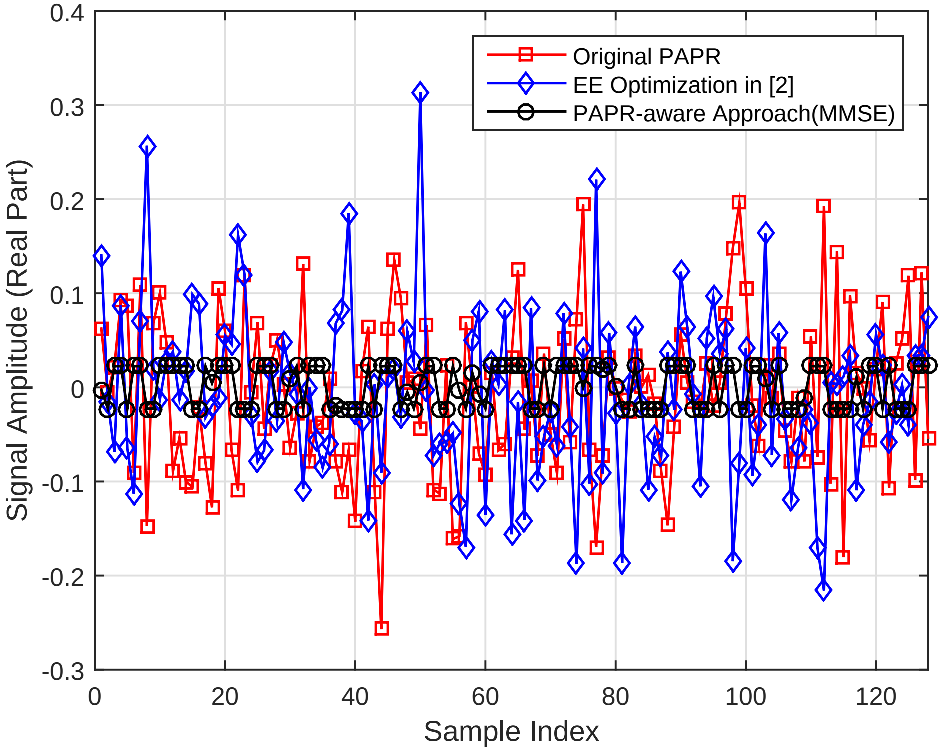


Figure 2.3: Signal representation in the time domain for original OFDM signal, the energy efficiency optimization presented in [1] and the proposed PAPR-aware energy efficiency optimization.

The simulation results are presented in this section to validate our analysis and the proposed algorithms. The isolated small cell is assumed to be circular with the radius of R to simplify the analysis. It could be extended to irregular shaped small cell with adjacent cell interference in future work. The distributed antenna OFDMA system with 128 subcarriers, 5-40 RRHs and 50 MSs has been chosen in this work to evaluate the performance of the PAPR reduction with limited influence on channel capacity, and up to 40 RRHs are considered to evaluate the performance of energy efficiency optimization with excessive DoF. The modulation scheme of QPSK is applied in the simulation. The MSs are randomly distributed within the range of the small cell. The C-RAN is located in the center of the small cell circle. The RRHs are uniformly distributed. The channels between each RRH and user are modeled as independent frequency selective channels. The time division duplex (TDD) is applied in this work to exploit the channel reciprocity and avoid the channel estimation overhead when the RRH number is large. Table I summarizes the simulation parameters.

Figure 2.3 illustrates the real part of one instance of the time-domain digital signal during a modulation symbol interval for different approaches. These simulation results show that our proposed scheme achieves a significant PAPR reduction when compared with the original signal generation and the EE optimization approach from [1].

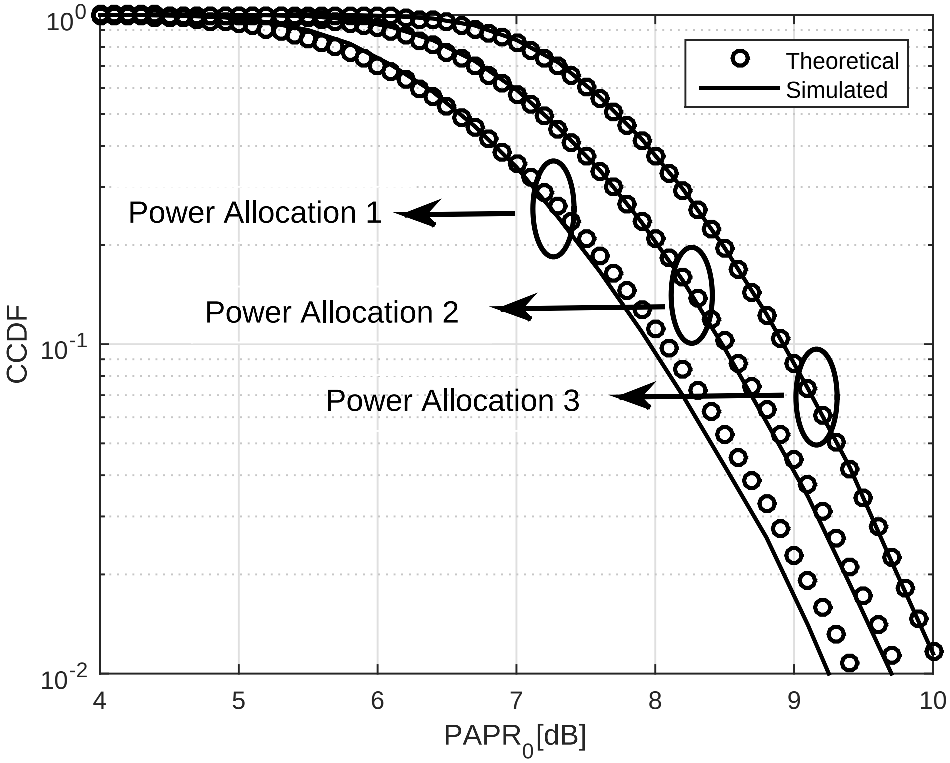


Figure 2.4: Comparison of approximation and simulation results of CCDF for different power allocation result.

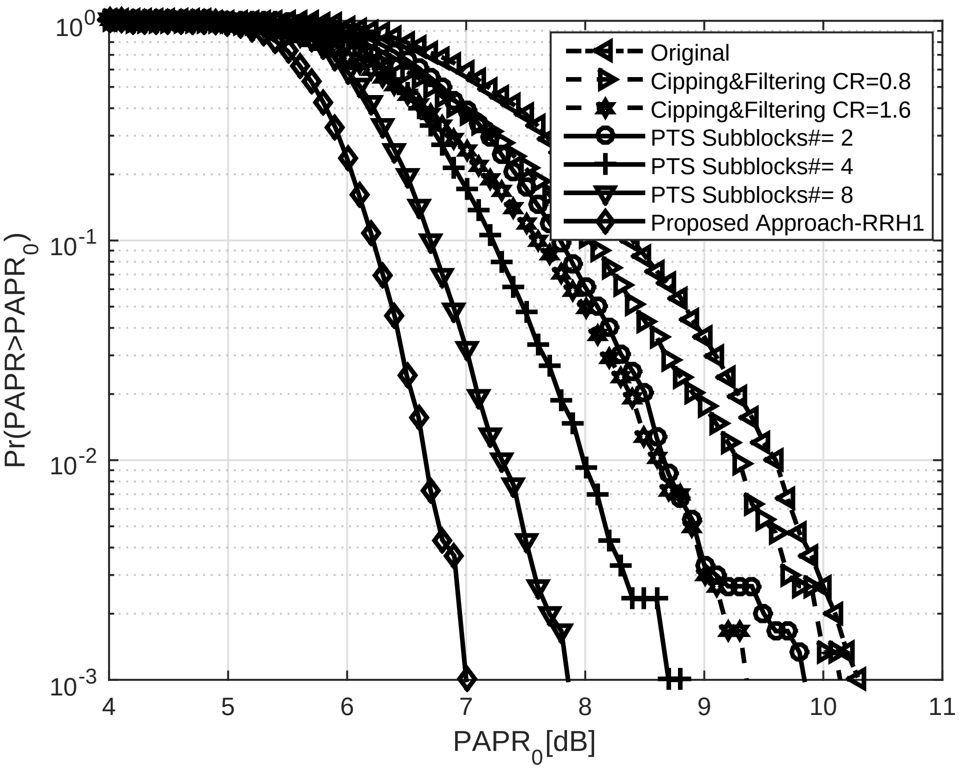


Figure 2.5: Performance comparison of different PAPR reduction schemes.

Figure 2.4 shows the accuracy of approximations of CCDF in (2.8) by comparing the simulation results. We can conclude from the figure that the PAPR of OFDM signals with UPA can be approximated with level cross rate which derived in Appendix A with high precision and accuracy. It also validates the idea of taking the PAPR into account during the energy-efficient resource allocation.

As one of the simplest PAPR reduction approaches, clipping and filtering [15] [16] has been implemented as the PAPR reduction performance baseline. The maximum of the transmit signal is limited by a pre-specified level A, and the clipped signal is applied to reduce the out-of-band radiation. The clipping ratio (CR) is the clipping level A normalized by the root mean square value σ of the OFDM signal, such that

$$\mathrm{CR} = \frac{A}{\sigma}.$$

The PTS scheme [20,21,79], one of the most extensively studied approaches, has been applied in the simulation to compare it with the proposed method in terms of PAPR reduction performance.

Figure 2.5 shows the performance of different PAPR reduction schemes. CR is chosen to be 0.8 and 1.6 in the simulations. The subblock numbers in PTS are chosen to be 2, 4 and 8 respectively, and each partitioned subblock is multiplied by a corresponding complex phase factor. The phase factors are randomly searched to find the optimum set of phase vectors. It should be noted that the search complexity increases exponentially with the number of subblocks. The binary phase factor $\{0, 1\}$ is selected in the simulation to reduce the search complexity and side channel overhead. The optimal phase factors are searched within $2^2, 2^4, 2^8$ enumeration respectively. $PAPR_0$ is compared for original OFDM signal, clipping and filtering with a CR of 0.8 and 1.6, PTS for subblock numbers of 2, 4 and 8 as well as the proposed approach. They are 10.28, 10.11, 9.38, 9.85, 8.83, 7.87 and 7.05 dB respectively, $PAPR_{tr}$ is defined as $Prob\{PAPR > PAPR_{tr}\} = 10^{-3}$. Figure 2.5 presents the results. It shows that the proposed PAPR-aware energy efficiency optimization approach can provide consistent improvement when compared with traditional clipping and filtering or PTS approaches.

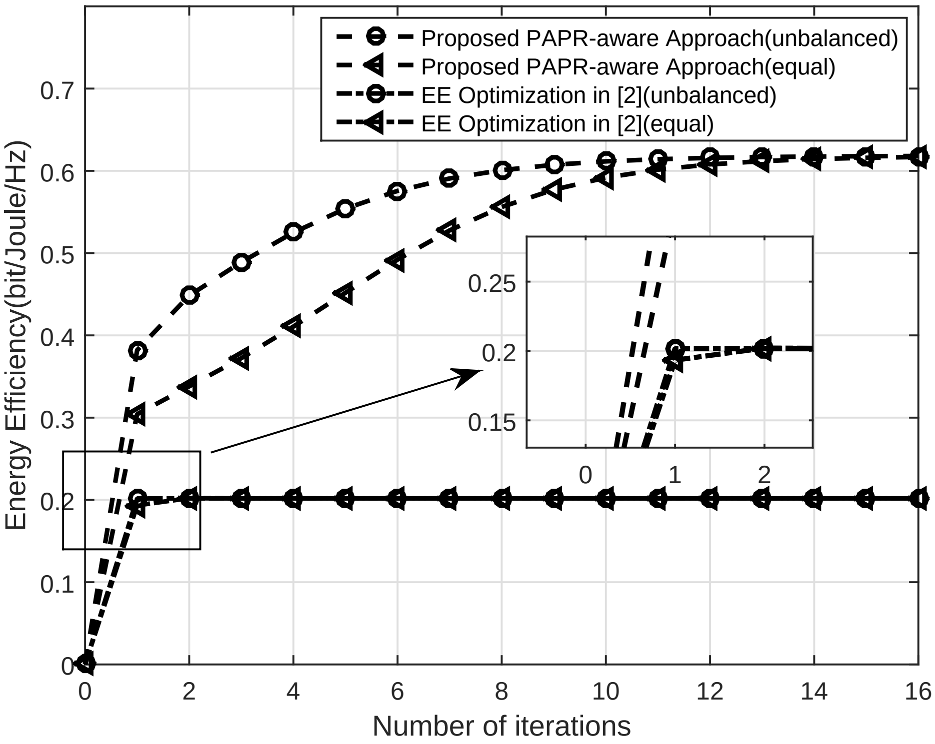


Figure 2.6: Energy efficiency versus number of iterations for the proposed PAPR-aware energy efficiency optimization and the energy efficiency optimization presented in [1].

Figure 2.6 plots the energy efficiency over the number of iterations for energy efficiency optimization with PAPR consideration as proposed in this work and without PAPR consideration as proposed in [1]. The convergence curves for different initial points of energy efficiency optimization (equal and unbalanced power allocation) are also plotted. Both of the algorithms converge to the optimal value after several iterations. Note that the resource allocation algorithm performed between the interval of the channel estimation, which is much larger than the extra time needed for computing the additional iterations in common BS hardware. Figure 2.6 indicates that although the convergence speed of the proposed algorithm is slightly slower than the method proposed in [1] because of the dual fractional programming algorithm, the optimal energy efficiency value of our approach is nearly three times higher than the result of the method in [1].

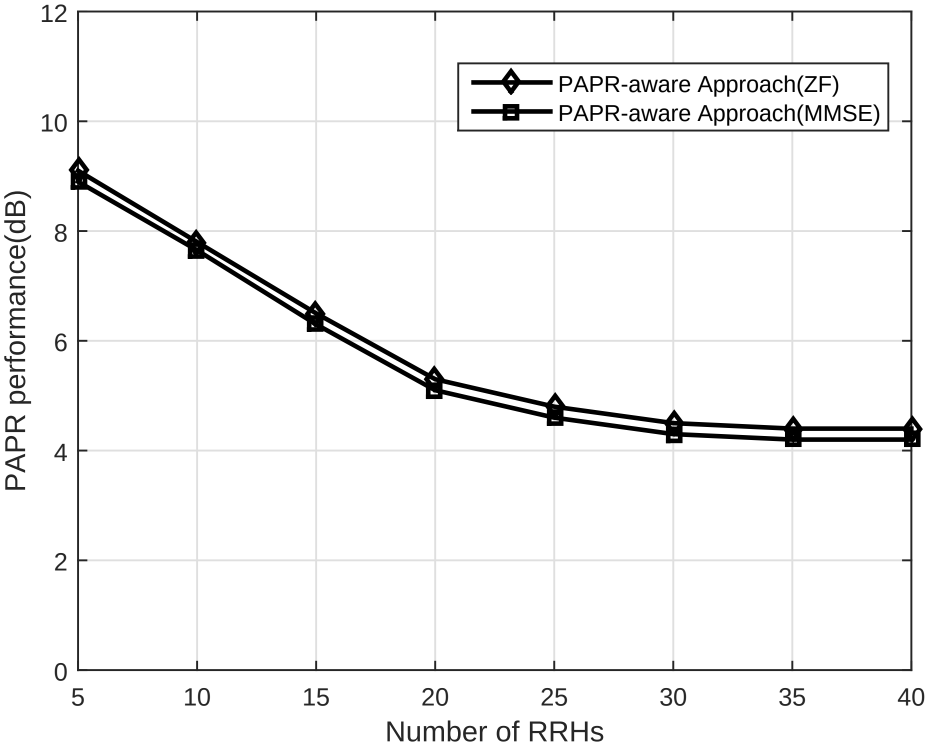


Figure 2.7: PAPR versus number of RRHs for the proposed PAPR-aware energy efficiency optimization and the energy efficiency optimization in [1].

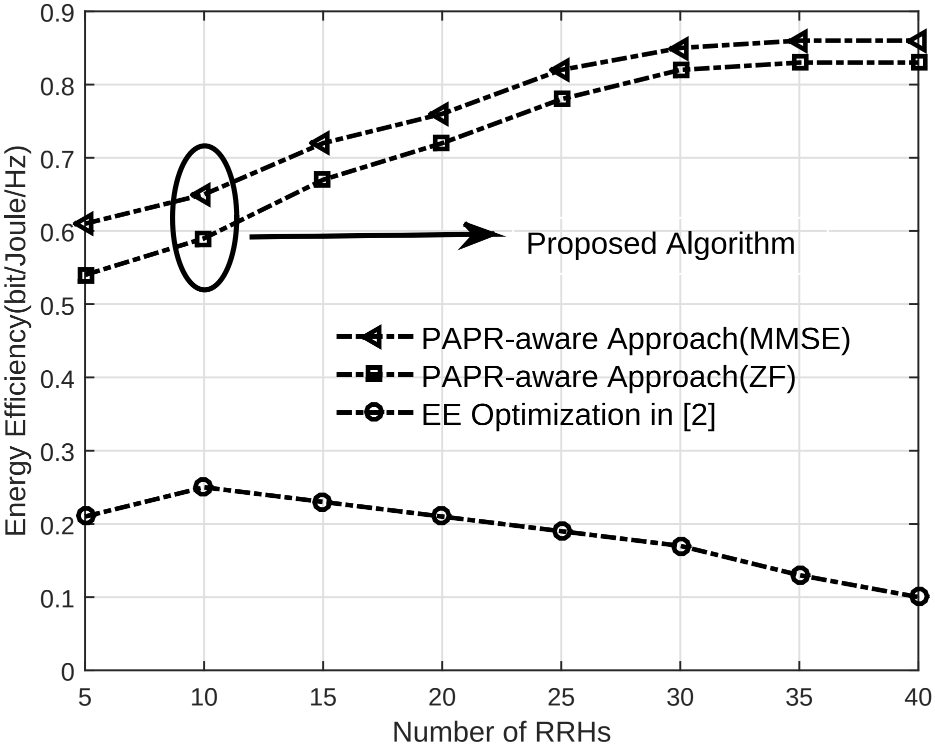


Figure 2.8: Energy efficiency over the number of RRHs for PAPR-aware energy efficiency optimization and the approach proposed in [1].

The simulation results in Figure 2.7 show that the performance of our approach improves significantly when we increase number of RRHs. The intuition of improved PAPR performance with increasing RRH number is that the PAPR-aware algorithm proposed in this work can exploit the excessive DoF of the distributed RRHs.

Figure 2.8 plots the energy efficiency over the number of RRHs for the proposed solution and the approach from [1]. The general trend shows that energy efficiency of the proposed method increases with the number of RRHs, especially when the number of RRHs is small. The PAPR-aware energy-efficient algorithm increases with the number of RRHs since the DoF is increased providing more space for optimization. It is shown in the simulation results that maximum energy efficiency is achieved for as many as 30 RRHs. The results also show that the energy efficiency proposed in [1] increases when the number of RRH is small, but decreases when increasing the number of RRHs beyond the optimal point.

2.6 Conclusion

This chapter has derived a PAPR-aware energy-efficient resource allocation scheme for the downlink of OFDMA/SDMA wireless communication systems in the DAS configuration. The expectation of PAPR-aware energy-efficient DAS structure is very promising in two aspects: 1) the performance of PAPR reduction enables the operator to equip the RRH with low-cost PA, which helps the operator to reduce the initial investment of the DAS deployment significantly and; 2) the operational expenditure of DAS in terms of energy consumption can be reduced up to 90% especially with the large scale DAS. This also helps the mobile RRH to extend the battery life to a great extent.

Based on the theory of PAPR-aware energy-efficient resource allocation in DAS which developed in this chapter, we will extend it to massive MIMO. With large scale collocated BS antenna, the energy-efficient resource allocation in massive MIMO system has more degree of freedom for both subcarrier and power allocation than the DAS.

Appendix A

LEVEL CROSSING RATE OF OFDM SIGNAL WITH UNEQUAL POWER ALLOCATION

In this appendix, we derive the level-crossing rate of OFDM signals with unequal power allocation, the OFDM signal can be formulated as

$$s(t) = \sum_{k=0}^{N-1} A_k e^{jw_k t} = x(t) + jy(t)$$

where $w_k = \frac{2\pi}{N} k\Delta f$ is the angular frequency of kth subcarrier and $E\{|A_k|^2\} = P_k$ is the power allocated to kth subcarrier. Thus, we have

$$x(t) = \sum_{k=0}^{N-1} (A_k^R \cos w_k t - A_k^I \sin w_k t)$$

where A_k^R and A_k^I are the real and imaginary components of A_k. Similarly, we can get

$$y(t) = \sum_{k=0}^{N-1} (A_k^R \sin w_k t + A_k^I \cos w_k t)$$

According to central limit theory, $x(t)$ and $y(t)$ are statistically independent zero-mean Gaussian random processes for large enough N, e.g. N is 64. By assuming $A_k^I = 0$ for BPSK and A_k^R is independent with A_k^I for higher order modulation scheme and assuming $x'(t) = dx(t)/dt$, thus the cross-correlation $R_{x'y}(t, t + \tau)$ with any order of modulation can be represented as

$$E\{y(t)x'(t+\tau)\} = E\{\sum_{k=0}^{N-1}(A_k^R \sin w_k t + A_k^I \cos w_k t)$$

$$\sum_{k=0}^{N-1}(-A_k^R w_k \sin w_k(t+\tau) - A_k^I w_k \cos w_k(t+\tau))\}$$

$$= -\sum_{k=0}^{N-1} P_k w_k \cos(w_k\tau)$$

and also the cross-correlation $R_{y'x}(t, t+\tau)$ can be represented as

$$E\{x(t)y'(t+\tau)\} = \sum_{k=0}^{N-1} P_k w_k \cos(w_k\tau)$$

Similarly, we can derive

$$E\{x(t)x(t+\tau)\} = E\{y(t)y(t+\tau)\} = \sum_{k=0}^{N-1} P_k \cos(w_k\tau)$$

and

$$E\{x'(t)x'(t+\tau)\} = E\{y'(t)y'(t+\tau)\} = \sum_{k=0}^{N-1} w_k^2 P_k \cos(w_k\tau)$$

As a result, the autocovariance and cross-covariance between each other are as follows

$$\sigma_{xx} = \sigma_{yy} = \sum P_k = \sigma_1^2$$

$$\sigma_{xy} = \sigma_{yx} = \sigma_{xx'} = \sigma_{yy'} = 0$$

$$\sigma_{x'x'} = \sigma_{y'y'} = \sum w_k^2 P_k = \sigma_2^2$$

$$\sigma_{xy'} = \sum w_k P_k = \sigma_3^2 = -\sigma_{x'y}$$

The covariance matrix is given by

$$R_{xy'yx'} = \begin{bmatrix} \sigma_1^2 & \sigma_3^2 & 0 & 0 \\ \sigma_3^2 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_1^2 & -\sigma_3^2 \\ 0 & 0 & -\sigma_3^2 & \sigma_2^2 \end{bmatrix}$$

Since x, y' and y, x' are independent with each other, the joint pdf of x,y',y,x' can be represented as

$$f_{xy'yx'} = f_{xy'} f_{yx'} = \frac{1}{(2\pi)^2\sigma_1^2\sigma_2^2(1-\rho^2)} \cdot$$
$$\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{x^2+y^2}{\sigma_1^2} + \frac{x'^2+y'^2}{\sigma_2^2} + 2\rho\frac{yx'-xy'}{\sigma_1\sigma_2}\right]\right\}$$

where $\rho = \sigma_3^2/\sigma_1\sigma_2$ is the correlation coefficient of x and y'.

Then the Cartesian coordinates (x, y) is transformed into polar coordinates (r, θ) by means of $r = \sqrt{x^2+y^2}$ and $\theta = \arctan(y/x)$ which leads to the following equations

$$x = r\cos\theta, \quad x' = r'\cos\theta - r\theta'\sin\theta$$
$$y = r\sin\theta, \quad y' = r'\sin\theta + r\theta'\cos\theta$$

According to the transformation rule, the result joint pdf in polar coordinates can be derived as

$$f_{rr'\theta\theta'} = |J|^{-1} f_{xy'yx'}(r\cos\theta, r'\sin\theta + r\theta'\cos\theta,$$
$$r\sin\theta, r'\cos\theta - r\theta'\sin\theta)$$

where

$$J = \begin{vmatrix} \partial x/\partial r & \partial x/\partial r' & \partial x/\partial \theta & \partial x/\partial \theta' \\ \partial y/\partial r & \partial y/\partial r' & \partial y/\partial \theta & \partial y/\partial \theta' \\ \partial x'/\partial r & \partial x'/\partial r' & \partial x'/\partial \theta & \partial x'/\partial \theta' \\ \partial y'/\partial r & \partial y'/\partial r' & \partial y'/\partial \theta & \partial y'/\partial \theta' \end{vmatrix}^{-1} = -\frac{1}{r^2}$$

represents for the Jacobian determinant. Thus, the joint pdf can be given as

$$f_{rr'\theta\theta'} = \frac{r^2}{(2\pi)^2\sigma_1^2\sigma_2^2(1-\rho^2)} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{r^2}{\sigma_1^2} + \frac{r'^2 + r^2\theta'^2}{\sigma_2^2} - 2\rho\frac{r^2\theta'}{\sigma_1\sigma_2}\right]\right\}$$

for $r \geq 0$, $|r'| < \infty$, $|\theta| \leq \pi$ and $|\theta'| < \infty$. Therefore, we can get the joint distribution of $f_{r,r'}$ by integrating $f_{rr'\theta\theta'}$ by θ from $-\pi$ to π, θ' from $-\infty$ to ∞.

$$\begin{aligned} f_{rr'} =& \int_{-\pi}^{\pi}\int_{-\infty}^{\infty} f_{rr'\theta\theta'} \mathrm{d}\theta' \mathrm{d}\theta \\ =& \frac{r^2}{2\pi\sigma_1^2\sigma_2^2(1-\rho^2)} \exp\left\{-\frac{r^2}{2\sigma_1^2} - \frac{r'^2}{2\sigma_2^2(1-\rho^2)}\right\} \\ & \int_{-\infty}^{\infty} \exp\left\{-\frac{r^2}{2(1-\rho^2)}\left(\frac{\theta'}{\sigma_2} - \frac{\rho}{\sigma_1}\right)^2\right\} \mathrm{d}\theta' \\ =& \frac{r}{\sqrt{2\pi}\sqrt{1-\rho^2}\sigma_1^2\sigma_2} \exp\left\{-\frac{r^2}{2\sigma_1^2} - \frac{r'^2}{2\sigma_2^2(1-\rho^2)}\right\} \end{aligned}$$

Since the level crossing rate of process $r(t)$ at level $r = l$ is generally defined by

$$N_C(l) := \int_0^{\infty} r' f_{rr'}(l, r') \mathrm{d}r', l \geq 0$$

As a result, the level crossing rate can be derived as

$$\begin{aligned}N_C(l) =& \frac{\sqrt{1-\rho^2}\sigma_2 l}{\sqrt{2\pi}\sigma_1^2}\exp\left\{-\frac{l^2}{2\sigma_1^2}\right\}\\ =& \frac{\sqrt{\sum P_k \sum w_k^2 P_k - (\sum w_k P_k)^2}}{\sqrt{2\pi}(\sum P_k)^{3/2}} l \exp\left\{-\frac{l^2}{2\sum P_k}\right\}\\ =& \frac{\sqrt{2\pi}\Delta f \sqrt{\sum P_k \sum k^2 P_k - (\sum k P_k)^2}}{N(\sum P_k)^{3/2}} l \exp\left\{-\frac{l^2}{2\sum P_k}\right\}\end{aligned}$$

Appendix B

Derivation of Relationship Between UPA Factor and PA Efficiency

Since the power efficiency varies with the PAPR of OFDM signal, the probability of PAPR $\gamma \leq f_e(\eta_0)$ should satisfy some predetermined value to guarantee the power amplifier operating with efficiency η_0, where the $f_e(\eta)$ represents for the function between PA efficiency and PAPR. The PAPR value with a specific UPA factor which satisfied the predetermined probability is derived in this appendix. Assume the predetermined probability is p_0, it can be obtained by equation (7) that

$$\nu\sqrt{\gamma_0}e^{-\frac{\gamma_0}{2\sum P_k}} = f_0$$

Thus, we can get the relationship between UPA factor and PAPR by Taylor expansion

$$\begin{aligned}
\nu^2 &= \frac{f_0^2 \exp\left\{\frac{\gamma_0}{\sum P_k}\right\}}{\gamma_0} \\
&\approx f_0^2 \left(\frac{a + b\eta_0}{a^2} + \frac{1}{\sum P_k} + \frac{a - b\eta_0}{2 \sum {P_k}^2} + o(\eta_0) \right) \\
&\approx f_0^2 \left\{ \frac{1}{a} + \frac{2 \sum P_k + a}{2 \sum {P_k}^2} - \left(\frac{b}{2 \sum {P_k}^2} - \frac{b}{a^2} \right) \eta_0 \right\}
\end{aligned}$$

where we assume the characteristic function of PA efficiency $f_e(\eta)$ is fitted with linear function

$$\gamma = f_e(\eta) = a - b\eta$$

with both a and b are positive constant numbers, and we can derive the relationship between η and ν^2 as follows

$$\eta = c_1 + c_2 \nu^2$$

where

$$c_1 = \left(\frac{1}{a} + \frac{2 \sum P_k + a}{2 \sum {P_k}^2} \right) \bigg/ \left(\frac{b}{2 \sum {P_k}^2} - \frac{b}{a^2} \right)$$

and

$$c_2 = 1 \bigg/ \left\{ f_0^2 \left(\frac{b}{2 \sum {P_k}^2} - \frac{b}{a^2} \right) \right\}$$

Chapter 3

Recurrent Neural Network Inspired PAPR-Aware Downlink Massive MIMO-OFDM Systems

3.1 Introduction

The massive deployments of heterogeneous wireless networks and the emerging 5G NR have motivated the demand for energy and spectrum efficiency in order to reduce operator costs. OFDM has been adopted in 4G LTE (and its evolution) [80]. However, OFDM suffers from the drawback of high PAPR at the transmitter [15–17]. The high peaks which are generated by the constructive addition of different subcarriers lead to the signal excursions into the nonlinear region of the PA. Typically, to avoid this nonlinear distortion, the input power to the PA is reduced or "backed-off". However, the amplifier is less power efficient and ultimately more expensive when operating in this input back-off region. This is not a realistic solution for the 5G NR system since the target energy efficiency improvement is

100x between 4G and 5G [5]. Large scale MIMO (a.k.a. massive MIMO) systems have been proposed as the key enabler of 5G due to their ability to mitigate the MUI and improve link reliability and spectrum efficiency [86]. While highly-linear PAs is desirable, the cost scales linearly with the number of antennas and is, thus, prohibitively expensive for massive MIMO BS deployment. Massive MIMO systems have the potential to reduce the PAPR on a symbol basis by exploiting the extra spatial DoFs and hence adopt less expensive PAs [2, 87–93].

