**Title**
On Optimizing the Performance of Interference-Limited Wireless Systems



**Author**
Abdelaal, Rana

**Publication Date**
2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

# On Optimizing the Performance of Interference-Limited Wireless Systems

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Electrical Engineering and Computer Science

by

Rana Abdelaziz Mohamed Abdelaal

Dissertation Committee:
Professor Ahmed Eltawil, Chair
Professor Ender Ayanoglu
Professor Hamid Jafarkhani

2017

# DEDICATION

To my beloved parents, my lovely husband, my little daughter, and son

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# ACKNOWLEDGMENTS

I would like to thank my advisor Professor Ahmed Eltawil who supported me during all aspects of my research including financial matters. He created an encouraging and inspiring environment to pursue research, and his broad knowledge significantly helped me execute my research projects. Special thanks go to Alireza S. Behbahani for his insightful discussions. The quality of this work was made possible by his assistance.

I would like to thank Professor Hamid Jafarkhani for his well-structured class in space time coding. He does a decent job teaching diversity techniques and coding concepts. He explains everything clearly, and his class slides have been of a great help.

I would like to thank Professor Ender Ayanoglu for his digital communications course which I took with him. His class is very well organized, and he explained everything elaborately. His notes have been very helpful and useful.

I would also like to express my gratitude to my beloved parents and brothers who have been always standing beside me. Last but not least, I would like to thank my lovely husband Muhammad Abdelghaffar who has always been there for me.

# CURRICULUM VITAE

## Rana Abdelaziz Mohamed Abdelaal

### EDUCATION

**Ph.D in Electrical Engineering and Computer Science** — **2017**
University of California, Irvine — *Irvine, CA*

**M.S. in Electrical Engineering** — **2012**
Cairo University — *Cairo, Egypt*

**B.S. in Electrical Engineering** — **2009**
Cairo University — *Cairo, Egypt*

### PROFESSIONAL EXPERIENCE

**Engineering Intern** — **June 2015–September 2015**
Qualcomm Technologies, Research and Development Department — *San Diego, California*

**Research Engineer** — **June 2013–December 2014**
NTT DOCOMO Innovations, Inc — *Palo Alto, California*

**Research Engineer** — **June 2010–August 2012**
NTRA — *Cairo, Egypt*

**Engineering Intern** — **June 2008–September 2008**
Legend technologies — *Cairo, Egypt*

**Engineering Intern** — **June 2007–September 2007**
Enppi — *Cairo, Egypt*

**Engineering Intern** — **June 2006–September 2006**
Etisalat — *Cairo, Egypt*

### RESEARCH EXPERIENCE

**Research Assistant** — **2012–2017**
University of California, Irvine — *Irvine, California*

**Research Assistant** — **2010–2012**
Cairo University — *Cairo, Egypt*

## TEACHING EXPERIENCE

**Teaching Assistant: Discrete Time Signal and Systems** **2016**
University of California, Irvine *Irvine, CA*

**Teaching Assistant: Senior Design Projects** **2015–2016**
University of California, Irvine *Irvine, CA*

## AWARDS AND HONORS

**The Center for Embedded Computer Systems fellowship** **2015**
University of California, Irvine *Irvine, CA*

**The Student research and travel grant award** **2014**
University of California, Irvine *Irvine, CA*

**NTT DOCOMO, Inc, scholarship** **2014**
NTT DOCOMO, Inc *Palo Alto, CA*

**The Center for Embedded Computer Systems fellowship** **2013**
University of California, Irvine *Irvine, CA*

**The Henry Samueli School of engineering graduate fellowship** **2012**
University of California, Irvine *Irvine, CA*

**Ranked 1st among over 100 students** **2011**
Cairo University, Master of Science in Electrical Engineering, *Cairo, Egypt*

**Honors Degree for Academic Excellence** **2009**
Cairo University, Bachelor of Science in Electrical Engineering *Cairo, Egypt*

## PUBLICATIONS

**Scheduling and Power Adaptation for Wireless Local Area Networks With Full-Duplex Capability** **2017**
Rana A. Abdelaal and Ahmed M. Eltawil, submitted to IEEE transactions of wireless communications

**Optimizing Energy Through Adaptive Bit Width Adjustment on Resistive Associative Processors** **2017**
Rana A. Abdelaal, Hasan Erdem Yantr, Ahmed M. Eltawil, and Fadi J. Kurdahi, submitted to IEEE Transactions on Circuits and Systems I

**Practical Framework for Downlink MU-MIMO for LTE Systems** **2017**
Rana A. Abdelaal, Alireza S. Behbahani, and Ahmed M. Eltawil, IEEE Wireless Communications Letters

**Advanced Base Station Precoding and User Receiver Designs for LTE-Advanced Networks** **2015**
Rana A. Abdelaal, Alireza S. Behbahani, and Ahmed M. Eltawil, IEEE ICNC

**On Optimizing the Performance of Interference-Limited Cellular Systems** **2014**
Rana A. Abdelaal, Alireza S. Behbahani, and Ahmed M. Eltawil, IEEE WTS

**On the Performance of Massive MIMO Cellular Systems with Power Amplifiers** **2014**
Rana A. Abdelaal, Alireza S. Behbahani, and Ahmed M. Eltawil, IEEE WTS

**Optimized Joint Power and Resource Allocation for Coordinated Multi-Point Transmission in LTE-Advanced Systems** **2014**
Rana A. Abdelaal, Khaled Elsayed, and Mahmoud H. Ismail, Wireless Personal Communications

**Joint Scheduling and Resource Allocation with Fairness Based on the Signal-to-Leakage-plus-Noise Ratio in the Downlink of CoMP Systems** **2013**
Rana A. Abdelaal, Khaled Elsayed, and Mahmoud H. Ismail, Wireless Personal Communications

**Cooperative Scheduling, Precoding, and Optimized Power Allocation for LTE-Advanced CoMP Systems** **2012**
Rana A. Abdelaal, Khaled Elsayed, and Mahmoud H. Ismail, IEEE WD

**Resource Allocation Strategies Based on the Signal-to-Leakage-plus-Noise Ratio in LTE-A CoMP Systems** **2012**
Rana A. Abdelaal, Mahmoud H. Ismail, and Khaled Elsayed, IEEE WCNC

# ABSTRACT OF THE DISSERTATION

On Optimizing the Performance of Interference-Limited Wireless Systems

By

Rana Abdelaziz Mohamed Abdelaal

Doctor of Philosophy in Electrical Engineering and Computer Science

University of California, Irvine, 2017

Professor Ahmed Eltawil, Chair

Multi Input Multi Output (MIMO) technology has seen prolific use to achieve higher data rates and an improved communication experience for cellular systems. However, one of the challenging problems in MIMO systems is interference. Interference limits the system performance in terms of rate and reliability. In this thesis, we analyze methods that provide high performance over interference-limited wireless networks such as Long Term Evolution (LTE) and WiFi. In this thesis, we tackle different sources of interference. One of the interference sources is the neighbouring interference, we propose methods that include an optimized solution that models the interference as correlated noise, and uses its statistical information to jointly optimize the base station precoding and user receiver design of LTE systems. We study the benefits of exploiting interference in terms of both probability of error and signal-to-noise ratio (SNR). In addition, we compare the proposed method with the conventional beamforming and maximum ratio combining (MRC).

One of the key challenges to enable high data rates in the downlink of LTE is the precoding and receiver design. We focus primarily on the UE and the base station (BS) processing, particularly on estimating and using the interference resulting from neighboring stations. We propose a receiver design that performs well in the presence of interference. Furthermore, we present a precoding scheme that the BS can use to maximize the signal-to-interference plus noise-ratio (SINR). The proposed algorithm performs well under high speed channels. The limitations of the Minimum

Mean Square Error (MMSE) receiver are discussed and it is used for comparison purposes with the proposed approach. An interference free scenario is used as a benchmark to evaluate the proposed system performance.

Performance of LTE is optimized by tackling practical considerations that affect system performance. We present a suboptimal practical way of estimating the interference and utilizing this information on the processing techniques used at both the UE and the eNodeB sides. We focus on managing both MU-MIMO interference and other cell interference. The proposed study improves system performance even under non-perfect channel knowledge, enabling the throughput gains promised by MU-MIMO.