The improvements brought by the existing PAPR reduction techniques either sacrifice spectral efficiency by degrading the effective data rate or causing spectral regrowth. This research proposes a new PAPR reduction method which mimics the structure of an artificial neural network (ANN). The proposed structure exploits the additional DoFs with a large number of antenna, and it has minimal effect on either data rate or MUI.

3.2 System Model and Problem Formulation

The purpose of the PAPR-aware massive MIMO-OFDM downlink precoding design is to find transmit signal to meet the predetermined PAPR and MUI targets. In order to quantify the MUI, the overall massive MIMO-OFDM downlink precoding constraint need to be formulated. Consider the downlink massive MIMO system of Fig. 3.1 which has M_r single antenna users and one BS which equipped with N_t antennas. The number of BS antennas is significantly larger than number of simultaneous users, $N_t \gg M_r$.

To illustrate the excess DoFs which is enabled by the large scale transmit antenna array, we start with a single-carrier multiuser MIMO system and then extend it to the case of N_c subcarriers. The single-carrier received signal $\dot{\mathbf{y}} \in \mathbb{C}^{M_r \times 1}$ can be represented as

$$\dot{\mathbf{y}} = \dot{\mathbf{s}} + \left(\dot{\mathbf{H}}\dot{\mathbf{X}} - \dot{\mathbf{s}}\right) + \dot{\mathbf{w}}, \tag{3.1}$$

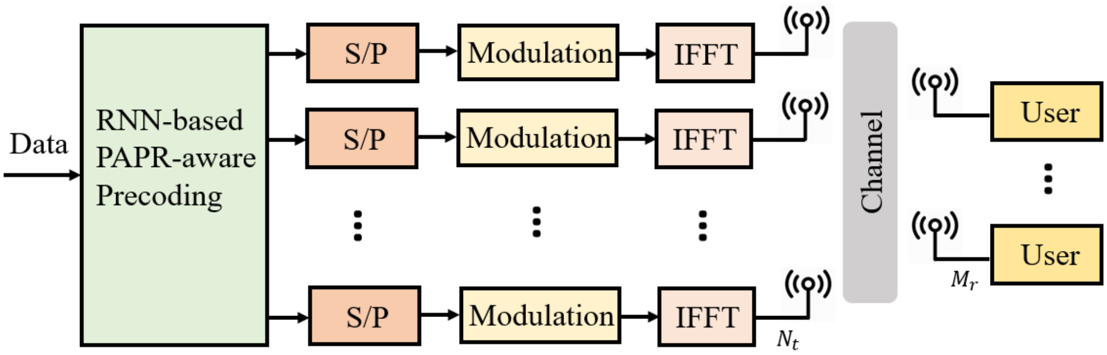


Figure 3.1: System diagram of the proposed PAPR-aware downlink massive MIMO-OFDM system based on recurrent neural network.

where $\dot{\mathbf{s}} \in \mathbb{C}^{M_r \times 1}$ represents the complex constellation before precoding, $\dot{\mathbf{H}} \in \mathbb{C}^{M_r \times N_t}$ the flat fading channel coefficients where the (m, n) entry represents the complex Gaussian channel tap between the mth user and nth BS antenna, $\dot{\mathbf{X}} \in \mathbb{C}^{N_t \times 1}$ the single-carrier transmit signal at the antennas, $\dot{\mathbf{w}} \in \mathbb{C}^{M_r \times 1}$ the additive white Gaussian noise, and $(\dot{\mathbf{H}}\dot{\mathbf{X}} - \dot{\mathbf{s}})$ the MUI term. The entries of $\dot{\mathbf{H}}$ and $\dot{\mathbf{w}}$ are independent and identically distributed (iid). The following precoding constraint must be satisfied to eliminate the MUI:

$$\dot{\mathbf{s}} = \dot{\mathbf{H}}\dot{\mathbf{X}}. \tag{3.2}$$

The channel matrix is underdetermined since we assume the number of BS antennas is significantly larger than the number of users. However, the formulation of an OFDM waveform is more complicated than the single-carrier case since PAPR is related to constructive and destructive addition of multiple carriers in the time domain, whereas precoding is related with desired spatial characteristics. The overall problem of multicarrier precoding across different antennas and subcarriers is formulated in Section 3.2.1.

3.2.1 Multicarrier Frequency-Space Precoding Formulation

Assuming perfect channel state information is available at the transmitter (CSIT), linear precoding can be applied at the BS to eliminate the MUI at the receiver. Zero forcing (ZF) precoding has been widely applied in massive MIMO due to its simplicity and exceptional performance [2]. By applying ZF precoding for the mth subcarrier, the transmit symbols satisfy $\mathbf{s}_m = \mathbf{H}_m\mathbf{X}_m (1 \leq m \leq N_c)$. To fully exploit DoFs of the large scale array, the user information symbols at the BS need to be mapped to the transmit antennas so that the information received by each user has no interference from the other users [2, 93].

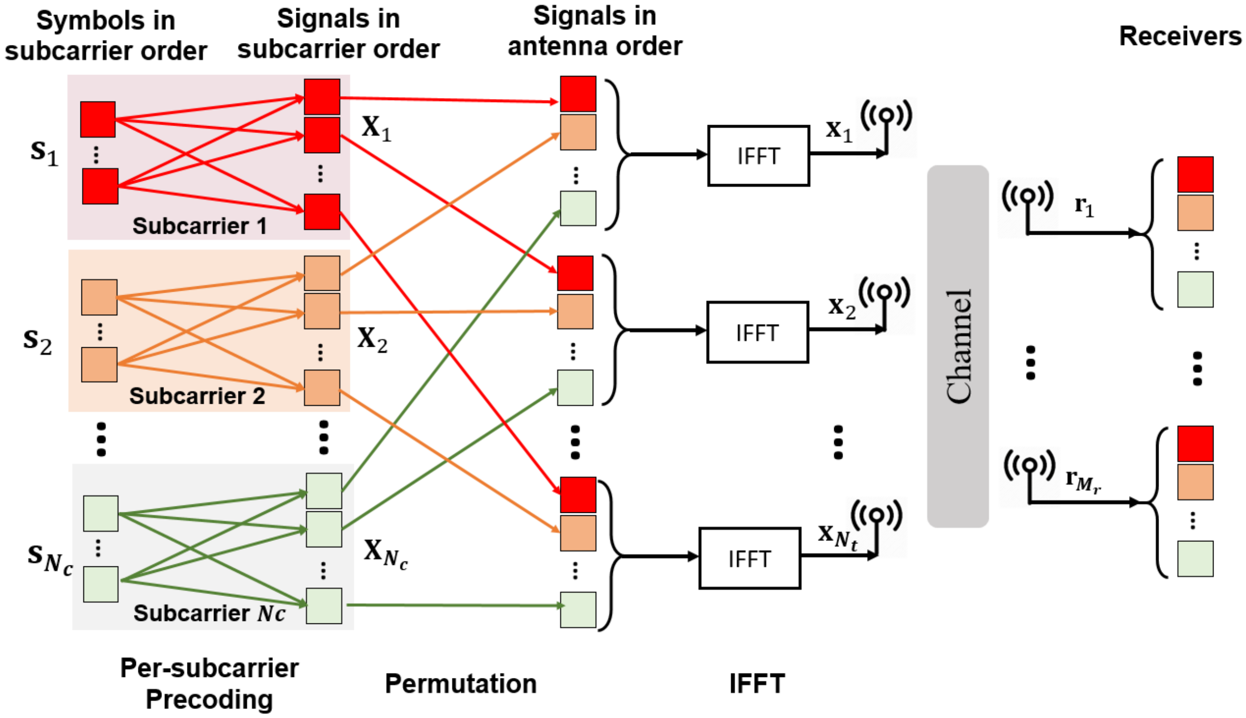


Figure 3.2: Multicarrier frequency-space precoding formulation with permutation and IFFT.

As shown in Fig. 3.2, the frequency domain transmit signals are permuted and N_c-point inverse fast Fourier transform (IFFT) operations are carried out for each antenna. As a result, the overall multicarrier frequency-space precoding can be represented as

$$\mathbf{s} = \underbrace{\begin{bmatrix} \mathbf{H}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{H}_{N_c} \end{bmatrix}}_{\triangleq \mathbf{H}} \underbrace{\begin{bmatrix} \mathbf{I}_{1,1} & \mathbf{I}_{1,2} & \dots & \mathbf{I}_{1,N_t} \\ \mathbf{I}_{2,1} & \mathbf{I}_{2,2} & \dots & \mathbf{I}_{2,N_t} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{I}_{N_c,1} & \mathbf{I}_{N_c,2} & \dots & \mathbf{I}_{N_c,N_t} \end{bmatrix}}_{\triangleq \bar{\mathbf{P}}} \underbrace{\begin{bmatrix} \mathbf{Q} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{Q} \end{bmatrix}}_{\triangleq \bar{\mathbf{Q}}} \underbrace{\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_{N_t} \end{pmatrix}}_{\mathbf{x}}, \quad (3.3)$$

where $\mathbf{s}$ represents the multicarrier information symbols, $\mathbf{H}_m$ the $M_r \times N_t$ channel matrix of mth subcarrier, $\bar{\mathbf{P}}$ the permutation matrix which is comprised by the $N_t \times N_c$ matrix $\mathbf{I}_{m,n}$ (the (n, m) entry is 1 for $1 \leq m \leq N_c$ and $1 \leq n \leq N_t$, and the other entries of $\mathbf{I}_{m,n}$ are 0s), $\bar{\mathbf{Q}}$ the $N_x \times N_x$ diagonalization matrix comprised by FFT matrix $\mathbf{Q}$, where $N_x \triangleq N_t N_c$, $\mathbf{x}$ the time-domain transmit signal. Note that the permutation matrix satisfies $\bar{\mathbf{P}}\bar{\mathbf{P}}^H = \mathbf{I}_{N_x}$ and $\bar{\mathbf{P}}^H$ can be considered as the reverse permutation matrix. Thus we have

$$\mathbf{s} = \mathbf{H}\bar{\mathbf{P}}\bar{\mathbf{Q}}\mathbf{x} = \mathcal{H}\mathbf{x}, \quad (3.4)$$

where $\mathcal{H} = \mathbf{H}\bar{\mathbf{P}}\bar{\mathbf{Q}}$ represents the mapping operation of the time domain transmit signal to the information symbols of the overall massive MIMO system.

3.2.2 Optimization Problem Formulation

Consider N_c subcarriers comprising an OFDM complex baseband signal, the PAPR of mth antenna signal $\mathbf{x}_m$ is defined as the ratio between the maximum power and the average power [7]

$$\text{PAPR}(\mathbf{x}_m) = \frac{N_c \left\|\mathbf{x}_m\right\|_\infty^2}{\left\|\mathbf{x}_m\right\|_2^2}. \quad (3.5)$$

Note that the PAPR of complex baseband signal $\mathbf{x}_m$ satisfy that $1 \leq \mathrm{PAPR}(\mathbf{x}_m) \leq N_c$ due to the standard vector norm relationships. We reduce PAPR by manipulating on the precoding vector. Assuming that the transmitted signal is normalized over subcarriers

$$\tilde{\mathbf{x}}_m = \frac{\mathbf{x}_m}{\sqrt{\sum_{m=1}^{N_c} \mathbf{x}_m}}, m = 1, \cdots, N_c. \tag{3.6}$$

With the auxiliary parameter y, the proposed PAPR can be relaxed as $\|\mathbf{x}_m\|_\infty$ and hence be represented by $[\tilde{\mathbf{x}}^H \odot \tilde{\mathbf{x}} - \mathbf{e}^T y]^+ + y$ during the iteration of the proposed recurrent neural network, where $\tilde{\mathbf{x}} \in \mathbb{C}^{NtN_c}$ and $y \in \mathbb{R}^+$.

Therefore, the overall massive MU-MIMO-OFDM downlink problem can be formulated as

$$\begin{aligned} &\arg\min_{\tilde{\mathbf{x}}} && [\tilde{\mathbf{x}}^H \odot \tilde{\mathbf{x}} - \mathbf{e}^T y]^+ + y && (3.7) \\ &\text{subject to} && \mathbf{s} = \mathcal{H}\tilde{\mathbf{x}} && (3.7\text{a}) \\ & && y \geq 0 && (3.7\text{b}) \end{aligned}$$

where $\odot$ denotes for element-wise multiplication, and $\mathbf{e}$ denotes for N_c-length vector which is $[1, 1, \cdots, 1]$, the maximum dynamic range y is defined as the recurrent neural network activation variable, which also represents for the crest factor of PA implemented by the activation function of recurrent neural network. $(\tilde{\mathbf{x}}, y)^T \in \mathbb{U}_0$ is the convex set where the solution exists. It is clear that the recurrent neural network variable y in this work represents the minimum PA efficiency among all the transmit antenna, we will prove that the optimal recurrent neural network variable exists in our proposed structure in the next section. For the given optimal recurrent neural network variable, the proposed structure guarantee that the signal $\tilde{\mathbf{x}}$ at all transmit antenna will not be distorted. In addition, we will prove that the optimal recurrent neural network variable is unique.

3.3 Stability and Convergence Analysis

The recurrent neural network has been widely shown to be suitable for real time implementation with finite and exponential convergence in various applications [82, 83]. In this section, we prove the global convergence of the proposed neural network for solving PAPR-aware downlink massive MIMO-OFDM problem and the uniqueness of the optimal PA efficiency solution.

3.3.1 Dynamic Equations

In order to formulate the dynamic equation to derive the optimal precoding vector and recurrent neural network activation variable, we have to restate the optimization in (3.7) in Lagrangian form. First of all, let's relax the precoding constraint $\mathbf{s} = \mathcal{H}\tilde{\mathbf{x}}$ in (3.7a) to the Euclidean norm form $\|\mathbf{s} - \mathcal{H}\tilde{\mathbf{x}}\|_2^2 \leq \eta$, noting that the relaxation does not significantly degrade the BER performance for small values of η [2]. Although the objective function in recurrent neural network is the same as the energy function for most applications [84, 85], it is not the case in this work since we have to make sure the optimal recurrent neural network activation variable should be unique. With an auxiliary parameter λ, the scalar-valued objective funciton of the recurrent neural network system is defined as the nonnegative Lagrangian function of the overall system

$$F(\tilde{\mathbf{x}}, y) = \lambda \left([\tilde{\mathbf{x}}^H \odot \tilde{\mathbf{x}} - \mathbf{e}^T y]^+ + y \right) + \|\mathbf{s} - \mathcal{H}\tilde{\mathbf{x}}\|_2^2 . \tag{3.8}$$

With the nonnegative Lagrangian function defined in (3.8), the dynamic equations for solving (3.7) can be derived with the help of the negative gradient as follows

$$\frac{du}{dt} = -\mu \nabla F(u) \tag{3.9}$$

where $u = (\tilde{\mathbf{x}}, y)^T \in \mathbb{U}_0$, the nonnegative Lagrangian function $F(u)$ defined in (3.8), $\nabla F(u)$ is the gradient of the objective function, and μ is a positive scalar constant which is used to scale the convergence rate of the recurrent neural network.

Lemma 3.1. *The derivative of* $\|\mathbf{s} - \mathcal{H}\tilde{\mathbf{x}}\|_2^2$ *with respect to* $\tilde{\mathbf{x}}$ *could be simply derived as* $2Re\left(\mathcal{H}^H(\mathcal{H}\tilde{\mathbf{x}} - \mathbf{s})\right)$.

Proof. See Appendix C □

Therefore, the dynamic equations for the recurrent neural network solving (3.7) are:

$$\frac{d\tilde{\mathbf{x}}}{dt} = -2\lambda\mu\tilde{\mathbf{x}}^H P_\Omega(\tilde{\mathbf{x}}^H \odot \tilde{\mathbf{x}} - \mathbf{e}^T y) + 2\mu \mathrm{Re}\left(\mathcal{H}^H(\mathcal{H}\tilde{\mathbf{x}} - \mathbf{s})\right) \tag{3.10}$$

and

$$\frac{dy}{dt} = -\lambda\mu P_{\hat{\Omega}}(\tilde{\mathbf{x}}^H \odot \tilde{\mathbf{x}} - \mathbf{e}^T y) \tag{3.11}$$

where the activation function $P_\Omega(x)$ and deactivation function $P_{\hat{\Omega}}(x)$ are defined respectively as

$$P_\Omega(x_i) = \begin{cases} 1, & x_i \geq 0 \\ 0, & x_i < 0 \end{cases} \tag{3.12}$$

and

$$P_{\hat{\Omega}}(x_i) = \begin{cases} 1, & x_i < 0 \\ 0, & x_i \geq 0 \end{cases} \tag{3.13}$$

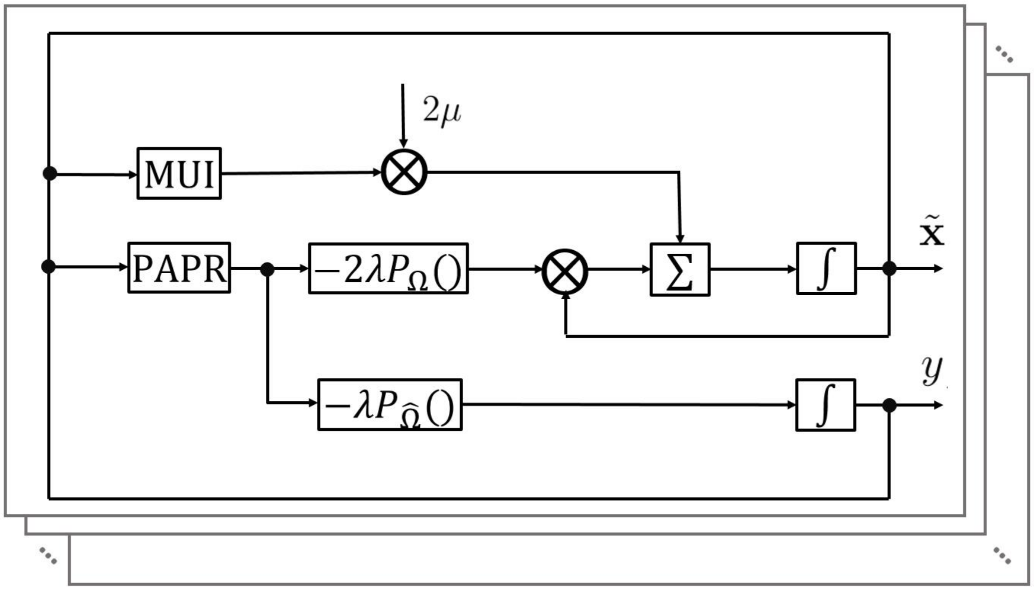


Figure 3.3: Block diagram of the proposed recurrent neural network.

The block diagram of the proposed recurrent neural network is shown in Figure 3.3, where PAPR block represents for the level of PAPR which is $(\tilde{\mathbf{x}}^H \odot \tilde{\mathbf{x}} - \mathbf{e}^T y)$, and MUI block for amount of MUI which is $\mathrm{Re}\left(\mathcal{H}^H(\mathcal{H}\tilde{\mathbf{x}} - \mathbf{s})\right)$, the layers in Figure 3.3 represent for different antennas. The activation function $P_{\Omega}(x_i)$ and deactivation function $P_{\hat{\Omega}}(x_i)$ are introduced in this research to emulate the ideal cut-off characteristic beyond the 1-dB compression point. The residue part of the signal after cutting-off is back-propagated to the perceptron which represented for the ZF precoding scheme, and $\tilde{\mathbf{x}}$ is trained by the signal residue to satisfy the precoding constraint in the mean time.

Algorithm 4 Recurrent neural network based massive MU-MIMO PAPR reduction algorithm

```
Input: Comprehensive channel state $\mathcal{H}$, complex constellation information symbol $\mathbf{s}$
Output: Transmit signal $\tilde{\mathbf{x}}$, activation variable $y$
    Initialization : $\lambda \leftarrow 0.4$, $\mu \leftarrow 0.05$, $\tilde{\mathbf{x}} \leftarrow \mathbf{0}$, $y \leftarrow 0$
while the desired accuracy is not satisfied do
    $\nabla\tilde{\mathbf{x}} \leftarrow (3.10)$
    $\nabla y \leftarrow (3.11)$
end while
return $\tilde{\mathbf{x}}$, $y$
```

Comparing with the existing solution [2], the proposed recurrent neural network solution has the advantages listed as follows:

1) The solution can be computed using parallelized operations and thus is especially suitable for tackling large-scale problems with parallel hardware.

2) With the global convergence property, the neural network solution converges to a unique and optimal solution.

3) The low complexity architecture of the proposed method can be implemented by reconfigurable hardware, e.g., an FPGA, which means that the neural network can be implemented in real time.

3.3.2 Neural Network Complexity Analysis and Hardware Implementation

Using the block diagram for the proposed recurrent neural network, Figure 3.3, the complexity of the neural network architecture is analyzed in this section. Assume a large-scale MIMO system equipped with N_t transmit antennas, M_r single antenna users, and N_c subcarriers. The recurrent neural network described in (3.10)-(3.11) contains $N_t \times N_c + 1$ neurons.

The asymptotic complexity of the proposed model is $O(N_t N_c M_t)$ for both multiplications and additions/subtractions per iteration. We shall see that the large quantity of neurons are well-suited for off-the-shelf hardware.

The computational intensive operations of RNN to be implemented in hardware are matrix/vector multiplications and nonlinear operations. Matrix/vector multiplication is inherently parallel and can be accelerated on parallel architectures. The FPGA-based coprocessor can efficiently implement pipelined operators with large-scale parallelism, and thus achieving high performance using programmable logic devices. The FPGA-based coprocessor comprises an array of processing elements called Multiply ACcumulate (MAC) unit which is implemented on programmable logic and interfaces with the modem via the high-speed bus. Input data in the form of modulated data symbols are streamed into the MAC array, one batch of data per clock cycle. The coprocessor uses memory to cache incoming data. Output data can then be rerouted to inputs to perform recurrent sequences of operations.

3.3.3 Energy Function

The energy function of the proposed recurrent neural network is defined as follows

$$\mathcal{E}(\tilde{\mathbf{x}}, y) = \lambda[\tilde{\mathbf{x}}^H \odot \tilde{\mathbf{x}} - \mathbf{e}^T y]^+ + \|\mathbf{s} - \mathcal{H}\tilde{\mathbf{x}}\|_2^2 \tag{3.14}$$

where $\lambda \in \mathbb{R}_+$ is a regulation factor, and $\mathbf{e}$ denotes for N_c-length vector which is $[1, 1, \cdots, 1]$. The first term in (3.14) represents for the amount of unwanted out-of-band radiation and signal distortion introduced by the high dynamic range of OFDM signal, the second term in (3.14) represents for the MUI. Clearly, the energy function (3.14) is convex and continuously differentiable. It can be seen that $\mathcal{E}(\tilde{\mathbf{x}}^*, y^*) = 0$ if and only if both the signal distortion and the MUI are eliminated.

Lemma 3.2. *Suppose that $\mathcal{E} : \mathbb{U} \subset \mathbb{R}^N \to \mathbb{R}$ is differentiable and convex set $\mathbb{U}_0 \subset \mathbb{U}$. Then $\mathcal{E}$ is convex on $\mathbb{U}_0$ if and only if*

$$(z - y)\nabla\mathcal{E}(y) \leq \mathcal{E}(z) - \mathcal{E}(y), \forall y, z \in \mathbb{U}_0 \tag{3.15}$$

where $\nabla\mathcal{E}(y)$ is the gradient of $\mathcal{E}(y)$.

Proof: First suppose $\mathcal{E}$ is convex. Then for all α, $0 \leq \alpha \leq 1$, by the definition of a convex function, we have

$$\mathcal{E}(\alpha z + (1 - \alpha)y) \leq \alpha\mathcal{E}(z) + (1 - \alpha)\mathcal{E}(y) \tag{3.16}$$

which implies that

$$\frac{\mathcal{E}(\alpha z + (1 - \alpha)y) - \mathcal{E}(y)}{\alpha} \leq \mathcal{E}(z) - \mathcal{E}(y). \tag{3.17}$$

By the mean value theorem, for a function $\mathcal{E}$ is differentiable on interval (y, z), there is a $\epsilon = \alpha z + (1 - \alpha)y$ such that $\mathcal{E}(z) - \mathcal{E}(y) = (z - y)^T\nabla\mathcal{E}(\epsilon)$. Hence, the left-hand side of the inequality equals

$$\frac{\mathcal{E}(\alpha z + (1 - \alpha)y) - \mathcal{E}(y)}{\alpha} = (z - y)^T\nabla\mathcal{E}(\epsilon). \tag{3.18}$$

As $\alpha \to 0$, $\epsilon \to y$, we have

$$(z - y)\nabla\mathcal{E}(y) \leq \mathcal{E}(z) - \mathcal{E}(y). \tag{3.19}$$

Conversely, suppose that (3.15) holds for all $y, z \in \mathbb{U}_0$. Setting $u = \alpha z + (1 - \alpha)y$ for $0 \leq \alpha \leq 1$, we obtain

$$(z - u)\nabla\mathcal{E}(u) \leq \mathcal{E}(z) - \mathcal{E}(u) \tag{3.20}$$

and

$$(y - u)\nabla\mathcal{E}(u) \leq \mathcal{E}(y) - \mathcal{E}(u). \tag{3.21}$$

Multiplying (3.20) by α and (3.21) by $1-\alpha$ respectively, and sum them together

$$\begin{aligned}\mathcal{E}(u) + (\alpha z + (1-\alpha)y - u)^T \nabla \mathcal{E}(u) \\ \leq \alpha \mathcal{E}(z) + (1-\alpha)\mathcal{E}(y).\end{aligned} \tag{3.22}$$

Since $u = \alpha z + (1-\alpha)y$, we have

$$\mathcal{E}(\alpha z + (1-\alpha)y) \leq \mathcal{E}(z) + (1-\alpha)\mathcal{E}(y) \tag{3.23}$$

for all α, $0 \leq \alpha \leq 1$, which implies $\mathcal{E}$ is convex. Hence, the lemma is proven.