Along the theme of enhancing spectral efficiency, we In-Band Full-Duplex (IBFD) when used in conjuction with Mu-MIMO. IBFD is very promising in enhancing wireless LANs, where full-duplex access points (APs) can support simultaneous uplink (UL) and downlink (DL) flows over the same frequency channel. One of the key challenges limiting IBFD benefits is interference. We propose a scheduling technique to manage interference in wireless LANs with full-duplex capability. We focus primarily on scheduling UL and DL stations (STAs) that can be efficiently served simultaneously.

Finally, we take a holistic view of performance by considering practical issues related to system performance, namely, a) Interference resulting from the non-linearity of power amplifiers, and b)the trade-offs between system performance and power consumption.

An important topic for practical communication systems is handling the interference due to the power amplifier nonlinearities, especially in Orthogonal Frequency-Division Multiple Access (OFDMA) based communication systems, due to the high peak to average power ratio. This problem becomes more compounded when a large number of PAs is required, as in Massive MIMO for example. In this thesis, we discuss the impact of PAs on cellular systems. We show the constraints that PAs introduce, and we take these constraints into consideration while searching for the optimum set of

transmitter and receiver filters. Moreover, we highlight how Massive MIMO cellular networks can relax PAs constraints resulting in low cost PAs, while maintaining high performance. The performance is evaluated by showing the probability of error curves and signal-to-noise-ratio curves for different transmit powers and different number of transmit antennas.

In terms of power consumption we investigate the use of emerging technologies (such as memristors) to enable highly efficient computation kernels for wireless communication systems. Specifically, we investigate the use of Associative processors (APs) to perform in-memory computation in the context of an FFT processor. To reduce power and power density, we investigate approximate computing in memristive based associative processors. A promising approach to save energy is through reducing the bit width, however reducing the bit width introduces errors that may affect the performance. In this thesis, our goal is to adjust the bit width based on the channel SNR, aiming at achieving good performance at reduced energy consumption. The mathematical approach that analytically describes the system performance under the reduced bit width noise is presented. Based on this model, an adaptive bit width adjustment algorithm is presented that utilizes the received SNR estimates to find the optimal bit width that achieves performance goals at reduced energy consumption. Simulation results show that the proposed algorithms can achieve up to 45% energy savings as compared to wireless communication systems with conventional FFT.

# Chapter 1

# Introduction

Mobile data demand has been rapidly rising over the past few years and it is expected to continue along the same trends due to the increase in data-hungry user equipments (UEs) in their various form-factors. This is especially problematic in dense urban cities which suffer from severe mobile congestion and network stress, thus requiring advanced interference management techniques to better manage the network, improve performance and spectral efficiency. Multiple input multiple-output (MIMO) technology, which uses multiple antennas at both sides of cellular systems, has emerged as a promising technology for achieving high data rates for wireless systems such as Long Term Evolution (LTE) and beyond. The key role which MIMO technology plays in the LTE standards testifies to its significant importance [1]. MIMO techniques can be used to their greatest extent and provide high data rates for negligible interference environments. However, it is much more challenging to provide high data rates in the presence of non-negligible interference. In other words, one of the main factors that limits the performance of MIMO systems is interference. The dimensions of the problem increase when MIMO is extended to simultaneous users in the Multiuser Multiple Input Multiple Output (MU-MIMO) schemes. MU-MIMO refers to serving multiple UEs on the same time-frequency channel resource by eliminating inter-user interference via spatial precoding, and advanced receiver designs. For correct operation of MU-MIMO, it is

essential that: 1) The transmissions intended to different UEs are well separated at the eNodeB side, and 2) The UEs receiver design has to be able to exploit the potentially available information about the interference.

## 1.1 Interference in Wireless Networks

There are several kinds of interference that wireless networks suffer from, including both inter-cell and intra-cell interference.