Lemma 3.3. *The optimal solution $u^* = (\tilde{\mathbf{x}}^*, y^*)$ satisfy that*

$$(u^* - u)^T \nabla \mathcal{E}(u) \leq -\mathcal{E}(u) \tag{3.24}$$

Proof: Since $\mathcal{E}(u) > 0$ is continuous differentiable for all u and convex, from Lemma 3.2, we have that $(u^* - u)^T \nabla \mathcal{E}(u) \leq \mathcal{E}(u^*) - \mathcal{E}(u)$, and the energy is all dissipated at the optimal point $\mathcal{E}(u^*) = 0$, we have the conclusion of Lemma 3.3.

3.3.4 Stability Analysis

Lemma 3.4. *The recurrent neural network which defined in (3.10)-(3.11) is globally stable and convergent to the optimal solution of the convex optimization problem*

Proof: Without loss of generality, let the positive scalar μ in the dynamic equation of recurrent neural network to be 1. Assume $u = (\tilde{\mathbf{x}}, y)$ and $u^* = (\tilde{\mathbf{x}}^*, y^*)$ is the KKT point. We define the Lyapunov function as

$$S(u) = \frac{1}{2} \left\| u^*(t) - u(t) \right\|_2^2. \tag{3.25}$$

Apparently, the Lyapunov function is zero at the optimal point $S(u^*) = 0$. *According to the chain rule, we have*

$$\begin{aligned} \frac{dS}{dt} &= \frac{dS}{du}\left(\frac{du}{dt}\right) \\ &= -(u^*(t) - u(t))^T \frac{du}{dt} \\ &= (u^*(t) - u(t))^T \nabla F(u) \\ &= (u^*(t) - u(t))^T \begin{bmatrix} \nabla \mathcal{E}_{\tilde{\mathbf{x}}}(u) \\ \nabla \mathcal{E}_y(u) + \lambda \end{bmatrix} \\ &= (u^*(t) - u(t))^T \nabla \mathcal{E}(u) + \lambda(y^*(t) - y(t)) \end{aligned} \tag{3.26}$$

To simplify the description, we neglect t here. According to Lemma 3.3, we have

$$\frac{dS}{dt} \leq -\mathcal{E}(u) + \lambda(y^* - y). \tag{3.27}$$

Therefore, we can obtain

$$\frac{dS}{dt} \leq \; -\lambda\left([\tilde{\mathbf{x}}^H \odot \tilde{\mathbf{x}} - \mathbf{e}^T y]^+ + y - y^*\right) - \|\mathbf{s} - \mathcal{H}\tilde{\mathbf{x}}\|_2^2 .$$

Since we have

$$[\tilde{\mathbf{x}}^H \odot \tilde{\mathbf{x}} - \mathbf{e}^T y]^+ + y = \max(\tilde{\mathbf{x}}^H \odot \tilde{\mathbf{x}}, \mathbf{e}^T y) \tag{3.28}$$

and the optimal recurrent neural network activation factor y^* *satisfy that*

$$\max(\tilde{\mathbf{x}}^H \odot \tilde{\mathbf{x}}, \mathbf{e}^T y) \geq y^*. \tag{3.29}$$

We can conclude that

$$\frac{dS}{dt} \leq 0. \tag{3.30}$$

Since $S(u) = \|u^*(t) - u(t)\|_2^2/2$ *is positive definite and radially unbounded, there exists a convergence subsequence* $\{u(t_k)\}$ *for any initial point* $u_0 \in \mathbb{U}$ *that* $\lim_{k\to\infty} u(t_k) = \hat{u}$*, where*

$$\frac{dS(\hat{u})}{dt} = 0 \tag{3.31}$$

Corollary 3.1. *We have unique optimal recurrent neural network activation variable* y *if* $dS(\hat{u})/dt = 0$ *is satisfied.*

Proof: See Appendix D for detail.

Lemma 3.5. *The convergence* $\lim_{k\to\infty} u(t_k) = \hat{u}$ *is satisfied if and only if* $\hat{u}$ *is the KKT point.*

Proof: Let's define the Lyapunov function as

$$\hat{S}(u(t)) = \frac{1}{2}\|\hat{u}(t) - u(t)\|_2^2 \tag{3.32}$$

where $\hat{u} = (\hat{\hat{\mathbf{x}}}, \hat{y})$ *is the KKT point.*
Then, $\hat{S}(\hat{u}) = 0$ *and we can get*

$$\lim_{k\to\infty} S(u(t_k)) = \hat{S}(\hat{u}) = 0 \tag{3.33}$$

$\forall \epsilon > 0$*, there* $\exists\, r > 0$ *such that* $\forall t_k > t_r$ *we have*

$$\hat{S}(u(t_k)) = \frac{1}{2}\|\hat{u}(t) - u(t_k)\|_2^2 < \epsilon. \tag{3.34}$$

According to the property of negative Lyapunov function derivative which derived in (3.30), *we have*

$$\frac{d}{dt}\hat{S}(u(t)) \leq 0. \tag{3.35}$$

Therefore, $\forall t > t_r$

$$\begin{aligned} \|\hat{u}(t) - u(t)\|_2^2 &= 2\hat{S}(u(t)) \\ &\leq 2\hat{S}(u(t_r)) \leq 2\epsilon \end{aligned} \tag{3.36}$$

We can then conclude that

$$\lim_{t\to\infty} \|\hat{u}(t) - u(t)\|_2^2 = 0 \tag{3.37}$$

and hence

$$\lim_{t\to\infty} u(t) = \hat{u}(t). \tag{3.38}$$

Therefore, the proposed neural network proposed in (3.10) and (3.11) is globally stable and converges to the KKT point.

In this section, we proposed a neural network solution for the PAPR-aware downlink massive MIMO-OFDM problem. The stability and convergence of the proposed structure are analyzed and proved in Lemma 3.4. In addition, we also proved that the PA efficiency can reach an unique optimal value, and the uniqueness of the optimal PA efficiency is guaranteed in Corollary 3.1.

3.4 Simulation Results

The simulation results are presented in this section to validate our analysis and the proposed algorithms for a small cell architecture. An isolated cell is assumed to be circular with the radius of R to simplify the analysis, but could be extended to irregular shape small cell with adjacent cell interference

3.4.1 Simulation Parameters

The recurrent neural network based large scale MU-MIMO-OFDM downlink system with 128 subcarriers, 120 antennas equipped in BS and 12 single antenna MSs is adopted in this work to evaluate the performance of PAPR reduction with limited influence to BER performance. The modulation is 16-QAM and the MSs are randomly distributed within the range of small cell. The BS or center unit locates in the pole which is the center of the small cell circle. The channels between each antenna and user are modeled as frequency selective channels and are independent from one another.

Table 3.1: Simulation parameters of the PAPR-aware massive MIMO system

Cell radius R	2 km
Reference distance	35 m
Carrier center frequency	3.55 GHz
Number of subcarriers	128
Subcarrier spacing	15kHz
Number of Antennas in BS	120
Number of MS	12
MS placement	Uniform
Modulation	16-QAM
Small scale fading distribution	Rayleigh fading
Noise power per subcarrier	-131 dBm
Oversampling factor L	4
Regulation factor λ	0.4

The simulation parameters can be found in Table 3.1, the recurrent neural network is simulated in this section is Algorithm 4 and is used to derive the optimal transmitted signal and activation variable. The PAPR reduction performance of different PAPR-aware algorithm in time domain is shown in Figure 3.4. The least square (LS) method is also shown as a baseline. The LS algorithm is one of the most prominent precoding method in MIMO system, and generates the transmitted signal with minimum l-2 norm while perfectly remove all the MUI. It is shown in Figure 3.4 that the LS algorithm gives a solution with less power but the PAPR reduction performance is clearly worse than the other approaches. The FITRA and the proposed RNN approach have similar performance in terms of PAPR reduction, while the maximum values of both methods are constant over time. On the other hand, the performance of MUI is also critical in massive MU-MIMO system, especially since the trade-off between the PAPR reduction and MUI is adopted in both approaches. The MUI performance is compared in terms of bit error rate (BER) in the rest of this chapter.

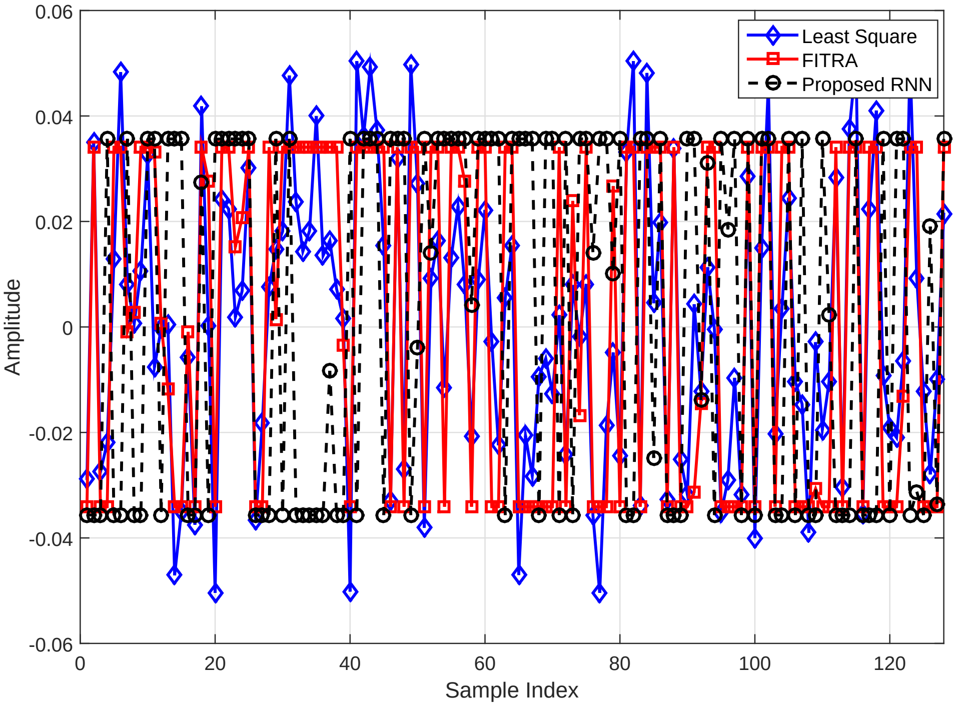


Figure 3.4: Signal waveform for different PAPR-aware precoding approaches in time domain.

3.4.2 Fast Iterative Truncation Algorithm (FITRA)

The FITRA method is simulated in this section especially to compare the convergence speed with the proposed RNN method. As shown in [2], FITRA gives a convergence with the rate of $O(1/k^2)$ as shown follows

$$F(\mathbf{x}_k) - F(\mathbf{x}^*) \leq \frac{2L \left\| \mathbf{x}_0 - \mathbf{x}^* \right\|_2^2}{(k+1)^2} \tag{3.39}$$

The performance of convergence rate of FITRA and proposed recurrent neural network are compared in Figure 3.5, it shows that the result of FITRA matches the theoretical convergence performance of $O(1/k^2)$. The FITRA takes more than 400 iterations to reach the optimal result. Thus for the time-varying channels, the slow convergence speed and high computational complexity algorithm rendering it less desirable for real-time implementation. Comparing the FITRA method in [2] with our proposed approach using recurrent neural network, it shows our new approach with near exponential convergence as shown in Figure 3.5, and only needs around 100 iterations comparing to about 300 iteration for the FITRA method.

The impact of antenna configuration to PAPR reduction performance is shown in Figure 3.6. The increasing number of antenna in BS yields better performance for both least square and proposed RNN method. Our approach outperform the least square method at each antenna configuration in the simulation.

3.5 Conclusion

A PAPR-aware massive MU-MIMO-OFDM system based on recurrent neural network is investigated in this chapter. We also proved the stability and convergence of the proposed

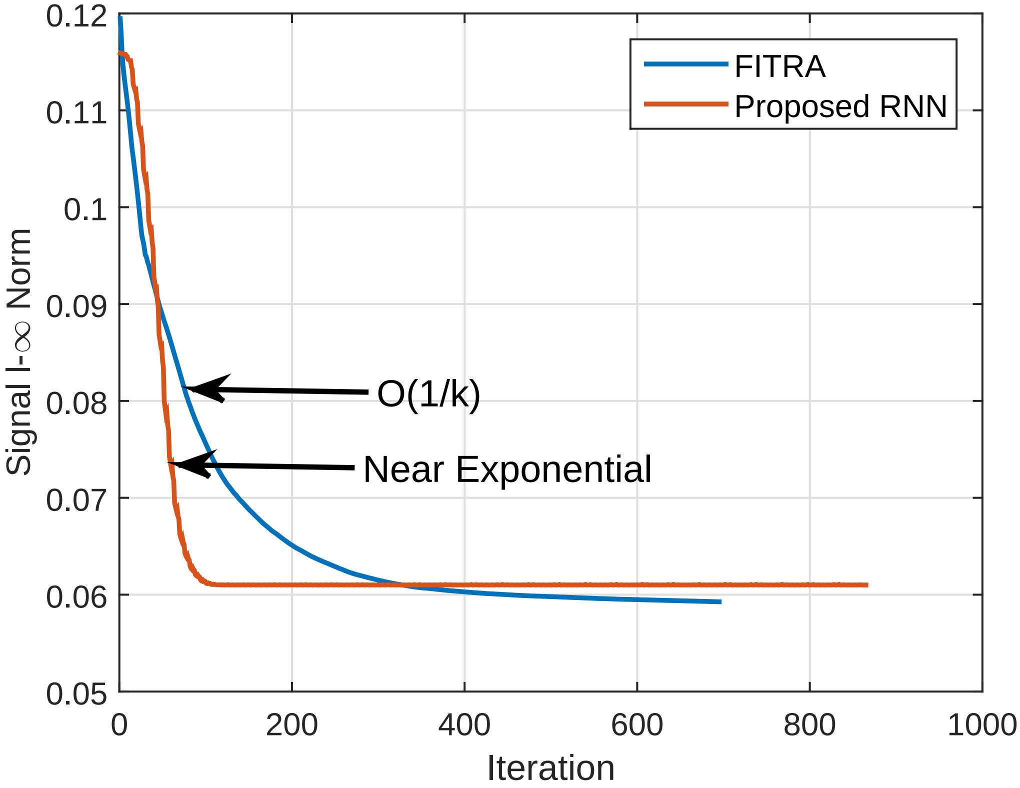


Figure 3.5: Convergence speed comparison for different PAPR-aware precoding approaches.

approach and the simulation results indicates that the convergence speed is much faster than the existing non-AI solution.

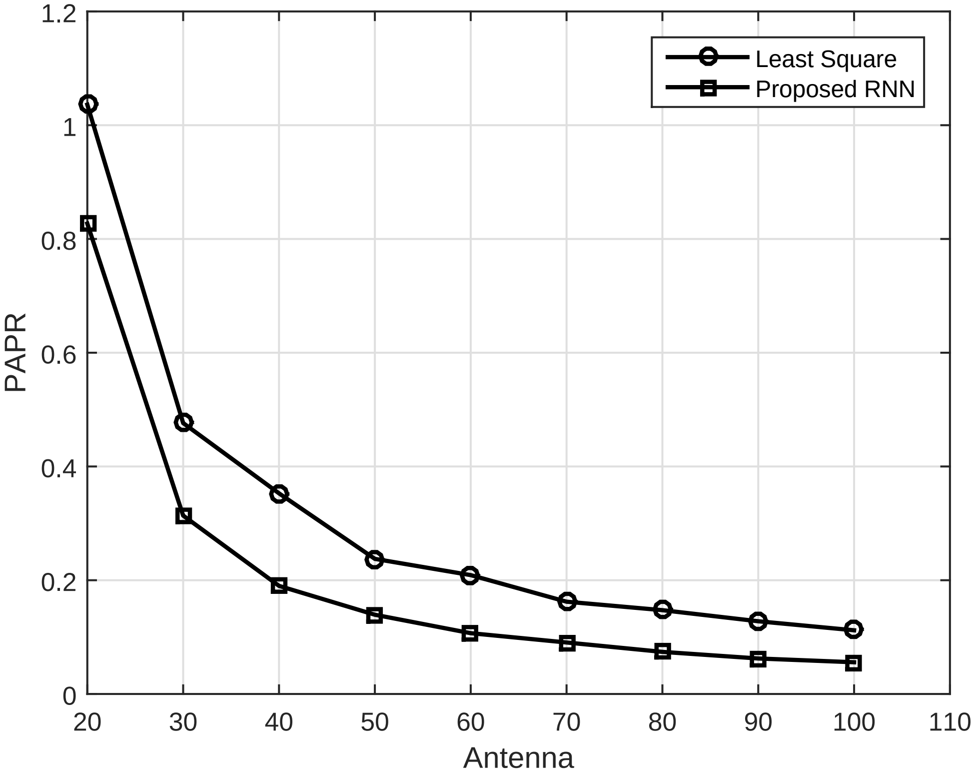


Figure 3.6: PAPR reduction performance of proposed RNN and least square precoding depends on number of transmit antennas.

Appendix C

Proof of Lemma 3.1

The gradient of $\mathcal{E}_\eta = \|\mathbf{s} - \mathcal{H}\tilde{\mathbf{x}}\|_2^2$ with respect to $\tilde{\mathbf{x}}$ can be derived as follows:

$$\begin{aligned}
\frac{\partial \mathcal{E}_\eta}{\partial \tilde{\mathbf{x}}} &= \frac{\partial \mathrm{tr}\left((\mathbf{s} - \mathcal{H}\tilde{\mathbf{x}})^H(\mathbf{s} - \mathcal{H}\tilde{\mathbf{x}})\right)}{\partial \tilde{\mathbf{x}}} \\
&= \frac{\partial \mathrm{tr}\left(\mathbf{s}^H\mathbf{s} - \tilde{\mathbf{x}}^H\mathcal{H}^H\mathbf{s} - \mathbf{s}^H\mathcal{H}\tilde{\mathbf{x}} + \tilde{\mathbf{x}}^H\mathcal{H}^H\mathcal{H}\tilde{\mathbf{x}}\right)}{\partial \tilde{\mathbf{x}}} \\
&= \mathcal{H}^H(\mathcal{H}\tilde{\mathbf{x}} - \mathbf{s}) + (\mathcal{H}\tilde{\mathbf{x}} - \mathbf{s})^H\mathcal{H} \\
&= 2\mathrm{Re}\left(\mathcal{H}^H(\mathcal{H}\tilde{\mathbf{x}} - \mathbf{s})\right)
\end{aligned}$$

The proof is completed.

Appendix D

Proof of Corollary 3.1

First, according to (3.31), there exists a convergence subsequence $\{u(t_k)\}$ for any initial point $u_0 \in \mathbb{U}$ so that $\lim_{k\to\infty} u(t_k) = \hat{u}$, where

$$\frac{dS(\hat{u})}{dt} = 0$$

which leads to

$$\begin{aligned} -\lambda\left([\tilde{\mathbf{x}}^H \odot \tilde{\mathbf{x}} - \mathbf{e}^T y]^+ + y - y^*\right) \\ - \|\mathbf{s} - \mathcal{H}\tilde{\mathbf{x}}\|_2^2 \geq 0 \end{aligned}$$

Therefore, $dS(\hat{u})/dt = 0$ if and only if

$$\begin{cases} \lambda\left([\hat{\tilde{\mathbf{x}}}^H \odot \hat{\tilde{\mathbf{x}}} - \mathbf{e}^T \hat{y}]^+ + \hat{y} - y^*\right) = 0 \\ \left\|\mathbf{s} - \mathcal{H}\hat{\tilde{\mathbf{x}}}\right\|_2^2 = 0 \end{cases} \tag{3.40}$$

Similarly, by using the symmetry principle, we also have

$$\begin{cases} \lambda\left([\tilde{\mathbf{x}}^{*H} \odot \tilde{\mathbf{x}}^* - \mathbf{e}^T y^*]^+ + y^* - \hat{y}\right) = 0 \\ \|\mathbf{s} - \mathcal{H}\tilde{\mathbf{x}}^*\|_2^2 = 0 \end{cases} \tag{3.41}$$

According to the nonnegativity of function $[\cdot]^+$, and from (3.40) and (3.41), we can obtain that

$$
\begin{aligned}
&[\hat{\tilde{\mathbf{x}}}^H \odot \hat{\tilde{\mathbf{x}}} - \mathbf{e}^T \hat{y}]^+ \cdot [\tilde{\mathbf{x}}^{*H} \odot \tilde{\mathbf{x}}^* - \mathbf{e}^T y^*]^+ \\
&= \mathbf{e}^T (y^* - \hat{y})(\hat{y} - y^*) \\
&= -\mathbf{e}^T \left\| \hat{y} - y^* \right\|_2^2 \geq 0
\end{aligned}
$$

As a result, $\|\hat{y} - y^*\|_2^2 = 0$ and $\hat{y} = y^*$ the optimal recurrent neural network activation variable is unique, and thus the optimal PA efficiency is unique. The proof is completed.

Chapter 4

Semidefinite Relaxation-Based PAPR-Aware Precoding for Massive MIMO-OFDM Systems

4.1 Introduction

There are two main kinds of PAPR-aware precoding scheme exploiting excess DoFs in the massive MIMO systems. One scheme is the constant envelope precoding [87–90], which has a flat amplitude transmit signal by using the phases of the transmitting signals to carry information. The authors in [87–89] propose MU precoding scheme for per-antenna constant envelope transmission to enable efficient implementation with nonlinear RF components. Moreover, the single-user constant envelope precoder in [90] is realized by unequal per-antenna power allocation. While CE precoding provides an attractive signal processing method to manage PA efficiency and reduce the implementation costs of the PA, it is proposed for the single carrier systems and not compatible with current OFDM waveform based cellular standards. The other scheme is PAPR-aware precoding in massive MIMO-OFDM system [2, 92, 93], which minimizes the combination of the MUI and the peak power of the signal. In this work, we take both the peak and average power of the transmit signal into account for the optimization.

Moreover, it is impractical to assume perfect CSI at the BS when channel uncertainty is presented. The imperfect CSI could be caused by inaccurate channel estimation, quantitization errors or the offsets between reciprocal channels in terms of time or frequency. The robust optimization of wireless communication systems has been extensively studied with imperfect channel knowledge [57, 94–102]. The imperfect CSI can be modeled by either deterministic or stochastic approaches. The deterministic robust MIMO precoding was studied for broadcasting channels [94], multi-cell systems [95] [96], and cognitive radio systems [97] [98] [99]. The stochastic robust precoding was studied in [57, 100–102].

Multicell processing based on the cooperation between BSs has emerged as a promising solution in recent years [103]. The optimal intercell coordination requires coherence between the signals from different BSs and the transmissions are implemented in a centralized manner. The coordination between neighboring BSs has been shown to be able to improve the capacity gain [104–109]. This work explores various formulations of the PAPR-aware precoding problem in a massive MIMO context including the practically important cases of imperfect CSI and intercell coordination.

4.2 System Model and Problem Formulation

The purpose of the PAPR-aware massive MIMO-OFDM downlink precoding design is to find transmit signal to meet the predetermined PAPR and MUI targets. In order to quantify the MUI, the overall massive MIMO-OFDM downlink precoding constraint need to be formulated. Consider the downlink massive MIMO system of Fig. 4.1 which has M_r single antenna users and one BS which equipped with N_t antennas. The number of BS antennas is significantly larger than number of simultaneous users, $N_t \gg M_r$. The overall multicarrier frequency-space precoding in Chapter 3 is rewritten here

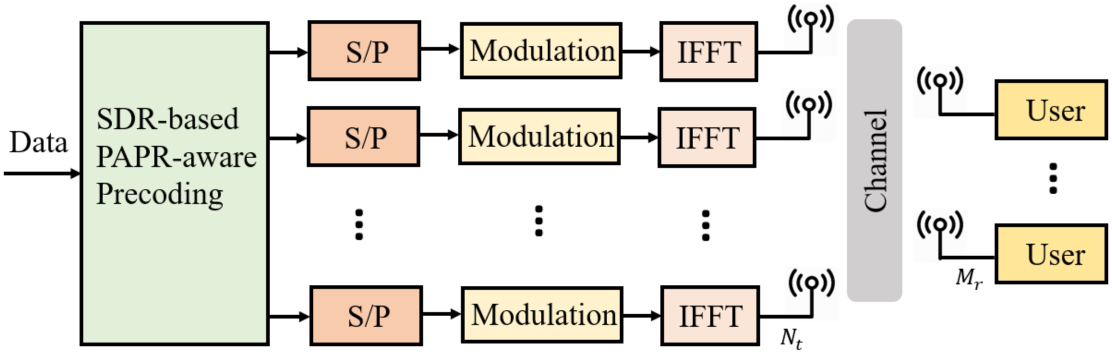


Figure 4.1: System diagram of the proposed PAPR-aware downlink massive MIMO-OFDM system based on SDR.

$$\mathbf{s} = \underbrace{\begin{bmatrix} \mathbf{H}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{H}_{N_c} \end{bmatrix}}_{\triangleq \mathbf{H}} \underbrace{\begin{bmatrix} \mathbf{I}_{1,1} & \mathbf{I}_{1,2} & \dots & \mathbf{I}_{1,N_t} \\ \mathbf{I}_{2,1} & \mathbf{I}_{2,2} & \dots & \mathbf{I}_{2,N_t} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{I}_{N_c,1} & \mathbf{I}_{N_c,2} & \dots & \mathbf{I}_{N_c,N_t} \end{bmatrix}}_{\triangleq \bar{\mathbf{P}}} \underbrace{\begin{bmatrix} \mathbf{Q} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{Q} \end{bmatrix}}_{\triangleq \bar{\mathbf{Q}}} \underbrace{\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_{N_t} \end{pmatrix}}_{\mathbf{x}}, \tag{4.1}$$

where $\mathbf{s}$ represents the multicarrier information symbols, $\mathbf{H}_m$ the $M_r \times N_t$ channel matrix of mth subcarrier, $\bar{\mathbf{P}}$ the permutation matrix which is comprised by the $N_t \times N_c$ matrix $\mathbf{I}_{m,n}$ (the (n,m) entry is 1 for $1 \le m \le N_c$ and $1 \le n \le N_t$, and the other entries of $\mathbf{I}_{m,n}$ are 0s), $\bar{\mathbf{Q}}$ the $N_x \times N_x$ diagonalization matrix comprised by FFT matrix $\mathbf{Q}$, where $N_x \triangleq N_t N_c$, $\mathbf{x}$ the time-domain transmit signal. Note that the permutation matrix satisfies $\bar{\mathbf{P}}\bar{\mathbf{P}}^H = \mathbf{I}_{N_x}$

and $\bar{\mathbf{P}}^H$ can be considered as the reverse permutation matrix. Thus we have

$$\mathbf{s} = \mathbf{H}\bar{\mathbf{P}}\bar{\mathbf{Q}}\mathbf{x} = \tilde{\mathbf{H}}\mathbf{x}, \tag{4.2}$$

where $\tilde{\mathbf{H}} = \mathbf{H}\bar{\mathbf{P}}\bar{\mathbf{Q}}$ represents the mapping operation of the time domain transmit signal to the information symbols of the overall massive MIMO system.

4.2.1 QCQP Problem Formulation With Perfect CSI

The optimization problem of transmit power minimization subject to PAPR and MUI constraints is formulated as QCQP problem in this subsection. We begin with the assumption of perfect CSI which will serve as a stepping stone towards the more realistic robust PAPR-aware precoding with imperfect CSI presented in the next section. The energy efficiency not only depends on the transmit power but also the PAPR [110]. To derive an energy-efficient PAPR-aware precoding solution, we follow the design principles below:

Principle 1: The transmit power should be reduced as much as possible, but still provides the required signal-to-noise ratio (SNR) at the receiver. Lower transmit power reduces power consumption of BS and better facilitates linear operation of the PA and hence lower out-of-band emissions.

Principle 2: Instead of minimizing the PAPR we prefer to set practical targets of PAPR for each PA since different PAPR ranges are allowed for different classes of PAs (e.g. class A, class AB, etc).

Principle 3: Due to the hardware impairments (e.g. amplifier nonlinearity, I/Q-imbalance, phase noise, quantization errors, etc), CSI uncertainties, and thermal noises [111], a small amount of MUI will not significantly degrade the performance of massive MIMO-OFDM downlink precoding.

Based on these principles, our optimization strategy is to minimize the transmit power while keeping the PAPR of each antenna and MUI below the predetermined thresholds $\alpha_m (1 \le m \le N_t)$ and δ_e, respectively. Therefore, the problem can be formulated as:

$$\begin{aligned} (\mathbf{P1}) : \min_{\mathbf{x}} \quad & \|\mathbf{x}\|_2^2 && (4.3a) \\ \text{subject to} \quad & \frac{\|\mathbf{F}_m\mathbf{x}\|_\infty^2}{\|\mathbf{F}_m\mathbf{x}\|_2^2} \le \frac{\alpha_m}{N_c}, \qquad \forall m \in \{1, \cdots, N_t\} && (4.3b) \\ & \left\|\mathbf{s} - \tilde{\mathbf{H}}\mathbf{x}\right\|_2^2 \le \delta_e, && (4.3c) \end{aligned}$$

where $\mathbf{x} = [\mathbf{x}_1^H \ \cdots \ \mathbf{x}_{N_t}^H]^H$, and $\mathbf{x}_m = \mathbf{F}_m\mathbf{x}$.[1] The MUI constraint of (4.3c) is derived from (4.2) when the precoding error δ_e is allowed. Note while the objective function of power minimization in (4.3a) and the MUI constraint in (4.3c) are convex, the set of PAPR constraints in (4.3b) are nonconvex. Although the constraints in (4.3b) can be relaxed to convex form by neglecting the average power [2][2], it leads to a sub-optimal solution of the PAPR-aware precoding problem. Instead of the sub-optimal convex relaxation in [2, 93], we convert (4.3b) to quadratic form:

$$\mathbf{x}^H\mathbf{F}_m^H \left(\mathbf{Y}_i - \frac{\alpha_m}{N_c}\mathbf{I}_{N_c}\right) \mathbf{F}_m\mathbf{x} \le 0, \forall i \in \{1, \cdots, N_c\}, m \in \{1, \cdots, N_t\}, \tag{4.4}$$

where $\mathbf{Y}_i \in \mathbb{R}^{N_c}$ is defined as

$$\mathbf{Y}_i(i, j) = \begin{cases} 1, j = i \\ 0, j \ne i \end{cases} . \tag{4.5}$$

---

[1] $\mathbf{F}_m = [\underbrace{\mathbf{0}_{N_c}, ..., \mathbf{I}_{N_c}, ..., \mathbf{0}_{N_c}}_{N_t \text{ submatrices}}]$, $\mathbf{I}_{N_c}$ is the mth submatrix of $\mathbf{F}_m$, the other submatrices of $\mathbf{F}_m$ are $\mathbf{0}_{N_c}$.

[2] In [2], constraint (4.3b) is relaxed as $\|\mathbf{F}_m\mathbf{x}\|_\infty^2 \le \beta_m, \forall m \in \{1, 2, ..., N_t\}$, which is convex.

It can be further simplified as

$$\mathbf{x}^H \mathbf{Q}_{im} \mathbf{x} \leq 0, \forall i, m, \tag{4.6}$$

by defining $\mathbf{Q}_{im} = \mathbf{F}_m^H \left(\mathbf{Y}_i - \frac{\alpha_m}{N_c} \mathbf{I}_{N_c} \right) \mathbf{F}_m$. Similarly, the constraints of (4.3c) can be reformulated as

$$\begin{bmatrix} \mathbf{x}^H & t \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{H}}^H \tilde{\mathbf{H}} & -\tilde{\mathbf{H}}^H \mathbf{s} \\ -\mathbf{s}^H \tilde{\mathbf{H}} & \mathbf{s}^H \mathbf{s} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ t \end{bmatrix} \leq \delta_e, \tag{4.7}$$

where $t = 1$.

By defining $\tilde{\mathbf{x}} \triangleq \begin{bmatrix} \mathbf{x} \\ t \end{bmatrix}$ and $\mathbf{G} \triangleq \begin{bmatrix} \tilde{\mathbf{H}}^H \tilde{\mathbf{H}} & -\tilde{\mathbf{H}}^H \mathbf{s} \\ -\mathbf{s}^H \tilde{\mathbf{H}} & \mathbf{s}^H \mathbf{s} \end{bmatrix}$, the constraints of (4.3c) can be reduced to

$$\tilde{\mathbf{x}}^H \mathbf{G} \tilde{\mathbf{x}} \leq \delta_e. \tag{4.8}$$

By combining (4.6) and (4.8), the optimization problem **P1** is equivalent to

$$\begin{aligned} (\mathbf{P2}) : \min_{\tilde{\mathbf{x}}} \quad & \|\tilde{\mathbf{x}}\|_2^2 && (4.9a) \\ \text{subject to} \quad & \operatorname{Tr}(\tilde{\mathbf{x}}^H \tilde{\mathbf{Q}}_{im} \tilde{\mathbf{x}}) \leq 0, \forall i, m && (4.9b) \\ & \operatorname{Tr}(\tilde{\mathbf{x}}^H \mathbf{G} \tilde{\mathbf{x}}) \leq \delta_e && (4.9c) \\ & \tilde{x}_{N_x+1} = 1, && (4.9d) \end{aligned}$$

where $\tilde{\mathbf{Q}}_{im} \triangleq \left[\begin{array}{c:c} \mathbf{Q}_{im} & \mathbf{0}_{N_x \times 1} \\ \hdashline \mathbf{0}_{1 \times N_x} & 0 \end{array}\right]$, $\tilde{x}_{N_x+1} = 1$ indicates that the last entry of $\tilde{\mathbf{x}}$ is 1. It is a nonconvex QCQP problem and can be relaxed to apply computationally efficient semidefinite programming (SDP) solutions. Note that the relationship between the objectives of **P1** and **P2** can be written as $\|\tilde{\mathbf{x}}\|_2^2 = \|\mathbf{x}\|_2^2 + 1$ and the solutions of **P1** and **P2** as $\tilde{\mathbf{x}}_{opt} = \begin{bmatrix} \mathbf{x}_{opt} \\ 1 \end{bmatrix}$.

4.3 Relaxed PAPR-Aware Massive MIMO Precoding by Semidefinite Programming

In order to solve the derived nonconvex problems, one of the most common approaches is relaxing the nonconvex constraints to obtain a convex problem that approximates the original problem [49]. This section shows that the derived QCQP for PAPR-aware massive MIMO precoding can be solved by the SDR method and then applying the randomization method for rank reduction.

4.3.1 PAPR-Aware Massive MIMO Precoding Relaxation

To derive the SDR of **P2**, a necessary step is to apply $\|\tilde{\mathbf{x}}\|_2^2 = \mathrm{Tr}(\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H)$, $\mathrm{Tr}(\tilde{\mathbf{x}}^H\tilde{\mathbf{Q}}_{im}\tilde{\mathbf{x}}) = \mathrm{Tr}(\tilde{\mathbf{Q}}_{im}\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H)$ and $\mathrm{Tr}(\tilde{\mathbf{x}}^H\mathbf{G}\tilde{\mathbf{x}}) = \mathrm{Tr}(\mathbf{G}\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H)$. In particular, all the objective and constraints in **P2** are linear to $\bar{\mathbf{x}}\tilde{\mathbf{x}}^H$ after applying the SDR. As a result, we define a new variable as $\tilde{\mathbf{X}} \triangleq \tilde{\mathbf{x}}\tilde{\mathbf{x}}^H$. Therefore, $\tilde{\mathbf{X}}$ is a rank-1 symmetric positive semidefinite (PSD) matrix represented by $\mathrm{rank}(\tilde{\mathbf{X}}) = 1$ and $\tilde{\mathbf{X}} \succeq 0$. The optimization problem **P2** is equivalent to

$$
\begin{aligned}
(\mathbf{P3}) : \min_{\tilde{\mathbf{X}}} \quad & \mathrm{Tr}(\tilde{\mathbf{X}}) && (4.10a)\\
\text{subject to} \quad & \mathrm{Tr}(\tilde{\mathbf{Q}}_{im}\tilde{\mathbf{X}}) \leq 0, \forall i, m && (4.10b)\\
& \mathrm{Tr}(\mathbf{G}\tilde{\mathbf{X}}) \leq \delta_e && (4.10c)\\
& \mathrm{Tr}(\mathbf{O}_x\tilde{\mathbf{X}}) = 1 && (4.10d)\\
& \tilde{\mathbf{X}} \succeq 0 && (4.10e)\\
& \mathrm{rank}(\tilde{\mathbf{X}}) = 1, && (4.10f)
\end{aligned}
$$

where $\mathbf{O}_x = \mathbf{o}_x\mathbf{o}_x^H$, $\mathbf{o}_x = [0, \cdots, 0, 1]^H \in \mathbb{R}^{(N_x+1)\times 1}$, and therefore $\mathrm{Tr}(\mathbf{O}_x\tilde{\mathbf{X}}) = \mathrm{Tr}(\mathbf{o}_x^H\tilde{\mathbf{X}}\mathbf{o}_x)$. Notice that the objective function in **P3** is the overall transmit power of the massive MIMO system, it can be easily converted to a per-antenna power minimization problem when considering the fairness between transmit antennas [112]. However, the reformulation in **P3** is just as difficult to solve as problem **P2** since the rank constraint (4.10f) is nonconvex although the objective function and all other constraints are convex in $\tilde{\mathbf{X}}$. By dropping the rank constraint we can obtain the following SDR version of **P2**:

$$\begin{aligned} (\mathbf{P4}) : \min_{\tilde{\mathbf{X}}} \quad & \mathrm{Tr}(\tilde{\mathbf{X}}) \\ \text{subject to} \quad & (4.10b) - (4.10e). \end{aligned}$$

The SDR problem in **P4** is convex and thus does not suffer from local minima. The optimization problem is relaxed and the optimal objective value of **P4** is always less or equal to the value of **P3**,

$$\mathrm{Tr}(\tilde{\mathbf{X}}_{opt}) \leq \mathrm{Tr}(\tilde{\mathbf{X}}'_{opt}), \tag{4.11}$$

where $\tilde{\mathbf{X}}_{opt}$ is the optimal value of **P4** and $\tilde{\mathbf{X}}'_{opt}$ the optimal value of **P3**. The equality holds when there exists a rank 1 optimal solution $\tilde{\mathbf{X}}_{opt} = \tilde{\mathbf{x}}_{opt}\tilde{\mathbf{x}}_{opt}^H$. There is a corresponding convex Lagrange dual problem for any given problem (primal problem), which yields a bound on the optimal value of the primal problem. The bound is tightest and the strong duality holds when the primal problem is convex. The robust extensions of **P4**, which accounts for imperfect channel knowledge based on bounded-error and statistical models, are discussed in Section 4.4.

4.3.2 Duality of the SDR

A SDR problem can be solved by employing the primal-dual path following algorithm, which has polynomial complexity [62]. Consider the Lagrange dual problem of **P2**, which can be written as

$$\begin{aligned}
&L(\boldsymbol{\lambda}, \nu_1, \nu_2, \mathbf{x})\\
=&\tilde{\mathbf{x}}^H\tilde{\mathbf{x}} + \sum_{i=1}^{N_c}\sum_{m=1}^{N_t} \lambda_{im}\tilde{\mathbf{x}}^H\tilde{\mathbf{Q}}_{im}\tilde{\mathbf{x}} + \nu_1(\tilde{\mathbf{x}}^H\mathbf{G}\tilde{\mathbf{x}} - \delta_e) - \nu_2(\tilde{\mathbf{x}}^H\mathbf{O}_x\tilde{\mathbf{x}} - 1)\\
=&\tilde{\mathbf{x}}^H\left(\mathbf{I}_{N_x+1} + \sum_{i=1}^{N_c}\sum_{m=1}^{N_t} \lambda_{im}\tilde{\mathbf{Q}}_{im} + \nu_1\mathbf{G} - \nu_2\mathbf{O}_x\right)\tilde{\mathbf{x}} + \nu_2 - \nu_1\delta_e,
\end{aligned}$$

where $\boldsymbol{\lambda}, \nu_1, \nu_2$ are non-negative Lagrangian multipliers. The Hessian matrix of the Lagrangian of the problem is derived as

$$\nabla_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H} = \mathbf{I}_{N_x+1} + \sum_{i=1}^{N_c}\sum_{m=1}^{N_t} \lambda_{im}\tilde{\mathbf{Q}}_{im} + \nu_1\mathbf{G} - \nu_2\mathbf{O}_x. \tag{4.12}$$

Therefore, both the dual problems of **P2** and **P4** are the same and can be written as

$$(\mathbf{P5}) : \max_{\mathbf{V}, \boldsymbol{\lambda}, \nu_1, \nu_2} \quad \nu_2 - \nu_1\delta_e \tag{4.13a}$$

$$\text{subject to} \quad \mathbf{V} = \mathbf{I}_{N_x+1} + \sum_{i=1}^{N_c}\sum_{m=1}^{N_t} \lambda_{im}\tilde{\mathbf{Q}}_{im} + \nu_1\mathbf{G} - \nu_2\mathbf{O}_x \succeq \mathbf{0}. \tag{4.13b}$$

It is clear that both **P4** and **P5** are convex and that the strong duality holds for **P4**. As a result, we have

$$\mathrm{Tr}(\tilde{\mathbf{X}}_{opt}) = \nu_2^* - \nu_1^*\delta_e, \tag{4.14}$$

where $\tilde{\mathbf{X}}_{opt}$ is the optimal solution of **P4**, and (ν_1^*, ν_2^*) is the optimal solution of **P5**. It is shown in (4.14) that given a larger allowable precoding error δ_e, a lower optimal transmit power $\mathrm{Tr}(\tilde{\mathbf{X}}_{opt})$ can be achieved. Intuitively, lower allowable precoding error means smaller feasible optimization solution region. Such a solution may increase the transmit power, but achieves less MUI at the receivers. Upon defining $\bar{\mathbf{X}} \triangleq \begin{bmatrix} \Re\{\tilde{\mathbf{X}}\} & -\Im\{\tilde{\mathbf{X}}\} \\ -\Im\{\tilde{\mathbf{X}}\} & \Re\{\tilde{\mathbf{X}}\} \end{bmatrix}$, $\bar{\mathbf{Q}}_{im} \triangleq \begin{bmatrix} \Re\{\tilde{\mathbf{Q}}_{im}\} & -\Im\{\tilde{\mathbf{Q}}_{im}\} \\ -\Im\{\tilde{\mathbf{Q}}_{im}\} & \Re\{\tilde{\mathbf{Q}}_{im}\} \end{bmatrix}$, $\bar{\mathbf{G}} \triangleq \begin{bmatrix} \Re\{\mathbf{G}\} & -\Im\{\mathbf{G}\} \\ -\Im\{\mathbf{G}\} & \Re\{\mathbf{G}\} \end{bmatrix}$, and $\bar{\mathbf{O}}_x \triangleq \begin{bmatrix} \mathbf{O}_x & \mathbf{0} \\ \mathbf{0} & \mathbf{O}_x \end{bmatrix}$, the SDR problem **P4** can be transformed to a real-valued problem that can be addressed by the interior-point method. The details of the interior-point method can be found in [49].

4.3.3 Approximation Error of the SDR

In the previous analysis, we relaxed the PAPR-aware massive MIMO problem by dropping the nonconvex rank-1 constraint to formulate the optimization problem **P4**, and derive the optimal value, $\tilde{\mathbf{X}}_{opt}$. The performance loss in terms of approximation error is evaluated in this subsection. If the rank of the optimal solution of **P4** is 1, there is no approximation error since we can further decompose it as $\tilde{\mathbf{X}}_{opt} = \tilde{\mathbf{x}}_{opt}\tilde{\mathbf{x}}_{opt}^H$, where $\tilde{\mathbf{x}}_{opt}$ is the optimal solution of **P2**. Without loss of generality, assuming that $\mathrm{rank}(\tilde{\mathbf{X}}_{opt}) > 1$, $\tilde{\mathbf{X}}_{opt}$ can be factorized with Cholesky decomposition

$$\tilde{\mathbf{X}}_{opt} = \begin{bmatrix} \mathbf{U^H} \\ \mathbf{u^H} \end{bmatrix} \begin{bmatrix} \mathbf{U} & \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{U^H U} & \mathbf{U^H u} \\ \mathbf{u^H U} & \mathbf{u^H u} \end{bmatrix}, \tag{4.15}$$

since being a Hermitian PSD matrix of rank N_x or lower, where $\mathbf{U} \in \mathbb{C}^{N_x \times N_x}$ and $\mathbf{u} \in \mathbb{C}^{N_x \times 1}$, $\|\mathbf{u}\|_2^2 = 1$. For now, suppose $\mathbf{x}_{opt}$ is the optimum of **P1**, we can derive that $\tilde{\mathbf{X}}'_{opt}$ is the optimum of problem **P3**, where

$$\tilde{\mathbf{X}}'_{opt} = \left[\begin{array}{c:c} \mathbf{x}_{opt}\mathbf{x}_{opt}^H & \mathbf{x}_{opt} \\ \hdashline \mathbf{x}_{opt}^H & 1 \end{array}\right]. \tag{4.16}$$

By comparing (4.15) and (4.16), we have the following lemma:

Lemma 4.1. *Suppose* $\Delta\mathbf{x}_u$ *is the direct approximation error of* $\mathbf{x}_{opt}$ *(*$\Delta\mathbf{x}_u = \mathbf{x}_{opt} - \mathbf{U}^{\mathbf{H}}\mathbf{u}$*), the upper bound of the approximation error satisfies* $\|\Delta\mathbf{x}_u\|_2^2 \leq 2\delta_e/|\lambda_{min}(\mathbf{H}\mathbf{H}^H)|$*, where* $\lambda_{min}(\cdot)$ *represents the minimum eigenvalue of a matrix.*

Proof. See Appendix E. □

Remarks:

- The upper bound of the approximation error is determined by both the maximum allowed precoding error (a.k.a. MUI allowance) δ_e and the minimum eigenvalue of $\mathbf{H}\mathbf{H}^H$, which are independent ($\mathbf{H}$ is defined in (4.1), the distribution of $\lambda_{min}(\mathbf{H}\mathbf{H}^H)$ can be found in *Proposition 3.6* of [113]);
- The approximation error of the SDR will be sufficiently low for a sufficiently low δ_e, which is a preset value in the aforementioned optimization problems **P1**-**P4**;
- Due to the duality of the optimization, a lower optimal transmit power is achieved with higher δ_e.

4.3.4 Rank-1 Solution via Randomization

We are able to use the SDR solution of **P4** to approximate the solution of the original problem **P1** directly with error $\Delta\mathbf{x}_u$. However, it is still necessary to find a more precise method by deriving the solution of **P1** through the solution of **P4**, this is especially true when the maximum allowed precoding error δ_e is large. Randomization is widely applied to extract an approximate QCQP solution from an SDR solution $\tilde{\mathbf{X}}_{opt}$ [49–52, 114]. The randomization generates a set of candidate vectors $\{\tilde{\mathbf{x}}_i\}$ using $\tilde{\mathbf{X}}_{opt}$ and chooses the best solution from these candidate vectors. The overall SDR and rank reduction algorithm based on randomization is described below:

Algorithm 5 SDR and rank reduction algorithm via randomization [114]

```
Input: $\tilde{\mathbf{Q}}_{im}$, $\mathbf{G}$, $\mathbf{O}_x$, $M$ (number of randomization iterations)
Output: $\tilde{\mathbf{x}}_{opt}$
    Solve the SDR problem (P4), and obtain its solution $\tilde{\mathbf{X}}_{opt}$;
    Factorize $\tilde{\mathbf{X}}_{opt} = \mathbf{V}^H\mathbf{V}$;
    for $i = 1, 2, ..., M$ do
        Generate $\boldsymbol{\xi}_i \sim \mathcal{N}(0, 1)$;
        Compute $\tilde{\mathbf{x}}_i = \mathbf{V}^H\boldsymbol{\xi}_i$;
    end for
    Choose $k = \arg\min_{i=1,...,M}(\|\tilde{\mathbf{x}}_i\|_2^2)$;
    return $\tilde{\mathbf{x}}_{opt} = \tilde{\mathbf{x}}_k$.
```

The mathematical proof and approximation quality of Algorithm 5 can be found in [114], which shows that the optimal rank-1 solution exists for a large enough M. Note that the complexity of randomization process is much smaller relative to the SDR approach [114], and hence its computational complexity can be ignored in the complexity comparison in Section 4.6.

4.4 Robust PAPR-Aware Precoding With Imperfect CSI

Robust optimization is usually developed to address either the bounded or statistical CSI errors. We propose three approaches to achieve robustness under channel uncertainties, namely coarse robust precoding (bounded-error), fine robust precoding via the S-procedure (bounded-error), and fine robust precoding via the Bernstein-type inequality (statistical error), respectively. In this work, the transmit signals are optimized to achieve given PAPRs $\alpha_1, \cdots, \alpha_{N_t}$ and maximum allowed precoding error δ_e for every possible CSI error, and so facilitate the robust precoding. In particular, the robust precoding can be achieved by operating under the bounded-error channel conditions if the CSI errors are bounded. Alternatively, the robust precoding can also be ensured in the probabilistic sense if the channel uncertainties are modeled as statistical distributions. The channel matrix is represented as below to model the channel estimation inaccuracies in massive MIMO systems

$$\tilde{\mathbf{H}} = \hat{\mathbf{H}} + \Delta\tilde{\mathbf{H}}, \tag{4.17}$$

where $\hat{\mathbf{H}}$ captures the channel matrix which is measured at the BS from the uplink channel, and $\Delta\tilde{\mathbf{H}}$ the channel uncertainty. Note that from (4.2), we get

$$\begin{aligned} \mathbf{E}\left(\Delta\tilde{\mathbf{H}}\Delta\tilde{\mathbf{H}}^H\right) &= \mathbf{E}\left(\Delta\mathbf{H}\bar{\mathbf{P}}\bar{\mathbf{Q}}\bar{\mathbf{Q}}^H\bar{\mathbf{P}}^H\Delta\mathbf{H}^H\right) \\ &= \mathbf{E}\left(\Delta\mathbf{H}\Delta\mathbf{H}^H\right), \end{aligned} \tag{4.18}$$

where $\Delta\mathbf{H} = \text{diag}\{\Delta\mathbf{H}_1, \cdots, \Delta\mathbf{H}_{N_c}\}$ and $\Delta\mathbf{H}_m$ indicates the $M_r \times N_t$ channel error matrix of the mth subcarrier. In this section, we consider both the deterministic and stochastic model of the channel uncertainty, and derive the respective robust precoding schemes.

- *Deterministic model (bounded-error):* Here we assume the error matrix $\Delta\tilde{\mathbf{H}}$ take values from the bounded set

  $$\Upsilon_d := \left\{\mathrm{Tr}(\Delta\tilde{\mathbf{H}}\Delta\tilde{\mathbf{H}}^H) \leq \epsilon_h^2\right\}, \tag{4.19}$$

  where $\epsilon_h > 0$ denotes for the radius of the feasible region Υ_d, which is associated with the degree of uncertainty of equivalent channel measurement $\hat{\mathbf{H}}$.

- *Stochastic model:* Here we assume the channel uncertainty subject to zero-mean complex Gaussian distribution with variance vector $\mathbf{R}_\epsilon$

  $$\Upsilon_p := \left\{\mathrm{vec}(\Delta\tilde{\mathbf{H}}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_\epsilon)\right\}, \tag{4.20}$$

  where $\mathrm{vec}(\cdot)$ represents the the column-by-column matrix vectorization.

The shape of the feasible region depends on the second-order statistics of the channel uncertainty, and the specific channel estimation approach. Since the channel information is only involved in the constraint (4.10c), we need to guarantee no constraint violation of (4.10c) to ensure robust precoding.

4.4.1 Coarse Robust Precoding

We first adopt the bounded-error model and develop the coarse robust precoding for the bounded-error channel uncertainty. Since the constraints need to be satisfied for all possible CSI errors, the upper bound of left side in (4.10c) should be less than the maximum allowed precoding error δ_e. As a result, by applying the Cauchy-Schwarz inequality, the constraint

(4.10c) under the circumstances of coarse robust precoding can be represented as

$$\begin{aligned}\left\|\mathbf{s}-\left(\hat{\mathbf{H}}+\Delta\tilde{\mathbf{H}}\right)\mathbf{x}\right\|_2^2 &\leq \left\|\mathbf{s}-\hat{\mathbf{H}}\mathbf{x}\right\|_2^2+\left\|\Delta\tilde{\mathbf{H}}\mathbf{x}\right\|_2^2\\ &\leq \mathrm{Tr}(\hat{\mathbf{G}}\tilde{\mathbf{X}})+\epsilon_h^2\left(\mathrm{Tr}(\tilde{\mathbf{X}})-1\right),\end{aligned} \tag{4.21}$$

where $\hat{\mathbf{G}} \triangleq \begin{bmatrix}\hat{\mathbf{H}}^H\hat{\mathbf{H}} & -\hat{\mathbf{H}}^H\mathbf{s}\\ -\mathbf{s}^H\hat{\mathbf{H}} & \mathbf{s}^H\mathbf{s}\end{bmatrix}$. Therefore, the constraint of (4.10c) in terms of coarse robust precoding becomes

$$\mathrm{Tr}\left((\hat{\mathbf{G}}+\epsilon_h^2\mathbf{I}_{N_x+1})\tilde{\mathbf{X}}\right)\leq \delta_e+\epsilon_h^2, \tag{4.22}$$

which is a standard SDR constraint, and hence the optimization problem **P4** with coarse robust precoding can be reformulated by replacing (4.10c) with (4.22):

$$\begin{aligned}(\mathbf{P7}):\min_{\tilde{\mathbf{X}}} \quad & \mathrm{Tr}(\tilde{\mathbf{X}})\\ \text{subject to} \quad & (4.10b),(4.10d),(4.10e),(4.22).\end{aligned}$$

The problem **P7** provides a robust optimization with a loose upper bound of precoding error. The advantage of coarse robust precoding lies in the fact that the change between the constraints (4.10c) and (4.22) is not significant.

4.4.2 Fine Robust Precoding via S-Procedure

The loose upper bound derived in (4.22) for coarse robust precoding degrades the performance of robust optimization because it changes the feasibility region. As a result, a fine robust precoding with a tighter upper bound is proposed here. The fine robust precoding is derived as

$$\left\|\mathbf{s}-\left(\hat{\mathbf{H}}+\Delta\tilde{\mathbf{H}}\right)\mathbf{x}\right\|_2^2 = \mathrm{Tr}\left(\begin{bmatrix}\hat{\mathbf{H}}+\Delta\tilde{\mathbf{H}} & -\mathbf{s}\end{bmatrix}\begin{bmatrix}\mathbf{x}\mathbf{x}^H & \mathbf{x}\\ \mathbf{x}^H & 1\end{bmatrix}\begin{bmatrix}(\hat{\mathbf{H}}+\Delta\tilde{\mathbf{H}})^H\\ -\mathbf{s}^H\end{bmatrix}\right)$$

$$= \mathrm{Tr}\left((\hat{\mathbf{H}}_s+\Delta\mathbf{H}_s)\mathbf{W}_x(\hat{\mathbf{H}}_s+\Delta\mathbf{H}_s)^H\right), \tag{4.23}$$

where we define $\hat{\mathbf{H}}_s \triangleq \begin{bmatrix}\hat{\mathbf{H}} & -\mathbf{s}\end{bmatrix}$, $\Delta\mathbf{H}_s = \begin{bmatrix}\Delta\tilde{\mathbf{H}} & \mathbf{0}\end{bmatrix}$ and $\mathbf{W}_x = \tilde{\mathbf{X}}|_{t=1}$. It does not sacrifice the performance degradation of the optimization compared against its coarse counterpart.

Lemma 4.2. *Given the bounded channel error matrix in (4.19), the constraint in (4.23) can be relaxed to (4.24a) and (4.24b)*

$$\delta_e - \lambda\epsilon_h^2 - Tr\left((\tilde{\mathbf{X}}-\tilde{\mathbf{Z}})\hat{\mathbf{H}}_s^H\hat{\mathbf{H}}_s\right) \geq 0 \tag{4.24a}$$

$$\tilde{\mathbf{X}}(-\tilde{\mathbf{X}}+\lambda\mathbf{I}_{N_x+1})^{-1}\tilde{\mathbf{X}} \succeq \tilde{\mathbf{Z}}, \tag{4.24b}$$

where $\lambda > 0$.

Proof. See Appendix F. □

By applying Schur's complement [115], (4.24b) can be further represented as

$$\begin{bmatrix}\tilde{\mathbf{X}}-\lambda\mathbf{I}_{N_x+1} & \tilde{\mathbf{X}}\\ \tilde{\mathbf{X}} & -\tilde{\mathbf{Z}}\end{bmatrix} \succeq \mathbf{0}. \tag{4.25}$$

Therefore, the optimization problem P4 with fine robust precoding can be represented as

$$
\begin{aligned}
(\mathbf{P8}): \min_{\tilde{\mathbf{X}},\mathbf{Y},\lambda>0} \quad & \mathrm{Tr}(\tilde{\mathbf{X}}) \\
\text{subject to} \quad & (4.10b), (4.10d), (4.10e), (4.24a), (4.25).
\end{aligned}
$$

4.4.3 Fine Robust Precoding via Bernstein-Type Inequality

With the assumption of statistical channel uncertainty model, we propose a less conservative reformulation with tractable probabilistic constraint of fine robust precoding using the Bernstein-type inequality [116]. The constraints of (4.37) with the stochastic model can be expressed as

$$
\Pr\Bigg(\mathrm{Tr}\left(\hat{\mathbf{H}}_s \mathbf{W}_x \hat{\mathbf{H}}_s^H\right) + \Delta\mathbf{h}_s^H (\mathbf{I}_{N_s} \otimes \mathbf{W}_x) \Delta\mathbf{h}_s + 2\mathrm{Re}\Big(\mathrm{vec}(\hat{\mathbf{H}}_s \mathbf{W}_x)^H \Delta\mathbf{h}_s\Big) \leq \delta_e \Bigg) \geq 1 - \gamma. \tag{4.26}
$$

Lemma 4.3. *The constraint (4.26) is equivalent to a set of constraints below*

$$
Tr\Big(\mathbf{R}'_\epsilon (\mathbf{I}_{N_s} \otimes \tilde{\mathbf{X}})\Big) - \sqrt{-2\theta}\lambda_1 - \theta\lambda_2 + \delta_e - Tr\left(\hat{\mathbf{H}}_s \tilde{\mathbf{X}} \hat{\mathbf{H}}_s^H\right) \geq 0 \tag{4.27a}
$$

$$
\sqrt{\left\| \mathbf{R}_\epsilon'^{1/2} (\mathbf{I}_{N_s} \otimes \tilde{\mathbf{X}}) \mathbf{R}_\epsilon'^{1/2} \right\|_2^2 + 2 \left\| \mathbf{R}_\epsilon'^{1/2} vec(\hat{\mathbf{H}}_s \mathbf{W}_x) \right\|_2^2} \leq \lambda_1 \tag{4.27b}
$$

$$
\lambda_2 \mathbf{I}_{N_x N_s} - \mathbf{R}_\epsilon'^{1/2} (\mathbf{I}_{N_s} \otimes \tilde{\mathbf{X}}) \mathbf{R}_\epsilon'^{1/2} \succeq \mathbf{0} \tag{4.27c}
$$

$$
\lambda_2 \geq 0. \tag{4.27d}
$$

Proof. See Appendix G. □

Therefore, the optimization problem with fine robust precoding in the statistical sense can be represented as

$$\begin{aligned}(\mathbf{P9}):\min_{\tilde{\mathbf{X}},\lambda_1,\lambda_2} \quad & \mathrm{Tr}(\tilde{\mathbf{X}}) \\ \text{subject to} \quad & (4.10b),(4.10d),(4.10e),(4.27a)-(4.27d).\end{aligned}$$

Note that the constraints in (4.27a), (4.27b) and (4.27c) are a linear constraint, a convex second-order cone (SOC) constraint, and a convex PSD constraint, respectively. This problem can be solved efficiently by using the interior-point method of [77].

4.5 PAPR-Aware Precoding with Intercell Coordination

In this section, we incorporate the interference between the co-channel users into the proposed SDR framework and formulate three different transmission schemes for PAPR-aware precoding with intercell coordination. The transmission schemes of coherent transmission, fast cell selection and interference coordination are illustrated in Fig. 4.2. The single cell massive MIMO system that has been discussed in this work so far applies well for cell-center users. However, a practical limitation of enabling the energy-efficient cellular network is the power consumption of BS while serving cell-edge users. The users at the cell edges not only suffer from high path loss, but also from severe co-channel interference when compared with the cell-center users [104] [117]. The PAPR-aware precoding with incercell coordination is based on the following observations:

Observation 1: PAPR-aware precoding is critical for serving cell-edge users. The BS allocates more power to improve SNR and satisfy quality of services (QoSs) of cell-edge users. PAPR reduction is especially necessary when the average power is high (high operation point), since expensive PAs are necessary to achieve high output power with linearity, and even then they may not be power efficient.

Observation 2: Downlink precoding is beneficial and feasible for intercell coordination when radio resource management is carried out in a centralized manner. Spatial DoFs available at the adjacent cells can be used to attenuate or mitigate the interferences (interference rejection) when the CSI of cell-edge users of adjacent cells are available at the centralized baseband processing unit (CBPU). Note that the interference is jointly suppressed across BSs by the CBPU rather than individually optimized at each BS.

The co-channel interference can be efficiently suppressed by the downlink beamforming optimization with the CSI at the BS. Centralized radio resource management allows for the optimization and scheduling to be done globally at the CBPU.

4.5.1 Coherent Massive MIMO Downlink Transmission

Coherent multipoint transmission is realized by a simultaneous transmission of signals to a user from multiple cooperating BSs and the contribution of each BS is coherently combined at the user terminal. The advantages of coherent transmission lies in the fact that the received SNR is increased with signal power combining and spatial diversity is enhanced. Spatial diversity that combines multiple independent fading paths reduces overall channel fading compared to a single link. The constraint (4.10c) in **P4** under the circumstances of coherent transmission is then given as

$$\left\| \mathbf{s} - \sum_{k=1}^{M_c} \tilde{\mathbf{H}}_k \mathbf{x} \right\|_2^2 \leq \delta_C, \tag{4.28}$$

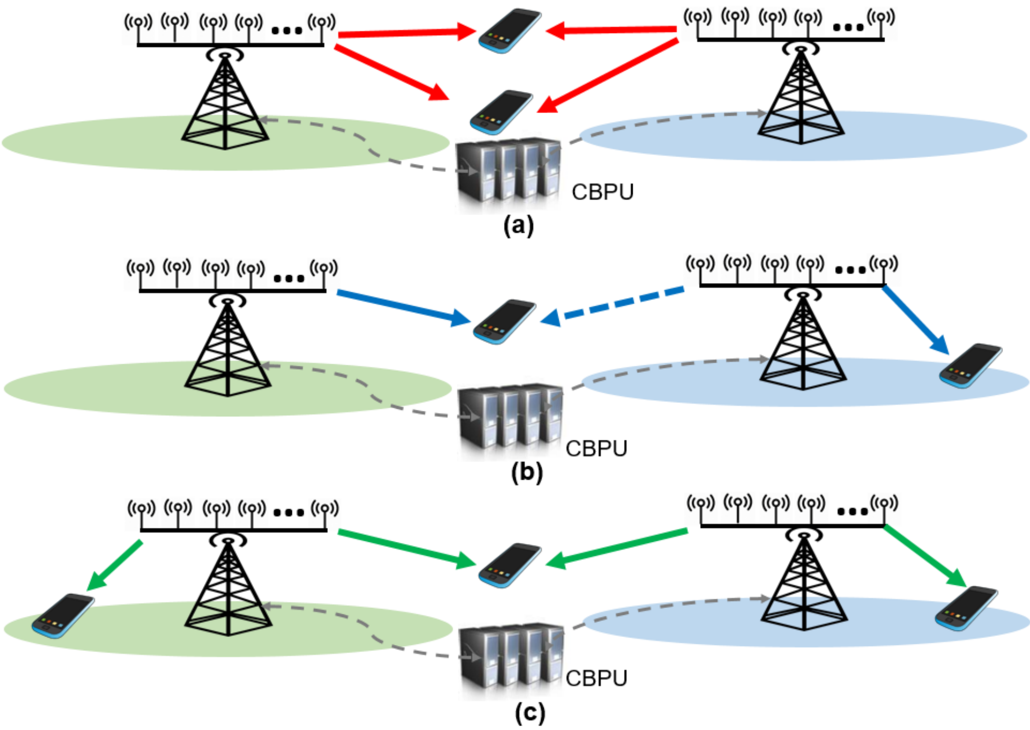


Figure 4.2: Intercell coordination for three transmission schemes, (a) coherent transmission, (b) fast cell selection, and (c) interference coordination, the BSs are coordinated by CBPU.

where δ_C denotes maximum precoding error allowance for the coherent transmission and $\tilde{\mathbf{H}}_k$ the channel information between BS k and all users. The equation can be reformulated as

$$\mathrm{Tr}(\mathbf{Z}_1\tilde{\mathbf{X}}) \leq \delta_C, \tag{4.29}$$

where

$$\mathbf{Z}_1 \triangleq \begin{bmatrix} (\sum_{k=1}^{M_c}\tilde{\mathbf{H}}_k)^H(\sum_{k=1}^{M_c}\tilde{\mathbf{H}}_k) & -(\sum_{k=1}^{M_c}\tilde{\mathbf{H}}_k)^H\mathbf{s} \\ -\mathbf{s}^H(\sum_{k=1}^{M_c}\tilde{\mathbf{H}}_k) & \mathbf{s}^H\mathbf{s} \end{bmatrix}. \tag{4.30}$$

The total number of BSs involved in the interference coordination is given by M_c. The constraint (4.10c) in **P4** is replaced by (4.29), and the centralized radio resource management is applied to coordinate the optimization of a cluster of BSs.

4.5.2 Fast Massive MIMO Cell Selection

The performance of cell-edge users are inevitably degraded by co-channel interference of adjacent cells because of the frequency reuse. However, cell-edge users can be served dynamically by a selected BS through fast scheduling at the CBPU as shown in Fig. 4.2 (b). Therefore, one possible problem formulation is to minimize the interference level of cell-edge users by constraining the interference level from adjacent cells below some threshold. This fast cell selection technique is similar to nulling, where all the precoders are forced to have nulls placed toward other users. The constraint of a group of cell-edge users under the circumstances of fast cell selection can be written as

$$\left\|\tilde{\mathbf{H}}_{e,m}\mathbf{x}\right\|_2^2 \leq \beta_s \lambda_{max}(\tilde{\mathbf{H}}_{e,m}^H \tilde{\mathbf{H}}_{e,m}) \mathrm{Tr}(\mathbf{x}\mathbf{x}^H), \tag{4.31}$$

where $\tilde{\mathbf{H}}_{e,m}$ represents the channel between the mth BS and the group of cell-edge users. The expected total interference power received by the neighboring cell users can thus be limited to a fraction β_s of the worst-case interference, and $\lambda_{max}(\cdot)$ represents the maximum eigenvalue of a matrix. Above constraint can be rewritten as

$$\beta_s \lambda_{max}(\tilde{\mathbf{H}}_{e,m}^H \tilde{\mathbf{H}}_{e,m}) \mathrm{Tr}(\tilde{\mathbf{X}}) - \mathrm{Tr}(\mathbf{Z}_{e,m}\tilde{\mathbf{X}}) \geq \beta_s \lambda_{max}(\tilde{\mathbf{H}}_{e,m}^H \tilde{\mathbf{H}}_{e,m}), \tag{4.32}$$

where

$$\mathbf{Z}_{e,m} \triangleq \begin{bmatrix} \tilde{\mathbf{H}}_{e,m}^H \tilde{\mathbf{H}}_{e,m} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix}. \tag{4.33}$$

Therefore, the constraint of (4.32) is added to the optimization problem **P4** as an additional constraint in each BS.

4.5.3 Massive MIMO Interference Coordination

In contrast to coherent transmission and fast cell selection, interference suppression based coordinated precoding is a more general case and refers to a coordinated selection of the transmit precoders in each cell. This technique aims at eliminating or reducing the effect of intercell interference. The constraint (4.10c) in **P4** under the circumstances of interference coordination precoding is given as

$$\left\| \mathbf{s} - \sum_{k=1}^{M_c} \tilde{\mathbf{H}}_k \mathbf{x}_k \right\|_2^2 \leq \delta_I, \tag{4.34}$$

where $\tilde{\mathbf{H}}_k$ captures the channel information between BS k and all users, and $\mathbf{x}_k$ the transmit signal of BS k. It can be further reformulated as

$$\mathrm{Tr}(\mathbf{Z}_3 \tilde{\mathbf{X}}) \leq \delta_I, \tag{4.35}$$

where δ_I denotes the maximum precoding error allowance for interference coordination, and

$$\mathbf{Z}_3 \triangleq \begin{bmatrix} \tilde{\mathbf{H}}_I^H \tilde{\mathbf{H}}_I & -\tilde{\mathbf{H}}_I^H \mathbf{s} \\ -\mathbf{s}^H \tilde{\mathbf{H}}_I & \mathbf{s}^H \mathbf{s} \end{bmatrix}, \tag{4.36}$$

where $\tilde{\mathbf{H}}_I \triangleq \mathrm{diag}\{\tilde{\mathbf{H}}_1, \cdots, \tilde{\mathbf{H}}_{M_c}\}$, $\mathbf{x} \triangleq [\mathbf{x}_1^H \ \cdots \ \mathbf{x}_{M_c}^H]^H$ and the relationship between $\mathbf{x}$ and $\tilde{\mathbf{X}}$ is the same as defined in Section II. The constraint (4.10c) in **P4** is replaced by (4.35) to formulate the optimization for this type of multicell coordinated transmission as a single optimization problem.

The intercell coordination for different scenarios with CBPU is summarized in Algorithm 6. The optimization of both coherent transmission and interference coordination are performed at the CBPU, the optimization of fast selection are performed at each BS. However, the selection of BS (cell-edge user assignment) in the scenario of fast selection is carried out by the CBPU.

Algorithm 6 Intercell Coordination for Different Scenarios with CBPU

```
Given the scenario of transmission scheme is determined by CBPU, and the index of selected cell if fast cell selection is applied;
if <coherent transmission> then
    Solve $\min_{\tilde{\mathbf{X}}} \mathrm{Tr}(\tilde{\mathbf{X}}) \quad \text{subject to} \quad (4.10b), (4.10d) - (4.10e), (4.29)$;
    Derive $\tilde{\mathbf{x}}_{opt}^{CT}$ with Algorithm 5;
    $\tilde{\mathbf{x}}_{opt} \leftarrow \tilde{\mathbf{x}}_{opt}^{CT}$;
else if <fast cell selection> then
    $j \leftarrow$index of selected cell;
    for $i = 1, \cdots, j-1, j+1, \cdots, M_c$ do
        Solve $\min_{\tilde{\mathbf{X}}_i} \mathrm{Tr}(\tilde{\mathbf{X}}_i) \quad \text{subject to} \quad (4.10b) - (4.10e), (4.32)$;
        Derive $\tilde{\mathbf{x}}_{opt,i}^{FS}$ with Algorithm 5;
    end for
    Solve $\min_{\tilde{\mathbf{X}}_j} \mathrm{Tr}(\tilde{\mathbf{X}}_j) \quad \text{subject to} \quad (4.10b) - (4.10e)$;
    Derive $\tilde{\mathbf{x}}_{opt,j}^{FS}$ with Algorithm 5;
    $\tilde{\mathbf{x}}_{opt} \leftarrow [(\tilde{\mathbf{x}}_{opt,1}^{FS})^H, \cdots, (\tilde{\mathbf{x}}_{opt,M_c}^{FS})^H]^H$;
else if <interference coordination> then
    Solve $\min_{\tilde{\mathbf{X}}} \mathrm{Tr}(\tilde{\mathbf{X}}) \quad \text{subject to} \quad (4.10b), (4.10d) - (4.10e), (4.35)$;
    Derive $\tilde{\mathbf{x}}_{opt}^{IC}$ with Algorithm 5;
    $\tilde{\mathbf{x}}_{opt} \leftarrow \tilde{\mathbf{x}}_{opt}^{IC}$;
end if
return $\tilde{\mathbf{x}}_{opt}$.
```

4.6 Simulation Results

This section presents numerical results based on Monte Carlo simulations of the proposed optimization techniques to validate our analysis and evaluate the proposed algorithms.

4.6.1 Simulation Parameters

The effects of PAPR-aware precoding are demonstrated by simulating single cell and multicell scenarios. The users are uniformly distributed in a geographically manner, the BSs are equipped with 100 antennas and serves 10 single antenna users. The distance between BS antenna elements is one wavelength of the carrier. The case of uncoded 16-QAM is used in the simulation. We simulate OFDM with 128 subcarriers and the multipath channel is modeled as a tapped delay line with 10 taps, each modeled as independently Rayleigh fading. All simulation results are averaged over 100 channel realizations and run with 2000 randomizations ($M = 2000$ in Algorithm 5) for rank reduction.

4.6.2 Complexity Analysis

We compare the computational complexity of the proposed approaches in a variety of scenarios. When the dual-scaling interior-point method is applied, the worst-case computational complexity is $\mathcal{O}(\max\{m, n\}^4 n^{0.5} log(1/\epsilon))$ [62], where m is the number of linear constraints, n is the order of the PSD constraint, and ϵ is the solution accuracy. The comparison of the computational complexities are shown in Table 4.1, where $N_x \triangleq N_t N_c$, $N_s \triangleq M_r N_c$, and M_c is the number of cooperating BSs. The first row of Table 4.1 is the baseline which represents the complexity of massive MIMO precoding without PAPR-awareness by dropping the constraint (4.3b) in problem $\mathbf{P1}$, the subsequent rows give the complexity for other approaches that have been proposed. In addition to the performance we have shown in the simulation section, it should be noted that the aforementioned complexity does not assume sparsity or any special structure in the data matrices $\tilde{\mathbf{Q}}_{im}$ and $\mathbf{G}$.

Table 4.1: Comparison of the Computational Complexities

Algorithms	Complexity
Massive precoding without PAPR-awareness[1,3]	$\mathcal{O}(N_x^{0.5} log(1/\epsilon))$
PAPR-aware precoding[1]	$\mathcal{O}(N_x^{4.5} log(1/\epsilon))$
Coarse *robust* PAPR-aware precoding[2]	$\mathcal{O}(N_x^{4.5} log(1/\epsilon))$
Deterministic fine *robust* PAPR-aware precoding[2]	$\mathcal{O}(N_x^{4.5} log(1/\epsilon))$
Stochastic fine *robust* PAPR-aware precoding[2]	$\mathcal{O}(N_x^{4.5} N_s^{0.5} log(1/\epsilon))$
Coherent transmission with PAPR-aware precoding[1]	$\mathcal{O}(N_x^{4.5} log(1/\epsilon))$
Fast cell selection with PAPR-aware precoding[1]	$\mathcal{O}(M_c N_x^{4.5} log(1/\epsilon))$
Interference coordination with PAPR-aware precoding[1]	$\mathcal{O}(M_c^{0.5} N_x^{4.5} log(1/\epsilon))$

[1] Perfect CSI is assumed.
[2] Imperfect CSI is assumed.
[3] The baseline is formulated as $\min_{\tilde{\mathbf{X}}} \operatorname{Tr}(\tilde{\mathbf{X}}) \quad \text{subject to} \quad (4.10c) - (4.10e)$.

4.6.3 Single-Cell Scenario

Fig. 4.3 compares the average instantaneous transmit power of our proposed method with the joint precoding, modulation and PAPR reduction (PMP) method of [2] and the per-antenna constant envelope precoding of [87]. Since the transmit signals are optimized on a symbol-by-symbol basis, the instantaneous transmit power is used to evaluate the performance of the PAPR-aware precoding with different precoding errors. The lower bound of the transmit power without PAPR-awareness which was shown in row 1 of Table 4.1 is also included. As an indicator of MUI allowance, the normalized maximum allowed precoding error (NMAE) used in Fig. 4.3 is defined as $\alpha = \delta_e/\delta_M$, where δ_e is the maximum allowed precoding error applied in the optimization and $\delta_M = 3.7 \times 10^{-4}$ is the precoding error when the transmit power upper bound is -5 dBm. The simulation results of PAPR-aware precoding with PMP in [2][3] and per-antenna constant envelope method in [87][4] are included to compare with the proposed approach. The NMAE values for plotting the performance of the PMP method in Fig. 4.3 are obtained by altering parameter λ, whereas the MUI is minimized by constant envelope method to find the optimal phases $\theta_1, \cdots, \theta_{N_t}$ for each subcarrier.

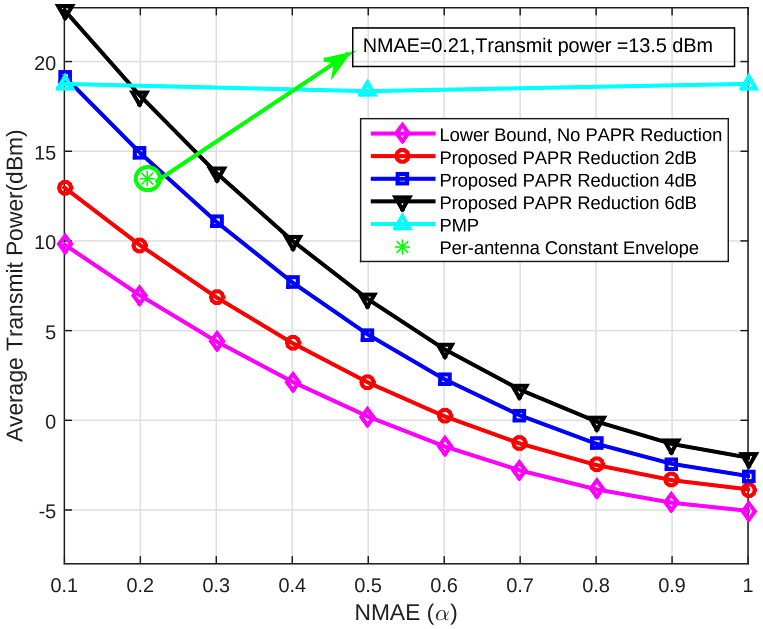


Figure 4.3: Instantaneous transmit power versus NMAE for different PAPR reduction targets, PMP method and constant envelope method (the minimized MUI and associated transmit power of constant envelope method is shown as a single point).

Fig. 4.3 shows that the instantaneous transmit power of the proposed scheme for different PAPR reduction targets[5] decreases when the NMAE value is increased, as derived in Section

---

[3] In [2], the PAPR-aware precoding problem is formulated as $\min_{\tilde{\mathbf{x}}} \left\| \mathbf{s} - \tilde{\mathbf{H}}\tilde{\mathbf{x}} \right\|_2^2 + \lambda \left\| \tilde{\mathbf{x}} \right\|_\infty$ where the signal peak ($\|\tilde{\mathbf{x}}\|_\infty$) instead of PAPR ($N_c \|\tilde{\mathbf{x}}\|_\infty^2 / \|\tilde{\mathbf{x}}\|_2^2$) is applied to realize the convex relaxation, $\lambda \geq 0$ is a regularization parameter to achieve the tradeoff between the PAPR reduction and precoding error (MUI).

[4] In [87], the per-antenna constant envelope method is formulated as $\min_{\theta_1,\cdots,\theta_{N_t}} \sum_{m=1}^{M_r} |\frac{\sum_{n=1}^{N_t} h_{m,n} e^{j\theta_n}}{\sqrt{N_t}} - \sqrt{E_m} u_m|^2$, where θ_n denotes the phase of the transmit signal of the nth antenna, u_m the information symbol of mth user, E_m the information symbol energy of mth user. Since it is designed for the single-carrier waveform, the parallel OFDM symbols are serialized and per-carrier constant envelope precoding is performed to compare with the performance of the proposed approach.

[5] The PAPR reduction target is achieved when the PAPR of 99% of OFDM symbols is reduced with the

4.3.2. The achievable lower bound of the transmit power is also plotted. The figure shows that the instantaneous transmit power stays constant for the PMP scheme of [2]. The instantaneous transmit power and NMAE which derived from the optimal phases of the per-antenna constant envelope method is also shown as a single point in Fig. 4.3.

Fig. 4.4 plots the absolute value of the time domain transmit signals of different schemes for the first antenna. It is observed from Fig. 4.4 that the transmit signal of the proposed method has a lower dynamic range. The signal generated by the PMP method has only an upper bound but the proposed SDR-based method has both an upper bound and lower bound. The PMP method aims to reduce the peak power but the proposed SDR-based method aims to reduce the dynamic range of the transmit signals.

Fig. 4.5 plots the PAPR complementary cumulative distribution functions (CCDFs) of the transmit signal for the approach without PAPR-awareness, the PMP approach, the clipping approach and the proposed SDR-based approach. The CCDF curves of all PAPR-aware precoded OFDM signals are shifted to the left with respect to the scheme that does not exploit the DoFs for PAPR reduction. By comparing with the CCDF of the transmit signal of the PMP method, our proposed approach achieves an additional PAPR reduction of 2 dB.

The channel estimation error $\Delta\tilde{\mathbf{H}} = \tilde{\mathbf{H}} - \hat{\mathbf{H}}$ is modeled by generating $\Delta\tilde{\mathbf{H}}$ from a zero-mean Gaussian distribution with $\mathbf{E}(\Delta\tilde{\mathbf{H}}\Delta\tilde{\mathbf{H}}^H) = \mathbf{R}_\epsilon = \sigma_\epsilon^2\mathbf{I}$, where we will use the same $\sigma_\epsilon^2 = 10^{-3}$ for all users and all subcarriers in the simulation. Moreover, to take different channels into account, the elements of the nominal channel $\hat{\mathbf{H}}$ are randomly generated according to zero-mean, unit-variance, i.i.d. Gaussian distributions. The philosophy of robust PAPR-aware precoding in this work is to guarantee the PAPR reduction and MUI level for any channel realization in the uncertainty region. In other words, we are interested in the behavior of a precoder with uncertainty of CSI.

---

amount of target.

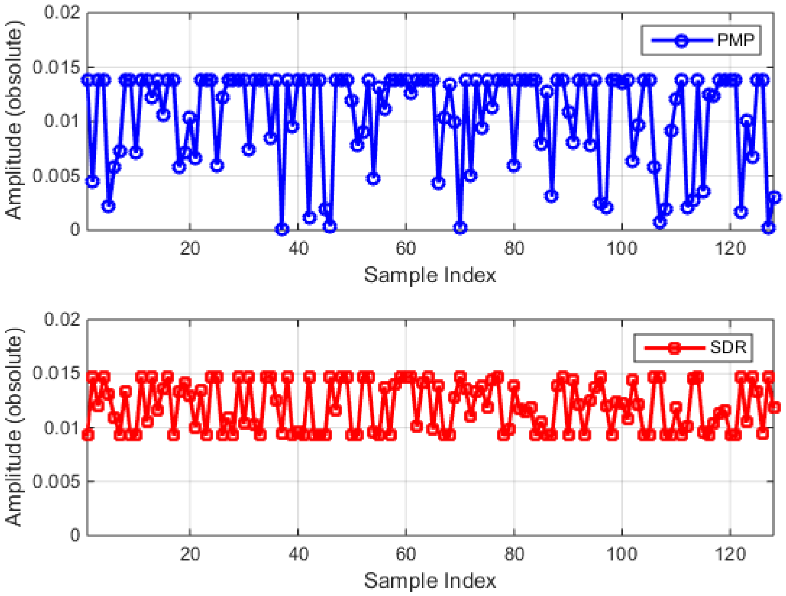


Figure 4.4: The time-domain transmit signals of the first antenna of the PMP method (upper) and the proposed SDR method (lower) ($\delta_e = 10^{-4}$).

Fig. 4.6 plots the performance of the proposed fine robust precoding strategy with probabilistic uncertainty. In order to show the importance of taking the channel uncertainty into account for PAPR-aware precoding, we begin with the non-robust PAPR-aware precoder design when channel uncertainty exists, but the precoder design assumes perfect CSI at the transmitter. Fig. 4.6 shows the percentage of constraint violations at different PAPR reduction targets. Note that a larger precoding error allowance δ_e is used here to achieve PAPR reduction as large as 12 dB. When the channel uncertainty is not considered, the probability of MUI target violation is high, especially with high PAPR reduction targets. However, for our proposed bounded-error fine robust PAPR-aware precoding algorithm, the constraints violation probabilities are low for different values of γ. As shown in Fig. 4.6, the constraints violation probability of the proposed robust precoding is approximately 10 times lower than the non-robust precoding when $\gamma = 0.02$.

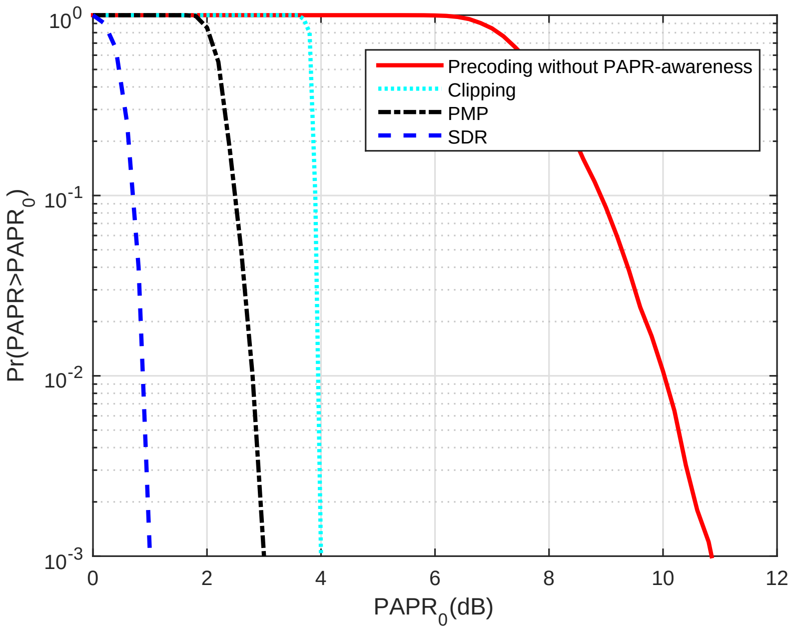


Figure 4.5: Comparison of the PAPR CCDF curves for the first antenna: instantaneous transmit power minimization without PAPR reduction, clipping, PMP and the proposed SDR-based PAPR-aware precoding ($\delta_e = 10^{-4}$).

An alternative way to investigate the robustness capability is the MUI[6] at the receivers for various error bounds ϵ_h. Fig. 4.7 plots the MUI changes with error bound ϵ_h for the PMP method, non-robust precoding based on SDR, coarse robust precoding and fine robust precoding with bounded uncertainty design. The results demonstrate that the proposed fine robust precoding using the S-procedure outperforms the coarse robust approach for PAPR reduction. As shown in Fig. 4.7, the proposed fine robust precoding with bounded uncertainty outperforms coarse robust precoding, non-robust precoding based on SDR and PMP method by 9.2 dB, 15.4 dB, 32.8 dB respectively when the PAPR reduction is 6 dB while outperforms by 7.3 dB, 12.9 dB and 31 dB respectively when the PAPR reduction is 8 dB.

---

[6] The MUI is defined to evaluate the multiuser interference at the receivers that MUI $= \frac{\|\mathbf{s}-\tilde{\mathbf{H}}\mathbf{x}\|_2^2}{\|\mathbf{s}\|_2^2}$.

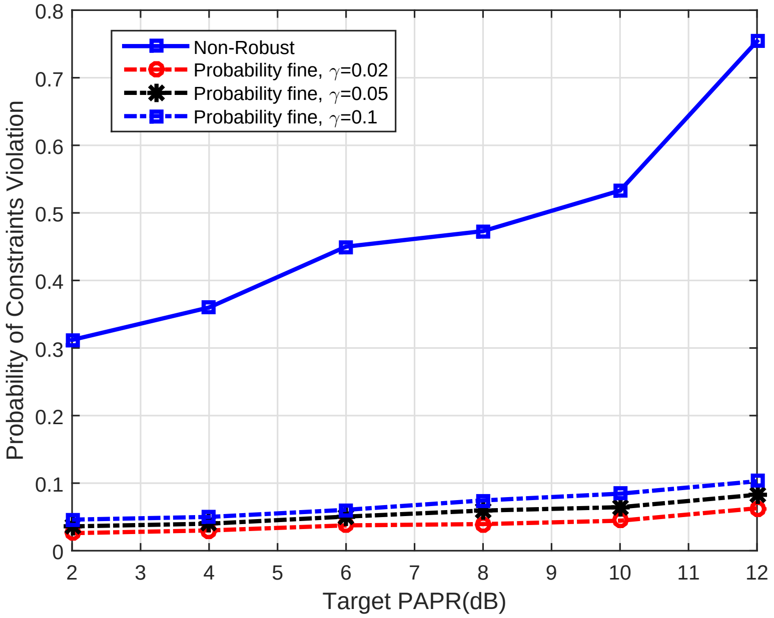


Figure 4.6: Comparison of the percentage of constraints violations for the non-robust precoding [2] and fine robust precoding with probabilistic channel uncertainty design ($\delta_e = 10^{-3}$).

4.6.4 Multi-Cell Scenario

The symbol error rate (SER) performance of the proposed PAPR-aware intercell coordination scheme here is evaluated for a multicell network with 2 cells. Each BS is equipped with 100 antennas. The aim of this numerical analysis is to quantify the benefit of coordinating resource allocation, including scheduling, precoding, and nulling, across multiple cells.

Fig. 4.8 plots the worst-case SER versus SNR without considering channel uncertainty (non-robust) for coherent transmission, fast cell selection and interference coordination precoding scenarios. The PAPR reduction for the BS antennas of both cells are 6 dB. The results show that the coherent transmission outperforms fast cell selection and interference coordination. However, it sacrifices performance in terms of effective transmission rate since both cells coherently transmit the same information symbols. Fast cell selection outperforms the interference coordination in terms of SER. The fine robust precoding with S-procedure results

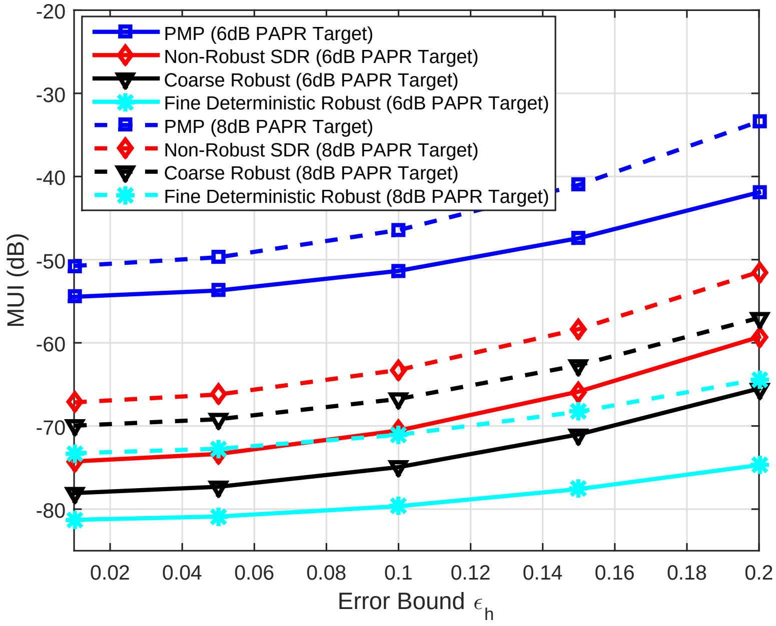


Figure 4.7: Comparison of the MUI for various error bounds ϵ_h for the PMP method, non-robust precoding, coarse robust precoding and fine robust precoding with bounded channel uncertainty design for different PAPR reduction targets ($\delta_e = 10^{-4}$).

for three different scenarios of intercell coordination are also plotted and show consistently better performance with respect to their non-robust counterparts.

4.7 Conclusion

We formulate the optimal downlink precoding problem as minimizing the instantaneous transmit power subject to PAPR and MUI constraints in this chapter. We also formulate robust versions to account for channel uncertainty leveraging the SDR method under various CSI uncertainties. Using numerical analysis, we demonstrate how our approaches reduce the PAPR of BS antennas for single cells and multiple cells serving cell-center and cell-edge users. Numerical results have been conducted to illustrate the performance and flexibility of the proposed framework.

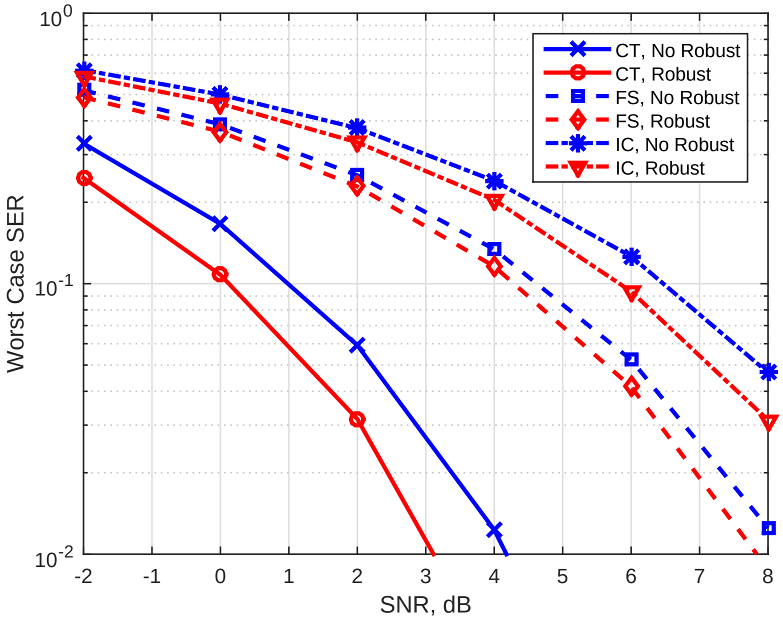


Figure 4.8: Worst-case SER for different scenarios of intercell coordination (16-QAM modulation, $\delta_e = 10^{-4}$, PAPR reduction target 6 dB, $\beta_s = 0.1$, CT: coherent transmission (20 cell-edge users between the two BSs), FS: fast selection (8 cell-center users randomly located within each cell and 4 cell-edge users between the two BSs), IC: interference coordination (8 cell-center users randomly located within each cell and 4 cell-edge users between the two BSs)).

Appendix E

Proof of Lemma 4.1

Proof. Assuming that $\mathbf{s} - \tilde{\mathbf{H}}\mathbf{x}_{opt} = \mathbf{g}$, thus we have $\|\mathbf{g}\|_2^2 \leq \delta_e$ and $\mathrm{Tr}(\mathbf{G}\tilde{\mathbf{X}}_{opt}) \leq \delta_e$ since $\mathbf{x}_{opt}$ and $\tilde{\mathbf{X}}_{opt}$ are the optimal values of **P1** and **P4** respectively. Note that $\|\mathbf{u}\|_2^2 = 1$ and the matrix $\mathbf{G} = \begin{bmatrix} \tilde{\mathbf{H}}^H \\ -\mathbf{s}^H \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{H}} & -\mathbf{s} \end{bmatrix}$. The constraint (4.10c) with optimal value $\tilde{\mathbf{X}}_{opt}$ can be represented as

$$
\begin{aligned}
\mathrm{Tr}(\mathbf{G}\tilde{\mathbf{X}}_{opt}) &= \mathrm{Tr}\left(\begin{bmatrix}\tilde{\mathbf{H}}^H \\ -\mathbf{s}^H\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{H}} & -\mathbf{s}\end{bmatrix}\begin{bmatrix}\mathbf{U}^{\mathbf{H}} \\ \mathbf{u}^{\mathbf{H}}\end{bmatrix}\begin{bmatrix}\mathbf{U} & \mathbf{u}\end{bmatrix}\right) \\
&= \mathrm{Tr}\left(\begin{bmatrix}\tilde{\mathbf{H}} & -\mathbf{s}\end{bmatrix}\begin{bmatrix}\mathbf{U}^{\mathbf{H}} \\ \mathbf{u}^{\mathbf{H}}\end{bmatrix}\begin{bmatrix}\mathbf{U} & \mathbf{u}\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{H}}^H \\ -\mathbf{s}^H\end{bmatrix}\right) \\
&= \left\|\tilde{\mathbf{H}}\mathbf{U}^{\mathbf{H}} - \mathbf{s}\mathbf{u}^{\mathbf{H}}\right\|_2^2 \\
&= \left\|\tilde{\mathbf{H}}\mathbf{U}^{\mathbf{H}} - (\tilde{\mathbf{H}}\mathbf{x}_{opt} - \mathbf{g})\mathbf{u}^{\mathbf{H}}\right\|_2^2 \\
&\geq \left\|\tilde{\mathbf{H}}(\mathbf{U}^{\mathbf{H}} - \mathbf{x}_{opt}\mathbf{u}^{\mathbf{H}})\right\|_2^2 - \left\|\mathbf{g}\mathbf{u}^{\mathbf{H}}\right\|_2^2 \\
&\geq \left\|\Delta\mathbf{x}_u\right\|_2^2 |\lambda_{min}(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H)| - \delta_e.
\end{aligned}
$$

As a result, we can derive that $\|\Delta\mathbf{x}_u\|_2^2 \leq \frac{2\delta_e}{|\lambda_{min}(\mathbf{H}\mathbf{H}^H)|}$, note that $\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H = \mathbf{H}\bar{\mathbf{P}}\bar{\mathbf{Q}}\bar{\mathbf{Q}}^H\bar{\mathbf{P}}^H\mathbf{H}^H = \mathbf{H}\mathbf{H}^H$, and $\mathbf{H} = \mathrm{diag}\{\mathbf{H}_1, \cdots, \mathbf{H}_{N_c}\}$. ☐

Appendix F

Proof of Lemma 4.2

Proof. The constraint of (4.23) is equivalent to

$$
\begin{aligned}
&\mathrm{Tr}\left((\hat{\mathbf{H}}_s + \Delta\mathbf{H}_s)\mathbf{W}_x(\hat{\mathbf{H}}_s + \Delta\mathbf{H}_s)^H\right) \\
=&\mathrm{Tr}\left(\hat{\mathbf{H}}_s\mathbf{W}_x\hat{\mathbf{H}}_s^H\right) + \mathrm{Tr}\left(\Delta\mathbf{H}_s\mathbf{W}_x\Delta\mathbf{H}_s^H\right) + 2\mathrm{Re}\left(\mathrm{Tr}(\hat{\mathbf{H}}_s\mathbf{W}_x\Delta\mathbf{H}_s^H)\right) \\
=&\mathrm{Tr}\left(\hat{\mathbf{H}}_s\mathbf{W}_x\hat{\mathbf{H}}_s^H\right) + \Delta\mathbf{h}_s^H(\mathbf{I}_{N_s}\otimes\mathbf{W}_x)\Delta\mathbf{h}_s + 2\mathrm{Re}\left(\mathrm{vec}(\hat{\mathbf{H}}_s\mathbf{W}_x)^H\Delta\mathbf{h}_s\right) \leq \delta_e, \qquad (4.37)
\end{aligned}
$$

where $\Delta\mathbf{h}_s = \mathrm{vec}(\Delta\mathbf{H}_s)$. Here we use the properties of the trace operator $\mathrm{Tr}(\mathbf{ABA}^H) = \mathrm{vec}(\mathbf{A})(\mathbf{I} \otimes \mathbf{B})\mathrm{vec}(\mathbf{A})^H$ and $\mathrm{Tr}(\mathbf{AB}^H) = \mathrm{vec}(\mathbf{A})\mathrm{vec}(\mathbf{B})^H$ [118]. Due to the property of the column-by-column matrix vectorization and (4.19), we have

$$\|\Delta\mathbf{h}_s\|_2^2 \leq \epsilon_h^2 \tag{4.38}$$

from (4.19). In order to further derive the precoding error constraint under the scenario of channel uncertainty, we adapt a S-procedure lemma from [77] to find the alternative of quadratic inequalities (4.37) and (4.38).

Lemma 4.4. *Consider* $\mathbf{A}_1, \mathbf{A}_2 \in \mathbb{H}^{n\times n}$, $\mathbf{b}_1, \mathbf{b}_2 \in \mathbb{C}^n$, $c_1, c_2 \in \mathbb{R}$, *and suppose there exists an* $\mathbf{u}_0$ *with [77]*

$$\mathbf{u}_0^H\mathbf{A}_2\mathbf{u}_0 + 2Re(\mathbf{b}_2^H\mathbf{u}_0) + c_2 < 0.$$

Then, the inequality

$$\begin{aligned} &\mathbf{u}^H\mathbf{A}_1\mathbf{u} + 2Re(\mathbf{b}_1^H\mathbf{u}) + c_1 \geq 0, \\ \forall\mathbf{u} \in &\Phi_u : \left\{\mathbf{u}|\mathbf{u}^H\mathbf{A}_2\mathbf{u} + 2Re(\mathbf{b}_2^H\mathbf{u}) + c_2 \leq 0\right\} \end{aligned} \tag{4.39}$$

holds if and only if there exists $\lambda \geq 0$ *such that*

$$\begin{bmatrix} \mathbf{A}_1 + \lambda\mathbf{A}_2 & \mathbf{b}_1 + \mathbf{b}_2 \\ (\mathbf{b}_1 + \mathbf{b}_2)^H & c_1 + \lambda c_2 \end{bmatrix} \succeq \mathbf{0}. \tag{4.40}$$

Since $\mathbf{W}_x = \tilde{\mathbf{X}}|_{t=1}$, by applying Lemma 4.4, there exists $\lambda \geq 0$ that the quadratic inequality constraints with considering the channel uncertainty in (4.37) and (4.38) can reformulated as

$$\begin{bmatrix} -\mathbf{I}_{N_s} \otimes \tilde{\mathbf{X}} + \lambda \mathbf{I}_{Ns(N_x+1)} & -\mathrm{vec}(\hat{\mathbf{H}}_s \tilde{\mathbf{X}}) \\ -\mathrm{vec}(\hat{\mathbf{H}}_s \tilde{\mathbf{X}})^H & \delta_e - \mathrm{Tr}\left(\hat{\mathbf{H}}_s \tilde{\mathbf{X}} \hat{\mathbf{H}}_s^H\right) - \lambda \epsilon_h^2 \end{bmatrix} \succeq \mathbf{0}, \tag{4.41}$$

where λ is a variable which needs to be found in the optimization. By applying Schur's complement [115], the equation (4.41) can be further represented as

$$\delta_e - \mathrm{Tr}\left(\hat{\mathbf{H}}_s \tilde{\mathbf{X}} \hat{\mathbf{H}}_s^H\right) - \lambda \epsilon_h^2 - \mathrm{vec}(\hat{\mathbf{H}}_s \tilde{\mathbf{X}})^H \left(-\mathbf{I}_{N_s} \otimes \tilde{\mathbf{X}} + \lambda \mathbf{I}_{Ns(N_x+1)}\right)^{-1} \mathrm{vec}(\hat{\mathbf{H}}_s \tilde{\mathbf{X}}) \geq 0. \tag{4.42}$$

Thus we have

$$\delta_e - \lambda \epsilon_h^2 - \mathrm{Tr}\left(\tilde{\mathbf{X}} \hat{\mathbf{H}}_s^H \hat{\mathbf{H}}_s\right) - \mathrm{Tr}\left(\tilde{\mathbf{X}}(-\tilde{\mathbf{X}} + \lambda \mathbf{I}_{N_x+1})^{-1} \tilde{\mathbf{X}} \hat{\mathbf{H}}_s^H \hat{\mathbf{H}}_s\right) \geq 0, \tag{4.43}$$

which can be relaxed as (4.24a) and (4.24b). □

Appendix G

Proof of Lemma 4.3

Proof. Since $\mathrm{vec}(\Delta\tilde{\mathbf{H}}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_\epsilon)$, we have that $\Delta \mathbf{h}_s \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}'_\epsilon)$, where $\mathbf{R}'_\epsilon = [\mathbf{R}_\epsilon, \mathbf{0}]$. Through the following conversion:

$$\Delta \mathbf{h}_s = \mathbf{R}_\epsilon'^{1/2} \mathbf{e}_h, \tag{4.44}$$

where $\mathbf{e}_h \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, the entries of $\mathbf{R}_\epsilon^{'1/2}$ are the square root of the entries of $\mathbf{R}'_\epsilon$. The probabilistic inequality (4.26) can be transformed as

$$\Pr\Big(\mathbf{e}_h^H \mathbf{W}_\epsilon \mathbf{e}_h + 2\mathrm{Re}\left(\mathbf{g}_\epsilon^H \mathbf{e}_h\right) + \delta_e - \mathrm{Tr}\left(\hat{\mathbf{H}}_s \tilde{\mathbf{X}} \hat{\mathbf{H}}_s^H\right) \geq 0\Big) \geq 1 - \gamma,$$

where we defined $\mathbf{W}_\epsilon$ and $\mathbf{g}_\epsilon$ as follows

$$\mathbf{W}_\epsilon \triangleq -\mathbf{R}_\epsilon^{'1/2}(\mathbf{I}_{N_s} \otimes \mathbf{W}_x)\mathbf{R}_\epsilon^{'1/2} \tag{4.45}$$

$$\mathbf{g}_\epsilon \triangleq -\mathbf{R}_\epsilon^{'1/2}\mathrm{vec}(\hat{\mathbf{H}}_s \mathbf{W}_x), \tag{4.46}$$

since $\mathbf{R}'_\epsilon$ is a Hermitian matrix, note that $\mathbf{W}_x = \tilde{\mathbf{X}}|_{t=1}$.

Lemma 4.5. *Consider* $\mathbf{W} \in \mathbb{H}^{n \times n}$, $\mathbf{g} \in \mathbb{C}^n$, $c \in \mathbb{R}$, $\mathbf{u} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ *and* $\gamma \in (0, 1]$ *and define* $\theta \triangleq -\log(\gamma)$ *[116]. The following condition*

$$Pr\left\{\mathbf{u}^H \mathbf{W} \mathbf{u} + 2Re(\mathbf{g}^H \mathbf{u}) + c \geq 0\right\} \geq 1 - \gamma \tag{4.47}$$

is satisfied if and only if all of the following inequalities hold:

$$\mathrm{Tr}(\mathbf{W}) - \sqrt{-2\theta}\lambda_1 - \theta\lambda_2 + c \geq 0 \tag{4.48a}$$

$$\sqrt{\|\mathbf{W}\|_2^2 + 2\|\mathbf{g}\|_2^2} \leq \lambda_1 \tag{4.48b}$$

$$\lambda_2 \mathbf{I} + \mathbf{W} \succeq \mathbf{0} \tag{4.48c}$$

$$\lambda_2 \geq 0, \tag{4.48d}$$

where $\lambda_1, \lambda_2 \in \mathbb{R}^+$ are slack variables. By applying Lemma 4.5, we obtain the constraints of (4.27a)-(4.27d). □

Chapter 5

Digital Predistortion for Massive MIMO Systems Exploiting Degrees-of-Freedom

5.1 Introduction

High IBO is often required to keep the signal within the amplifier's linear region and avoid in-band and out-of-band distortions and spectral regrowth that give rise to both amplitude/amplitude modulation (AM/AM) and amplitude/phase modulation (AM/PM) nonlinear distortion. However, operating with high IBO results in low energy efficiency especially considering the high heat dissipation and associated cooling. Therefore, the tradeoff between power efficiency and linearity of the PA has motivated the development of linearization techniques for energy-efficient cellular communications. DPD, which is used for PA nonlinearity compensation, is a popular solution among the linearization schemes because of its flexibility, accuracy and ability to adapt to the time-variant characteristic of the PA. It enables the application of low-cost PAs and operation of higher signal levels for increased energy efficiency without sacrificing linearity. This is especially needed for the BS of 5G NR since it uses 1-2 orders of magnitude more PAs than 4G LTE.

The DPD aims to realize a linear response for the overall DPD-PA blocks by cascading the PA and its inverse response. It can be generally categorized into two groups: polynomial based scheme [13] and LUT based scheme [59]. The Volterra series represents one of the most popular polynomial models of a nonlinear system. Moreover, the PA power response is time-variant when dealing with wideband signal inputs such as 10 or 20 MHz LTE. Therefore, the memory-based Volterra model is widely applied to represent the behavior of PA in wideband scenarios. Literature shows that a wide range of PA nonlinearities can be approximated with considerable precision by the Volterra series-based filter of sufficient order and memory depth in both single-input single-output [13] and MIMO systems [119, 120]. The Volterra series-based method can be a complete representation of an unknown nonlinear dynamic system, but has the drawback of requiring a large number of basis functions. This limits its practical application especially for large-scale MIMO (a.k.a. massive MIMO) systems of the emerging 5G NR technology because of the computational complexity. MIMO technology has gained popularity in the development and deployment of modern wireless communication systems, because of the improved signal to noise ratio (SNR), link coverage and spectral efficiency. To accommodate for increasing demand of smart devices in emerging economies as well as heterogeneous networks, proposals for 5G NR wireless communication standards consider massive MIMO with up to hundreds of transmit antennas. In massive MIMO systems, a digital predistorter is needed for each low-cost PA, one per antenna. The implementation of digital predistorters in massive MIMO system is challenging for the following reasons:

- The use of full-order and hence highly precise Volterra series-based DPD is computationally impractical;

- The required number of Volterra series grows exponentially with the number of antenna when we are considering the notable RF/antenna crosstalk in MIMO systems [119]; and

- The input of the PA is derived from the precoder/beamformer, which combines multiple data streams, and hence creates an interplay between dynamic precoder and DPD.

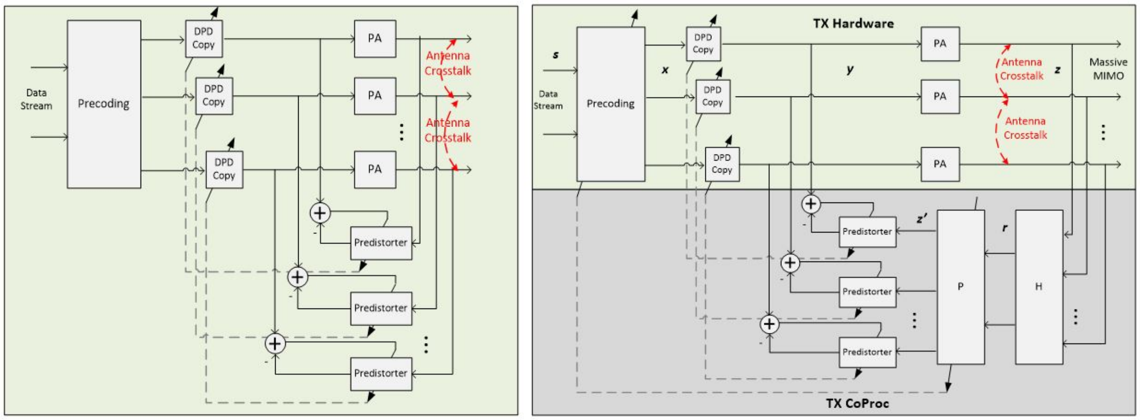


Figure 5.1: Conventional DPD for massive MIMO systems (left) and proposed, low-complexity and precoding-aware DPD solution (right, TX hardware: transmitter hardware, TX CoProc: transmitter coprocessor, the feedback loop includes PAs, channel matrix H which derived from TDD uplink channel estimation, adaptive precoding matrix P, and adaptive predistorter).

Therefore, it is critical to extend the operation of the PA into weakly nonlinear region of the low-cost PAs in massive MIMO systems and hence reduce the required order of the Volterra series. This work proposes a low-complexity and scalable DPD algorithm for massive MIMO system that takes advantage of spatial DoFs and facilitates the energy-efficient BS of 5G NR.

5.2 System Model and Problem Formulation

Consider a downlink massive MIMO system which has one BS equipped with N_t antennas as shown in Fig. 5.1 (left). The number of BS antennas is significantly larger than the number of single antenna users, $N_t \gg M_r$. When the BS transmits signals to M_r users in parallel by means of spatial multiplexing, the received signal of user k can be represented as

$$r_k = \mathbf{h}_k \Big[f_1(\mathbf{p}_1\mathbf{s}), f_2(\mathbf{p}_2\mathbf{s}), \cdots, f_{N_t}(\mathbf{p}_{N_t}\mathbf{s}) \Big]^T + n_k, \tag{5.1}$$

where $\mathbf{h}_k$ represents the $1 \times N_t$ channel vector from the BS to user k which is the kth row of the channel matrix, $\mathbf{p}_k$ and $\mathbf{s}$ respectively the $1 \times M_r$ precoding/beam-weight vector and $M_r \times 1$ information symbols, $f_i(\cdot)$ the nonlinear amplification operation of the ith PA, n_k the zero-mean complex circular-symmetric additive white Gaussian noise (AWGN) at user k, and the superscript T the matrix transpose operation. The information symbols at the BS are mapped to the appropriate transmit antennas so that the information received by each user has minimal interference from the signals of the other users. A larger number of spatial DoFs then necessary for zero forcing allows selecting from a larger signal space for precoding to minimize the multiuser interference. It allows the PAs to work near the nonlinear operating region. In this section, we start with an ideal PA model by assuming that the PAs are perfectly linear and ignoring the memory effect to simplify the analysis. Zero forcing precoding (ZFPC) is employed for multiuser signal transmission because of its simplicity and outstanding performance, the beam-weights of ZFPC need to satisfy the following constraint

$$\mathbf{HP} = \mathbf{I}_{M_r}, \tag{5.2}$$

where $\mathbf{H} = [\mathbf{h}_1^T, \cdots, \mathbf{h}_{M_r}^T]^T$ denotes the $M_r \times N_t$ channel matrix between the BS and all users, $\mathbf{P} = [\mathbf{p}_1^T, \cdots, \mathbf{p}_{N_t}^T]^T$ the $N_t \times M_r$ precoding matrix and $\mathbf{I}_{M_r}$ an $M_r \times M_r$ identity matrix.

Without loss of generality, the PA nonlinearity and the RF coupling between the different paths of the transmitters, which are the two major sources of ZFPC condition violation and the transmitter impairment, need to be considered for massive MIMO deployments. Therefore, an extra nonlinear function is needed to preprocess the input signal of the PA and thus linearize the overall cascaded DPD-PA amplified signal. In order to determine this nonlinear function, the behavioral model and inverse behavioral model of PAs are needed. Suppose $f_i(x_i)$ is the nonlinear model of the transmitter for the baseband input signal $x_i = \mathbf{p}_i\mathbf{s}$ and z_i is the equivalent complex signal at the output of the ith PA, the predistortion function $g_i(x_i)$ has to satisfy the following set of relations

$$z_i = f_i(g_i(x_i)) = G_0 x_i, \tag{5.3}$$

where G_0 is the equivalent linear gain of the DPD-PA cascaded system. A Volterra series-based memory polynomial predistorter is applied to compensate the dynamic nonlinear behavior in this section. This model, known as parallel Hammerstein model in literature [119], is a parallelization of a nonlinear function followed by a linear memory. Moreover, apart from the typical odd-order polynomial representation, even-order polynomials are also included to enrich the basis set and improve the modeling accuracy and DPD performance [119]. The discrete baseband-equivalent form of the Volterra series with memory effect consists of a sum of multidimensional convolutions that can be written as [119]

$$y(n) = \sum_{q=0}^{Q} \sum_{k=1}^{K} \omega_{k,q} |x(n-q)|^{k-1} x(n-q), \tag{5.4}$$

where $x(n)$ and $y(n)$ are the input and output complex envelope of the signals, $\omega_{k,q}$ the polynomial coefficient of the filter tap for kth order and qth delay, and K and Q the nonlinear degree and memory depth, respectively. By defining a new sequence

$$r_{k,q}(n) = |x(n-q)|^{k-1} x(n-q), \tag{5.5}$$

we can rewrite (5.4) as

$$\mathbf{y} = \mathbf{R}\boldsymbol{\omega}, \tag{5.6}$$

where $\mathbf{y} = [y(n), y(n+1), \cdots, y(n+N-1)]^T$, $\boldsymbol{\omega} = [\omega_{10}, \cdots, \omega_{11}, \cdots, \omega_{KQ}]^T$, $\mathbf{R} = [\mathbf{r}_{10}, \cdots, \mathbf{r}_{11}, \cdots, \mathbf{r}_{KQ}]$, and $\mathbf{r}_{kq} = [r_{kq}(n), r_{kq}(n+1), \cdots, r_{kq}(n+N-1)]^T$. Notice that the proposed Volterra-based models are linear with the coefficients $\boldsymbol{\omega}$. We can then estimate the model parameter $\boldsymbol{\omega}$ by minimizing the following cost function [119]

$$C(\boldsymbol{\omega}) = \arg\min_{\boldsymbol{\omega}} \|\mathbf{y} - \mathbf{R}\boldsymbol{\omega}\|^2 . \tag{5.7}$$

In this work, we apply the least mean squares (LMS) adaptive algorithms for DPD, where the DPD coefficients are updated every sample, and compare the linearization performance between the conventional approach and our proposed approach. The LMS algorithm is based on the minimum mean square error rule with steepest descent which can be described as [119]

$$\mathbf{e}(m) = \mathbf{y}(m) - \mathbf{R}(m)\boldsymbol{\omega}(m-1), \tag{5.8a}$$

$$\boldsymbol{\omega}(m) = \boldsymbol{\omega}(m-1) + \mu\mathbf{e}^H(m)\mathbf{R}(m). \tag{5.8b}$$

The weights of the LMS filter $\boldsymbol{\omega}(m) = \mathrm{LMS}(\mathbf{x}(m), \mathbf{y}(m))$ are a function of the input $\mathbf{x}(m)$ and output $\mathbf{y}(m)$ of the LMS algorithm and is initialized as $\boldsymbol{\omega}(0) = \boldsymbol{\omega}_0$. Expression $\mathbf{e}^H(m)$ is the Hermitian transpose of the error vector and μ the step size.

5.3 Proposed Architecture and Algorithm

The predistorter is trained using the system identification architecture, where the PA characteristics are identified in the feedback path [119]. A conventional indirect learning structure applied for massive MIMO system is shown in Fig. 5.1 (left). It duplicates the DPD-PA structure for each RF path and the LMS filters associated with each PA are updated independently. This architecture fails to consider the influence of ZFPC on DPD. Since the number of BS antennas is significantly larger than number of users in our model, the precoding matrix is underdetermined. This implies that there is more than one solution for the precoding matrix satisfying the constraint (5.2). The number of solutions depends on the rank of the channel matrix. Then the objective of our proposed scheme is finding the appropriate precoding matrix which enables low-complexity DPD with low order basis functions. Therefore, instead of the traditional one-stage indirect learning structure of Fig. 5.1 (left), we propose a novel learning structure which adapts the channel and PA distortion iteratively by cascading adaptive ZFPC and DPD.

The conventional indirect learning architecture in which feedback paths are independent among each other can be modified by incorporating the channel matrix into the feedback path. To be more specific, the feedback loop in Fig. 5.1 (right) includes the channel matrix, adaptive ZFPC, adaptive DPD and PAs. This new indirect learning architecture lends itself to efficient, yet flexible implementation for 5G NR. The forward paths in the proposed structure, which are always functional, can be implemented in hardware (e.g. ASIC or FPGA), exploiting parallel processing. The feedback paths, which need to be functional during the weight updating phase only, can be implemented by a general purpose processor-based coprocessor in serial manner. It is critical to reduce the complexity of the forward paths, especially the number of complex multipliers, to enable the ultra-low latency communications in 5G NR. The proposed scheme exploits the enormous DoFs in a massive MIMO system to reduce the number of basis functions in Volterra series and to mitigate its deleterious effects of crosstalk. The algorithm flow of successive refinement filtering is depicted below:

Algorithm 7 Successive Refinement Filtering

```
Input: channel matrix $\mathbf{H}$, modulated symbol $\mathbf{s}$
Output: predistortion matrix $\mathbf{G}$, precoding matrix $\mathbf{P}$
  Initialization: $\mathbf{P} = \mathbf{P}_0$
while error $\geq$ threshold $\epsilon$ do
    $\mathbf{G}(m) = \mathrm{LMS}(\mathbf{z}'(m), \mathbf{y}(m))$;
    DPD update: DPD $\leftarrow \mathbf{G}(m)$;
    $\mathbf{P}(m) = \mathrm{LMS}(\mathbf{r}(m), \mathbf{x}(m))$;
    ZFPC update: ZFPC $\leftarrow \mathbf{P}(m)$;
    $m \leftarrow m + 1$;
end while
return G and P
```

As shown in Fig. 5.1 (right), $\mathbf{y}(m)$ represents for the input signal of PA for mth iteration, $\mathbf{r}(m)$ the input of module P (ZFPC), $\mathbf{z}'(m)$ the input of the adaptive DPD, and $\mathbf{x}(m)$ the output of the precoding in the forward path of the proposed structure.

To give a quantitative measure of complexity of the transmitter hardware in terms of the hardware implementation in 5G NR, we evaluate the floating point operations (FLOPs) [121] required for the architecture of Fig. 5.1 (right). The complex filtering operation requires six FLOPs per filter tap, four real multiplications and two summations. For a given delay tap q, the output of DPD for $(K+1)$-order Volterra series can be represent as

$$\begin{aligned} y_q^{K+1}(n) &= \sum_{k=1}^{K+1} \omega_{k,q} |x(n-q)|^{k-1} x(n-q) \\ &= \omega_{1,q} x(n-q) + |x(n-q)| \underbrace{\sum_{k=1}^{K} \omega_{k+1,q} |x(n-q)|^{k-1} x(n-q)}_{\tilde{y}_q^K(n)} . \end{aligned}$$

There are one complex-real multiplication for $|x(n-q)|\tilde{y}_q^K(n)$, one complex multiplication for

$\omega_{1,q}x(n-q)$ (pre-calculated, negligible), and one complex summation[1]. Therefore, the total saving of the proposed DPD while comparing with the conventional DPD in terms of FLOPs is $4(K_C - K_P)QN_t$, where K_C denotes the order of conventional independent DPD scheme, K_P the order of the proposed scheme, Q the memory depth, and N_t the total number of transmit antennas in the downlink. Notice that N_t is usually extremely large for massive MIMO configurations and hence considerable FLOPs are saved.

5.4 Simulation and Experimental Results

5.4.1 Simulation Results

Simulations are performed with the polynomial PA model containing memory effects. A 100×10 massive MIMO configuration with 100 transmit antenna elements is used to evaluate the proposed DPD algorithm. The simulation setup consists of RF sources that represent the transmitting paths in the MIMO links to accurately capture the effects of the crosstalk on the performance of the massive MIMO transmitter. The crosstalk effect is simulated by coupling the signal in each RF path to its adjacent paths with 20 dB attenuation. The PA input and output are assumed to obey the Saleh model with the parameters $\alpha_a = 2, \beta_a = 2.2, \alpha_\phi = 2, \beta_\phi = 1$ [122]. The center frequency is set to be 3.5 GHz and the baseband bandwidth is set to 10 MHz. Fig. 5.2 shows the PSD output of the PA for the conventional polynomial predistorter with polynomial orders $K = 3$ and $K = 9$ along with the proposed polynomial predistorter with polynomial order $K = 3$.

The highest trace (purple) shows the spectral regrowth of the PA output without predistortion and is seen to be approximately -50 dB below the in-band power level. The black trace

---

[1] The computational complexities of $\tilde{y}_q^K(n)$ and $y_q^K(n)$ are the same.

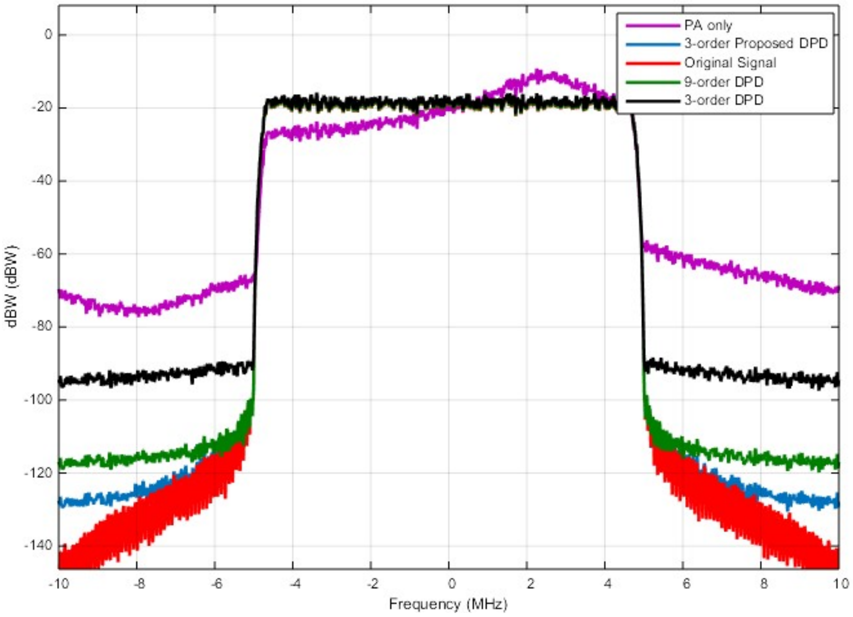


Figure 5.2: Predistortion linearization performance in terms of spectral regrowth suppression. The PA output PSD is shown for different cases with crosstalk set to -20 dB for adjacent antennas. Also shown is the original signal without crosstalk.

shows the output spectrum with the conventional memory polynomial model with nonlinear degree $K = 3$ and memory depth $Q = 5$. It shows an out-of-band emission of -70 dB with respect to the in-band signal power level. The green trace is for the conventional memory polynomial model with nonlinear degree $K = 9$ and memory depth $Q = 5$, which reduces the regrowth to -90 dB relative to the in-band power level. The blue trace shows the output spectrum for the proposed DPD scheme described in Algorithm 7 with nonlinear degree $K = 3$ and memory depth $Q = 5$. The proposed algorithm outperforms the conventional memory polynomial solution $(K = 9, Q = 5)$ and reduces the out-of-band emission by another 10 dB.

5.4.2 Hybird Experimental and Simulation Results

A more practical setup for hybrid experiment and simulation is proposed in this subsection, which gives the calculated normalized mean square error (NMSE) result. It is clear that the hardware implementation of the entire massive MIMO system requires a large number of PAs (and digital converters, mixers, etc.) since each antenna element needs an independent PA as shown in Fig. 5.1. In this subsection, we introduce a hybrid experimental and simulation setup which exploits both the real and theoretical PA models to examine the proposed DPD performance with limited hardware resources (Fig. 5.3).

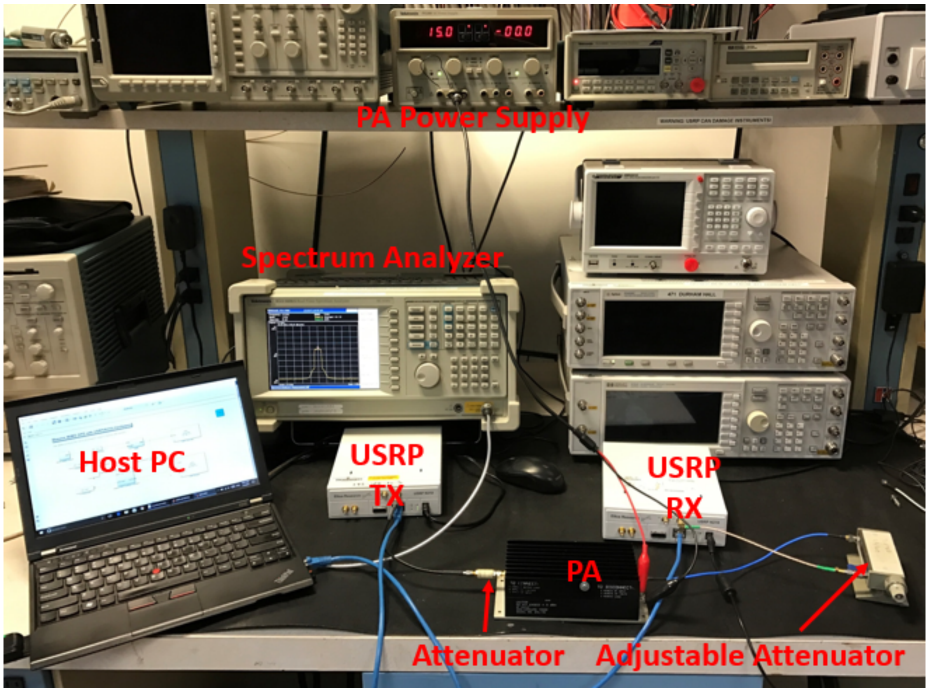


Figure 5.3: Hybrid experimental and simulation setup.

To be more specific, we use a real PA to serve as one of the 100 transmit antenna elements and the others are emulated using a theoretical PA model in the host PC. Synchronization is necessary between the real path and virtual paths. The full array experimental effort will be carried out in the future work. The proposed hybird experimental and simulation setup consists of a mini-circuits ZHL-42 PA (33dB gain@3.5GHz&15V) and two USRPs (N210) which are equipped with SBX daughterboard covers from 400MHz up to 4.4GHz. A real-time spectrum analyzer (RSA3408A) is used to show the output spectrum of ZHL-42 in the experiment. MATLAB Simulink is connected to USRPs to run Algorithm 7 and emulates the Saleh model for the other PAs. The USRPs are instantiated in Simulink. As shown in Fig. 5.3, one transmit signal is sent to an USRP via Ethernet and from USRP to the ZHL-42 PA after the 3.5GHz up-conversion. The amplified signal is attenuated and sent to another USRP and its output is sent to the feedback path in the host PC via an Ethernet-PCIe card after the down-conversion. The rest of the transmit signals are processed by the Saleh model and go to the internal feedback path directly. NMSE between the output of the ZHL-42 and the undistorted input signal, for the conventional and the proposed DPD, are calculated and presented in Table 5.1. It also shows the complexity comparison in terms of the number of FLOPS.

Table 5.1: Performance comparison of no DPD, conventional DPD and the proposed DPD ($N_t = 100$ and $Q = 5$).

Scheme	NMSE	Complexity
ZHL-42 only (No DPD)	-18.41 dB	N/A
Conventional 3rd order DPD	-27.38 dB	7×10^3 FLOPS
Conventional 9th order DPD	-40.17 dB	1.9×10^4 FLOPS
Conventional 11th order DPD	-45.30 dB	2.3×10^4 FLOPS
Proposed 3rd order DPD	-45.89 dB	7×10^3 FLOPS

The proposed 3rd order DPD scheme achieves gains of 18.5, 5.7 and 0.6 dB over the conventional 3rd order, 9th order and 11th order DPD, respectively. Note that only 30% of computational complexity is needed for the forward path of the proposed 3rd order DPD scheme to achieve the same NMSE performance as the conventional 11th order DPD scheme.

5.5 Conclusion

By exploiting excess DoFs, our proposed DPD approach outperforms the conventional DPD approach in terms of accuracy and complexity. Both simulation and hardware experiments confirm the proposed DPD model can linearize PAs better than conventional approach.

Chapter 6

Summary and Future Works

In this dissertation, we focus on the PA-centric research topics of multiple antenna systems, including the DAS and massive MIMO, to improve the energy efficiency.

6.1 Summary for PAPR-aware DASs

This work has derived a PAPR-aware energy-efficient resource allocation scheme for the downlink of OFDMA/SDMA wireless communication systems in the DAS configuration. The proposed scheme considers the overall data rate as well as the PAPR influence on the PA and, thus, the energy consumption. Dual fractional programming is applied to derive the optimal power allocation method for DASs and an iterative solution is proposed to reduce the computational complexity. We also adopted the iterative power allocation method for DAS to solve the key optimization problem of dual fractional programming. The improvement of the PAPR reduction is demonstrated by comparing with traditional energy-efficient resource allocation without PAPR awareness. In addition, the simulation results also show that both PAPR reduction and energy efficiency are improved significantly when compared with traditional schemes at the price of slightly increased convergence time. This, however, is not an issue for real-time implementation in modern hardware.

This work studies the effect of joint channel capacity and PAPR optimization in OFDMA-based DAS. The contributions of this work are outlined below:

1. We introduce a closed-form formula that quantifies the drain efficiency of PA based on unequal subcarrier power allocation;
2. We create a new greedy joint OFDMA/SDMA scheme for DASs to increase the spectral efficiency in the scenario of rich spatial diversity and that avoids the degraded performance due to correlated channel in SDMA;
3. We present a dual fractional programming algorithm to deal with a general case of DAS with unequal and non-uniform PA efficiency, it also provides a low complexity iterative power allocation approach for distributed antennas that is especially efficient for massive antennas;
4. We use the closed-form UPA formula to develop a resource allocation algorithm that jointly optimizes data rate and PAPR, and ultimately maximizes the energy efficiency of a DAS.

6.2 Summary for PAPR-aware Massive MIMO Precoding Using RNN

This work investigates a novel AI based solution for PAPR-aware massive MU-MIMO-OFDM system. Noting the similarity between precoding-PA architecture of massive MIMO system and the recurrent neural network motivates this new approach. The parallelizable structure using recurrent neural networks is especially beneficial as the number of transmit antenna increases. The constant envelope of input signal is achieved by exploiting the extra DoF of transmit antennas. We also proved the stability and convergence of the proposed approach and the simulation results indicates that the convergence speed is much faster than the existing non-AI solution. This result is encouraging for real-time hardware implementations. This work provides a method for enhancing 5G network PA efficiency with AI techniques. The contributions of this work are outlined below:

1. This work presents a PAPR-aware massive MIMO-OFDM downlink system that is inspired by recurrent neural networks. It draws the analogy that linear combination in precoding and the nonlinearity in PA to the linear weighting in neurons and the nonlinearity in activation function of artificial neural networks.
2. The stability and convergence of the proposed recurrently neural network were analyzed in this work, and we proved that there exists an optimal and unique solution for the proposed algorithm, and it converges to the Karush-Kuhn-Tucker (KKT) point.
3. The proposed recurrent neural network structure gives a parallel solution which especially suitable for tackling large-scale problems such as massive MU-MIMO. In addition, it requires fewer multiplication/division and addition/subtraction operations per iteration, and it provides nearly-exponential convergence and is much faster than previous non-AI solutions.

6.3 Summary for PAPR-aware Massive MIMO Precoding Using SDR

This work develops efficient SDR-based approaches for PAPR-aware massive MIMO-OFDM system and analyzes it in different scenarios of robust precoding and intercell coordination. We formulate the optimal downlink precoding problem as minimizing the instantaneous transmit power subject to PAPR and MUI constraints. When the number of BS antennas is much larger than the number of users in a massive MIMO system, the proposed SDR-based method exploits the spatial DoFs to yield a per-antenna PAPR-constrained OFDM signal. The randomization based method is then applied for rank reduction of the SDR solution. We also formulate robust versions to account for channel uncertainty leveraging the SDR method under various CSI uncertainties. Finally, we develop a PAPR-aware precoding solution for intercell coordination. Using numerical analysis, we demonstrate how our approaches reduce the PAPR of BS antennas for single cells and multiple cells serving cell-center and cell-edge users. Numerical results have been conducted to illustrate the performance and flexibility of the proposed framework. The contribution can be summarized as follows:

1. Basic PAPR-aware precoding optimization: We first formulate an optimization framework to minimize instantaneous transmit power with the assumption of perfect CSI to achieve a predefined PAPR and MUI at the transmitter and receiver, respectively. To circumvent the nonconvexity, we *approximate the solution using SDR and apply rank reduction approach to derive the rank-1 optimal solution.*

2. Robust PAPR-aware precoding optimization: We then incorporate both bounded and statistical CSI errors into a more realistic PAPR-aware precoding design by considering robust optimization techniques for *coarse robust precoding, fine robust precoding via S-*

procedure and fine robust precoding via Bernstein-type inequality. We show that the PAPR-aware robust precoding problems that incorporate the channel uncertainties can be formulated as SDR problems and efficiently be solved.

3. PAPR-aware precoding for intercell coordination: We also show the importance of PAPR-aware precoding to serve cell-edge users in three typical scenarios of intercell coordination, *coherent transmission, fast cell selection, and interference coordination and formulate them as SDR problems.* Finally, we show reductions in cell-edge interference and compare the computational complexities between the baseline and a variety of proposed approaches.

6.4 Summary for DPD in Massive MIMO

This work introduces a low-complexity predistorter for nonlinear PAs with memory that is computationally practical for massive MIMO systems. By exploiting excess DoFs, the successive refinement filtering-based approach outperforms the conventional approach in terms of accuracy and complexity. Both our simulation and hardware experiment results show the proposed DPD model can linearize a PA better than the standard conventional approach. The proposed solution achieves high performance at low complexity and reduces the NMSE by more than 27dB compared to the same order conventional approach. The contribution can be summarized as follows:

1. We introduce a DPD architecture which exploits the massive DoFs in a large scale MIMO system to significantly reduce the number of basis function in the Volterra series;

2. We cascade the separated adaptive precoding and predistortion in the indirect learning

architecture and introduce an architecture that is aware of the RF crosstalk and able to overcome it;

3. Enabled by this architecture, we propose a method that successively refines the cascaded precoding and DPD solution to avoid explosive increase of basis functions and corresponding increase in computation.

6.5 Future Works

The excess DoFs in DAS and massive MIMO provide very rich topics for future research. Particularly for PAPR-aware DAS, further works include: (1) PAPR-aware power allocation in DAS using semidefinite relaxation, the power allocation problem with individual SINR constraints can be relaxed to be solved by semidefinite programming; (2) PAPR-aware intercell coordination in DAS using geometric programming; (3) Analyze the impact of channel uncertainties and develop robust PAPR-aware DASs. As for PAPR-aware precoding in massive MIMO systems, further works include: (1) PAPR-aware power allocation in massive MIMO systems which is performed in channel coherence time rather than symbol duration; (2) Explore the sparsity of the matrices in the proposed SDR-based approach to further reduce the computational complexity; (3) Apply our proposed approach to commercial mmWave massive MIMO systems by taking the channel model of mmWave band into account; (4) MUI relaxation justification and quantization: Massive MIMO is attractive especially when each antenna element consists of inexpensive hardware. Hardware impairments are known to fundamentally limit the precise control in the high-power regime. Therefore, the MUI allowance without degrading the capacity is introduced in this work as a feature that can be exploited. However, the selection of MUI allowance δ_e is quite intuitive. By fully parameterize and quantize the time-varying hardware impairments (e.g., PA nonlinearities, I/Q-imbalance, phase noise, and quantization errors) in the future works, the MUI allowance can be selected in a more justified manner. Future works for DPD in massive MIMO systems include: (1) Full array hardware implementation of the proposed DPD algorithm with USRPs; (2) Low-complexity cognitive LUT-based DPD, which stores the channel information and the associated DPD parameters to realize a DPD scheme that can quickly adapt to the environment; (3) Integrated PAPR-aware precoding and DPD in massive MIMO systems.

6.6 Journal Publications

Under Review:

Yao, M., Carrick, M., Sohul, M., M, Marojevic, V., Patterson, C. & Reed, J. H., Semidefinite relaxation based PAPR-aware precoding for massive MIMO-OFDM systems, IEEE Transactions on Vehicular Technology, under review

Yao, M., Sohul, M., M, Marojevic, V., & Reed, J. H. Recurrent neural network inspired PAPR-aware downlink massive MIMO-OFDM system, IEEE Transactions on Wireless Communications, under review

Yao, M., Nealy, R., Sohul, M., M, Marojevic, V., &Reed, J. H., A digital predistortion scheme exploiting degree-of-freedom for massive MIMO systems, IEEE Communication Letters, , under review

Yao, M., Sohul, M., M, Marojevic, V., & Reed, J. H. Energy-Efficient Radio Resource Management for 5G New Radio Exploiting Degrees of Freedom, Wireless Personal Communications, under review

Yao, M., Sohul, M., M, Marojevic, V., & Reed, J. H. Sustainable Green Networking: Exploiting Spatial Degree of Freedom towards Energy-efficient 5G Systems, Wireless Networks, under review

Published:

Sohul, M. **Yao, M.**, M., Yang, T., & Reed, J. H. (2015). Spectrum access system for the citizen broadband radio service. IEEE Communications Magazine, 53(7), 18-25.

Sohul, M., **Yao, M.**, M., Ma, X., Imana, E. Y., Marojevic, V., & Reed, J. H. (2016). Next generation public safety networks: a spectrum sharing approach. IEEE Communications Magazine, 54(3), 30-36.

Sohul, M., **Yao, M.**, Abdallah, A, Carrick, M, Marojevic, V., & Reed, J. H. Quality of service assurance-based auction for spectrum sharing systems, Journal of Analog Integrated Circuits & Signal Processing, pp1-14

Bibliography

[1] C. He, G. Y. Li, F.-C. Zheng, and X. You, “Energy-efficient resource allocation in OFDM systems with distributed antennas,” *IEEE Transactions on Vehicular Technology*, vol. 63, no. 3, pp. 1223–1231, 2014.

[2] C. Studer and E. G. Larsson, “PAR-aware large-scale multi-user MIMO-OFDM downlink,” *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 303–313, 2013.

[3] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume, *et al.*, “How much energy is needed to run a wireless network?,” *IEEE Wireless Communications*, vol. 18, no. 5, pp. 40–49, 2011.

[4] T. Group *et al.*, “Smart 2020: Enabling the low carbon economy in the information age,” *Global e-Sustainibility Initiative (GeSI), Report at http://www. gesi. org*, 2008.

[5] ITU, “Report on the twenty-first meeting of working party 5D,” 2015.

[6] T. Pappaport, A. Annamalai, R. Buehrer, *et al.*, “Wireless communications: past events and a future perspectire,” *IEEE Communications Magzine*, vol. 5, pp. 148–161, 2002.

[7] T. Jiang and Y. Wu, “An overview: peak-to-average power ratio reduction techniques for OFDM signals,” *IEEE Transactions on Broadcasting*, vol. 54, no. 2, p. 257, 2008.

[8] W. Vereecken, W. Van Heddeghem, M. Deruyck, B. Puype, B. Lannoo, W. Joseph, D. Colle, L. Martens, and P. Demeester, “Power consumption in telecommunication networks: overview and reduction strategies,” *IEEE Communications Magazine*, vol. 49, no. 6, pp. 62–69, 2011.

[9] S. K. Mohammed and E. G. Larsson, “Single-user beamforming in large-scale MISO systems with per-antenna constant-envelope constraints: The doughnut channel,” *IEEE Transactions on Wireless Communications*, vol. 11, no. 11, pp. 3992–4005, 2012.

[10] C. Xiong, G. Y. Li, S. Zhang, Y. Chen, and S. Xu, “Energy-efficient resource allocation in OFDMA networks,” *IEEE Transactions on Communications*, vol. 60, no. 12, pp. 3767–3778, 2012.

[11] H. Zhu, “Performance comparison between distributed antenna and microcellular systems,” *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 6, pp. 1151–1163, 2011.

[12] C. He, B. Sheng, P. Zhu, X. You, and G. Y. Li, “Energy-and spectral-efficiency tradeoff for distributed antenna systems with proportional fairness,” *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 5, pp. 894–902, 2013.

[13] D. Mirri, G. Luculano, F. Filicori, G. Pasini, G. Vannini, and G. Gabriella, “A modified volterra series approach for nonlinear dynamic systems modeling,” *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 49, no. 8, pp. 1118–1128, 2002.

[14] P. Gilabert, G. Montoro, and E. Bertran, “On the wiener and hammerstein models for power amplifier predistortion,” in *Microwave Conference Proceedings, 2005. APMC 2005. Asia-Pacific Conference Proceedings*, vol. 2, pp. 4–pp, IEEE, 2005.

[15] H. Ochiai and H. Imai, “Performance of the deliberate clipping with adaptive symbol selection for strictly band-limited OFDM systems,” *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 11, pp. 2270–2277, 2000.

[16] G. Ren, H. Zhang, and Y. Chang, “A complementary clipping transform technique for the reduction of peak-to-average power ratio of OFDM system,” *IEEE Transactions on Consumer Electronics*, vol. 49, no. 4, pp. 922–926, 2003.

[17] S. B. Slimane, “Reducing the peak-to-average power ratio of OFDM signals through precoding,” *IEEE Transactions on Vehicular Technology*, vol. 56, no. 2, pp. 686–695, 2007.

[18] S. Gazor and R. AliHemmati, “Tone reservation for OFDM systems by maximizing signal-to-distortion ratio,” *IEEE Transactions on Wireless Communications*, vol. 11, no. 2, pp. 762–770, 2012.

[19] A. Aggarwal and T. H. Meng, “Minimizing the peak-to-average power ratio of OFDM signals using convex optimization,” *IEEE Transactions on Signal Processing*, vol. 54, no. 8, pp. 3099–3110, 2006.

[20] A. Alavi, C. Tellambura, and I. Fair, “PAPR reduction of OFDM signals using partial transmit sequence: an optimal approach using sphere decoding,” *IEEE Communications Letters*, vol. 9, no. 11, pp. 982–984, 2005.

[21] J.-C. Chen, "Partial transmit sequences for PAPR reduction of OFDM signals with stochastic optimization techniques," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 3, pp. 1229–1234, 2010.

[22] M. Breiling, S. H. Müller-Weinfurtner, and J. B. Huber, "SLM peak-power reduction without explicit side information," *IEEE Communications Letters*, vol. 5, no. 6, pp. 239–241, 2001.

[23] C. Y. Wong, R. S. Cheng, K. B. Lataief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 10, pp. 1747–1758, 1999.

[24] H. Zhu and J. Wang, "Chunk-based resource allocation in OFDMA systems-part I: chunk allocation," *IEEE Transactions on Communications*, vol. 57, no. 9, pp. 2734–2744, 2009.

[25] H. Zhu and J. Wang, "Chunk-based resource allocation in OFDMA systemspart II: joint chunk, power and bit allocation," *IEEE Transactions on Communications*, vol. 60, no. 2, pp. 499–509, 2012.

[26] G. Miao, N. Himayat, Y. G. Li, and D. Bormann, "Energy efficient design in wireless OFDMA," in *Communications, 2008. ICC'08. IEEE International Conference on*, pp. 3307–3312, IEEE, 2008.

[27] Y. Chen, S. Zhang, S. Xu, and G. Y. Li, "Fundamental trade-offs on green wireless networks," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 30–37, 2011.

[28] C. Xiong, G. Y. Li, S. Zhang, Y. Chen, and S. Xu, "Energy-and spectral-efficiency tradeoff in downlink OFDMA networks," *IEEE Transactions on Wireless Communications*, vol. 10, no. 11, pp. 3874–3886, 2011.

[29] C. Li, S. Song, J. Zhang, and K. B. Letaief, "Maximizing energy efficiency in wireless networks with a minimum average throughput requirement," in *Wireless Communications and Networking Conference (WCNC), 2012 IEEE*, pp. 1130–1134, IEEE, 2012.

[30] Y. Wang, W. Xu, K. Yang, and J. Lin, "Optimal energy-efficient power allocation for OFDM-based cognitive radio networks," *IEEE Communications Letters*, vol. 16, no. 9, pp. 1420–1423, 2012.

[31] S. Wang, M. Ge, and W. Zhao, "Energy-efficient resource allocation for OFDM-based cognitive radio networks," *IEEE Transactions on Communications*, vol. 61, no. 8, pp. 3181–3191, 2013.

[32] S. Wang, W. Shi, and C. Wang, "Energy-efficient resource management in OFDM-based cognitive radio networks under channel uncertainty," *IEEE Transactions on Communications*, vol. 63, no. 9, pp. 3092–3102, 2015.

[33] G. Miao, N. Himayat, and G. Y. Li, "Energy-efficient link adaptation in frequency-selective channels," *IEEE Transactions on Communications*, vol. 58, no. 2, pp. 545–554, 2010.

[34] G. Miao, N. Himayat, G. Y. Li, and S. Talwar, "Low-complexity energy-efficient scheduling for uplink OFDMA," *IEEE Transactions on Communications*, vol. 60, no. 1, pp. 112–120, 2012.

[35] D. W. K. Ng, E. S. Lo, and R. Schober, "Energy-efficient resource allocation in OFDMA systems with large numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 11, no. 9, pp. 3292–3304, 2012.

[36] S. Wei, D. L. Goeckel, P. Kelly, *et al.*, "Convergence of the complex envelope of bandlimited OFDM signals," *IEEE Transactions on Information Theory*, vol. 56, no. 10, pp. 4893–4904, 2010.

[37] Y. Jabrane, V. P. G. Jiménez, A. G. Armada, B. A. E. Said, and A. A. Ouahman, "Reduction of power envelope fluctuations in OFDM signals by using neural networks," *IEEE Communications Letters*, vol. 14, no. 7, pp. 599–601, 2010.

[38] I. Sohn, "A low complexity PAPR reduction scheme for OFDM systems via neural networks," *IEEE Communications Letters*, vol. 18, no. 2, pp. 225–228, 2014.

[39] I. Sohn and S. C. Kim, "Neural network based simplified clipping and filtering technique for PAPR reduction of OFDM signals," *IEEE Communications Letters*, vol. 19, no. 8, pp. 1438–1441, 2015.

[40] M. Gardner and S. Dorling, "Artificial neural networks (the multilayer perceptron)a review of applications in the atmospheric sciences," *Atmospheric environment*, vol. 32, no. 14, pp. 2627–2636, 1998.

[41] M. J. Orr *et al.*, "Introduction to radial basis function networks," 1996.

[42] F. J. Pineda, "Generalization of back-propagation to recurrent neural networks," *Physical Review Letters*, vol. 59, no. 19, p. 2229, 1987.

[43] D. W. Tank and J. J. Hopfield, "Simple 'neural' optimization networks: An A/D converter, signal decision circuit, and a linear programming circuit," *IEEE Transactions on Circuits and Systems*, vol. 33, no. 5, pp. 533–541, 1986.

[44] Y. Huang and D. P. Palomar, "Rank-constrained separable semidefinite programming with applications to optimal beamforming," *IEEE Transactions on Signal Processing*, vol. 58, no. 2, pp. 664–678, 2010.

[45] A. Wiesel, Y. C. Eldar, and S. Shamai, "Semidefinite relaxation for detection of 16-QAM signaling in MIMO channels," *IEEE Signal Processing Letters*, vol. 12, no. 9, pp. 653–656, 2005.

[46] A. Mobasher, M. Taherzadeh, R. Sotirov, and A. K. Khandani, "A near-maximum-likelihood decoding algorithm for MIMO systems based on semi-definite programming," *IEEE Transactions on Information Theory*, vol. 53, no. 11, pp. 3869–3886, 2007.

[47] Y.-C. Wang, J.-L. Wang, K.-C. Yi, and B. Tian, "PAPR reduction of OFDM signals with minimized EVM via semidefinite relaxation," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 9, pp. 4662–4667, 2011.

[48] H. Zhang, Y. Yuan, and W. Xu, "PAPR reduction for DCO-OFDM visible light communications via semidefinite relaxation," *IEEE Photonics Technology Letters*, vol. 26, no. 17, pp. 1718–1721, 2014.

[49] W.-K. Ma, P.-C. Ching, and Z. Ding, "Semidefinite relaxation based multiuser detection for M-ary PSK multiuser systems," *IEEE Transactions on Signal Processing*, vol. 52, no. 10, pp. 2862–2872, 2004.

[50] W.-K. Ma, T. N. Davidson, K. M. Wong, Z.-Q. Luo, and P.-C. Ching, "Quasi-maximum-likelihood multiuser detection using semi-definite relaxation with application to synchronous CDMA," *IEEE Transactions on Signal Processing*, vol. 50, no. 4, pp. 912–922, 2002.

[51] M. X. Goemans and D. P. Williamson, "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming," *Journal of the ACM (JACM)*, vol. 42, no. 6, pp. 1115–1145, 1995.

[52] N. D. Sidiropoulos and Z.-Q. Luo, “A semidefinite relaxation approach to MIMO detection for high-order QAM constellations,” *IEEE Signal Processing Letters*, vol. 13, no. 9, pp. 525–528, 2006.

[53] J.-L. Yu and C.-C. Yeh, “Generalized eigenspace-based beamformers,” *IEEE Transactions on Signal Processing*, vol. 43, no. 11, pp. 2453–2461, 1995.

[54] L. Chang and C.-C. Yeh, “Performance of DMI and eigenspace-based beamformers,” *IEEE Transactions on Antennas and Propagation*, vol. 40, no. 11, pp. 1336–1347, 1992.

[55] J. Huang and A. L. Swindlehurst, “Robust secure transmission in MISO channels based on worst-case optimization,” *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 1696–1707, 2012.

[56] S. A. Vorobyov, A. B. Gershman, and Z.-Q. Luo, “Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem,” *IEEE Transactions on Signal Processing*, vol. 51, no. 2, pp. 313–324, 2003.

[57] B. K. Chalise, S. Shahbazpanahi, A. Czylwik, and A. B. Gershman, “Robust downlink beamforming based on outage probability specifications,” *IEEE Transactions on Wireless Communications*, vol. 6, no. 10, 2007.

[58] A. Pezeshki, B. D. Van Veen, L. L. Scharf, H. Cox, and M. L. Nordenvaad, “Eigenvalue beamforming using a multirank mvdr beamformer and subspace selection,” *IEEE transactions on signal processing*, vol. 56, no. 5, pp. 1954–1967, 2008.

[59] H. Zhi-yong, G. Jian-hua, G. Shu-Jian, and W. Gang, “An improved look-up table predistortion technique for HPA with memory effects in OFDM systems,” *IEEE Transactions on Broadcasting*, vol. 52, no. 1, pp. 87–91, 2006.

[60] Z. Gao and P. Gui, "A look-up-table digital predistortion technique for high-voltage power amplifiers in ultrasonic applications," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 59, no. 7, 2012.

[61] D. R. Morgan, Z. Ma, J. Kim, M. G. Zierdt, and J. Pastalan, "A generalized memory polynomial model for digital predistortion of RF power amplifiers," *IEEE Transactions on signal processing*, vol. 54, no. 10, pp. 3852–3860, 2006.

[62] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20–34, 2010.

[63] L. M. Correia, D. Zeller, O. Blume, D. Ferling, Y. Jading, I. Gódor, G. Auer, and L. Van Der Perre, "Challenges and enabling technologies for energy aware mobile radio networks," *IEEE Communications Magazine*, vol. 48, no. 11, pp. 66–72, 2010.

[64] H. Ochiai and H. Imai, "On the distribution of the peak-to-average power ratio in OFDM signals," *IEEE Transactions on Communications*, vol. 49, no. 2, pp. 282–289, 2001.

[65] P. Vandenameele, L. Van Der Perre, M. G. Engels, B. Gyselinckx, and H. J. De Man, "A combined OFDM/SDMA approach," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 11, pp. 2312–2321, 2000.

[66] S. Thoen, L. Van der Perre, M. Engels, and H. De Man, "Adaptive loading for OFDM/SDMA-based wireless networks," *IEEE Transactions on Communications*, vol. 50, no. 11, pp. 1798–1810, 2002.

[67] M. Fuchs, G. Del Galdo, and M. Haardt, "Low-complexity space–time–frequency scheduling for MIMO systems with SDMA," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 5, pp. 2775–2784, 2007.

[68] T. F. Maciel and A. Klein, "A convex quadratic SDMA grouping algorithm based on spatial correlation," in *Communications, 2007. ICC'07. IEEE International Conference on*, pp. 5342–5347, IEEE, 2007.

[69] E. Hossain, V. K. Bhargava, and G. P. Fettweis, *Green radio communication networks.* Cambridge University Press, 2012.

[70] 3GPP, *TR 36.931 version 9.0.0 Release 9.* May 2011.

[71] Z. Shen, J. G. Andrews, and B. L. Evans, "Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints," *IEEE Transactions on Wireless Communications*, vol. 4, no. 6, pp. 2726–2737, 2005.

[72] W. C. Jakes and D. C. Cox, *Microwave mobile communications.* Wiley-IEEE Press, 1994.

[73] D. Wulich, "Definition of efficient PAPR in OFDM," *IEEE Communications Letters*, vol. 9, no. 9, pp. 832–834, 2005.

[74] S. Schaible and J. Shi, "Fractional programming: the sum-of-ratios case," *Optimization Methods and Software*, vol. 18, no. 2, pp. 219–229, 2003.

[75] Y. Nesterov, A. Nemirovskii, and Y. Ye, *Interior-point polynomial algorithms in convex programming*, vol. 13. SIAM, 1994.

[76] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for gaussian vector multiple-access channels," *IEEE Transactions on Information Theory*, vol. 50, no. 1, pp. 145–152, 2004.

[77] S. Boyd and L. Vandenberghe, *Convex optimization.* Cambridge University Press, 2004.

[78] R. R. Kumar and J. Gurugubelli, "How green the lte technology can be?," in *Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology (Wireless VITAE), 2011 2nd International Conference on*, pp. 1–5, IEEE, 2011.

[79] L. Yang, R. Chen, Y. Siu, and K. Soo, "PAPR reduction of an OFDM signal by use of PTS with low computational complexity," *Broadcasting, IEEE Transactions on*, vol. 52, no. 1, pp. 83–86, 2006.

[80] R. v. Nee and R. Prasad, *OFDM for wireless multimedia communications.* Artech House, Inc., 2000.

[81] J.-J. Fuchs, "Spread representations," in *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on*, pp. 814–817, IEEE, 2011.

[82] Y. Xia, G. Feng, and J. Wang, "A recurrent neural network with exponential convergence for solving convex quadratic program and related linear piecewise equations," *Neural Networks*, vol. 17, no. 7, pp. 1003–1015, 2004.

[83] Y. Zhang, D. Jiang, and J. Wang, "A recurrent neural network for solving sylvester equation with time-varying coefficients," *IEEE Transactions on Neural Networks*, vol. 13, no. 5, pp. 1053–1063, 2002.

[84] Y. Zhang and S. S. Ge, "Design and analysis of a general recurrent neural network model for time-varying matrix inversion," *IEEE Transactions on Neural Networks*, vol. 16, no. 6, pp. 1477–1490, 2005.

[85] Y. Xia and J. Wang, “A recurrent neural network for nonlinear convex optimization subject to nonlinear inequality constraints,” *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 51, no. 7, pp. 1385–1394, 2004.

[86] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, “Energy and spectral efficiency of very large multiuser MIMO systems,” *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1436–1449, 2013.

[87] S. K. Mohammed and E. G. Larsson, “Per-antenna constant envelope precoding for large multi-user MIMO systems,” *IEEE Transactions on Communications*, vol. 61, no. 3, pp. 1059–1071, 2013.

[88] J. Zhang, Y. Huang, J. Wang, B. Ottersten, and L. Yang, “Per-antenna constant envelope precoding and antenna subset selection: A geometric approach,” *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6089–6104, 2016.

[89] S. Mukherjee and S. K. Mohammed, “Constant-envelope precoding with time-variation constraint on the transmitted phase angles,” *IEEE Wireless Communications Letters*, vol. 4, no. 2, pp. 221–224, 2015.

[90] J. Pan and W.-K. Ma, “Constant envelope precoding for single-user large-scale MISO channels: efficient precoding and optimal designs,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 982–995, 2014.

[91] J.-C. Chen, C.-J. Wang, K.-K. Wong, and C.-K. Wen, “Low-complexity precoding design for massive multiuser MIMO systems using approximate message passing,” *IEEE Transactions on Vehicular Technology*, vol. 65, no. 7, pp. 5707–5714, 2016.

[92] H.-S. Cha, H. Chae, K. Kim, J. Jang, J. Yang, and D. K. Kim, “Generalized inverse aided PAPR-aware linear precoder design for MIMO-OFDM system,” *IEEE Communications Letters*, vol. 18, no. 8, pp. 1363–1366, 2014.

[93] H. Bao, J. Fang, Z. Chen, H. Li, and S. Li, “An efficient bayesian PAPR reduction method for OFDM-based massive MIMO systems,” *IEEE Transactions on Wireless Communications*, vol. 15, no. 6, pp. 4183–4195, 2016.

[94] N. Vucic, H. Boche, and S. Shi, “Robust transceiver optimization in downlink multiuser MIMO systems,” *IEEE Transactions on Signal Processing*, vol. 57, no. 9, pp. 3576–3587, 2009.

[95] A. Tajer, N. Prasad, and X. Wang, “Robust linear precoder design for multi-cell downlink transmission,” *IEEE Transactions on Signal Processing*, vol. 59, no. 1, pp. 235–251, 2011.

[96] C. Shen, T.-H. Chang, K.-Y. Wang, Z. Qiu, and C.-Y. Chi, “Distributed robust multicell coordinated beamforming with imperfect CSI: An ADMM approach,” *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2988–3003, 2012.

[97] G. Zheng, K.-K. Wong, and B. Ottersten, “Robust cognitive beamforming with bounded channel uncertainties,” *IEEE Transactions on Signal Processing*, vol. 57, no. 12, pp. 4871–4881, 2009.

[98] J. Wang, G. Scutari, and D. P. Palomar, “Robust MIMO cognitive radio via game theory,” *IEEE Transactions on Signal Processing*, vol. 59, no. 3, pp. 1183–1201, 2011.

[99] E. A. Gharavol, Y.-C. Liang, and K. Mouthaan, “Robust downlink beamforming in multiuser MISO cognitive radio networks with imperfect channel-state information,” *IEEE Transactions on Vehicular Technology*, vol. 59, no. 6, pp. 2852–2860, 2010.

[100] Y. Rong, S. A. Vorobyov, and A. B. Gershman, "Robust linear receivers for multiaccess space-time block-coded MIMO systems: A probabilistically constrained approach," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1560–1570, 2006.

[101] K.-Y. Wang, T.-H. Chang, W.-K. Ma, A. M.-C. So, and C.-Y. Chi, "Probabilistic sinr constrained robust transmit beamforming: A bernstein-type inequality based conservative approach," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pp. 3080–3083, IEEE, 2011.

[102] P.-J. Chung, H. Du, and J. Gondzio, "A probabilistic constraint approach for robust transmit beamforming with imperfect channel information," *IEEE Transactions on Signal Processing*, vol. 59, no. 6, pp. 2773–2782, 2011.

[103] Y. Huang, G. Zheng, M. Bengtsson, K.-K. Wong, L. Yang, and B. Ottersten, "Distributed multicell beamforming with limited intercell coordination," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 728–738, 2011.

[104] M. Sawahashi, Y. Kishiyama, A. Morimoto, D. Nishikawa, and M. Tanno, "Coordinated multipoint transmission/reception techniques for LTE-advanced [coordinated and distributed MIMO]," *IEEE Wireless Communications*, vol. 17, no. 3, 2010.

[105] W. Choi and J. G. Andrews, "The capacity gain from intercell scheduling in multi-antenna systems," *IEEE Transactions on Wireless Communications*, vol. 7, no. 2, 2008.

[106] O. Somekh, B. M. Zaidel, and S. Shamai, "Sum rate characterization of joint multiple cell-site processing," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4473–4497, 2007.

[107] F. Rashid-Farrokhi, K. R. Liu, and L. Tassiulas, "Transmit beamforming and power control for cellular wireless systems," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1437–1450, 1998.

[108] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 1, pp. 18–28, 2004.

[109] O. Somekh, O. Simeone, Y. Bar-Ness, A. M. Haimovich, and S. Shamai, "Cooperative multicell zero-forcing beamforming in cellular downlink channels," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3206–3219, 2009.

[110] N. Michailow and G. Fettweis, "Low peak-to-average power ratio for next generation cellular systems with generalized frequency division multiplexing," in *Intelligent Signal Processing and Communications Systems (ISPACS), 2013 International Symposium on*, pp. 651–655, IEEE, 2013.

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

[112] W. Yu and T. Lan, "Transmitter optimization for the multi-antenna downlink with per-antenna power constraints," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2646–2660, 2007.

[113] A. Man-Cho So, "Probabilistic analysis of the semidefinite relaxation detector in digital communications," in *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms*, pp. 698–711, Society for Industrial and Applied Mathematics, 2010.

[114] A. M.-C. So, Y. Ye, and J. Zhang, “A unified theorem on SDP rank reduction,” *Mathematics of Operations Research*, vol. 33, no. 4, pp. 910–920, 2008.

[115] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge University Press, 2012.

[116] I. Bechar, “A bernstein-type inequality for stochastic processes of quadratic forms of gaussian variables,” *arXiv preprint arXiv:0909.3595*, 2009.

[117] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele, and V. Jungnickel, “Coordinated multipoint: Concepts, performance, and field trial results,” *IEEE Communications Magazine*, vol. 49, no. 2, pp. 102–111, 2011.

[118] X. Zhang, “Matrix analysis and applications,” *Tsinghua and Springer Publishing House, Beijing*, pp. 71–100, 2004.

[119] P. M. Suryasarman and A. Springer, “A comparative analysis of adaptive digital predistortion algorithms for multiple antenna transmitters,” *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 62, no. 5, pp. 1412–1420, 2015.

[120] A. Abdelhafiz, L. Behjat, F. M. Ghannouchi, M. Helaoui, and O. Hammi, “A high-performance complexity reduced behavioral model and digital predistorter for MIMO systems with crosstalk,” *IEEE Transactions on Communications*, vol. 64, no. 5, pp. 1996–2004, 2016.

[121] S. Afsardoost, T. Eriksson, and C. Fager, “Digital predistortion using a vector-switched model,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 4, pp. 1166–1174, 2012.

[122] M. O’droma, S. Meza, and Y. Lei, “New modified saleh models for memoryless nonlinear power amplifier behavioural modelling,” *IEEE Communications Letters*, vol. 13, no. 6, 2009.