### 1.1.1 Other cell / Neighboring Interference

Other cell interference is due to users served by neighboring cells over the same time and frequency resources. Of particular importance are UEs located at the cell edge of a wireless system that receive strong interference from neighboring cells. In addition to the strong interference that the cell-edge UEs face, they receive a weak desired signal due to the propagation loss from their home Base station (BS). Thus, in order to enable cell-edge UEs to establish a reliable connection with their BS, it becomes imperative to design precoding and receiver algorithms that can achieve a certain desired performance under the conditions mentioned.

There are several techniques in literature that aim at mitigating interference from neighboring cells, such as [2–9]. Some of these techniques have been proposed for reducing interference by adding overhead at the BS side, such as frequency reuse [2] and BS coordination [3]. Other techniques are focused on the receiver design at the UE side such as the maximal likelihood (ML) receiver [4, 5], linear minimum mean square error (MMSE) receiver [6], Advanced MMSE with successive interference cancellation (MMSE-SIC) [7], and receivers based on nonlinear decision feedback equalizers (DFE), which perform an iterative interference cancellation [8, 9].

Some of these techniques can significantly reduce the interference, however, they suffer from one or more of the following drawbacks: spectral efficiency reduction, inefficient use of the available bandwidth, high latency, high complexity, performance degradation under high interference, and/or the requirement of a large number of iterations to achieve a reliable solution. Thus, part of this thesis is motivated to overcome these drawbacks by using interference statistics to jointly design the precoder and the receiver at the BS and the UE respectively in a closed form solution.

### 1.1.2 Intra-cell / Multi-user MIMO Interference

Intra-cell interference is driven by reusing the same resources for multiple users within the same cell, typically referred to as Multi-User MIMO (MU-MIMO). MU-MIMO is a promising wireless technique where the same time-frequency channel resources are allowed to be used by multiple UEs simultaneously through spatial precoding. The performance of LTE systems critically depends on how the interference either across different cells or due to MU-MIMO is managed.

### 1.1.3 In-band Full Duplex Interference

The explosive growth of wireless data traffic, spurred by data-hungry stations (STAs) such as smart phones and tablets, is draining current wireless LANs resources, requiring a new paradigm shift such as in-band full-duplex (IBFD). Recent work [15–19] has demonstrated IBFD, which is the ability to transmit and receive simultaneously, in the same band, via self-interference cancellation which can be suppressed significantly, enabling near-ideal IBFD capability. To take full advantage of IBFD, it is essential to carefully select the STAs to be served simultaneously, since IBFD leads to new network interference scenarios compared to current half-duplex (HD) communication based Wireless LANs described above.

## 1.2 Interference Challenges in Wireless Networks

### 1.2.1 Interference Estimation and Mitigation

Estimating interference poses considerable challenges in LTE. In this thesis, we study estimating the interference information and utilizing it in both the UE and the eNodeB processing. We propose an interference estimation approach for the design of precoders and receivers in LTE. We estimate the interference covariance matrix and design precoders to maximize signal-to-interference plus noise ratio (SINR). Performance of the proposed approach is benchmarked against the precoding and receiver designs that are currently considered for LTE systems.

**Related Work in Literature**

Interference in wireless networks has been treated to various degrees in existing literature. The current state-of-the-art on interference management techniques encompasses investigations of the throughput achievable by a number of different joint transmission strategies, including Dirty Paper Coding (DPC) [20, 21], coordinated beamforming (linear precoding) [22, 23], and coordinated scheduling [3, 24]. In [23], coordinated beamforming in multicell scenarios was investigated to minimize the total power consumption of all eNodeBs to meet with UEs individual signal-to-interference-plus-noise ratio (SINR) targets, based on the uplink-downlink beamforming duality.

In [3, 25], coordinated processing was investigated, where interference is managed through coordination between several eNodeBs. This imposes restrictions on what resources in time and/or frequency are available to each cell, or what transmit power may be used in a certain time/frequency resource. In another coordinated approach, antennas from multiple eNodeBs act as a single antenna array. This technique requires signal level coordination (exchange of data streams) [26]. This approach is well-known as a solution for strong neighboring interference (cell-edge prob-

lem). However, the promising results for cooperative processing are only achievable if multiple eNodeBs can maintain a sufficiently accurate time and phase synchronization for cooperation, which is generally challenging to implement [27]. A number of issues remain to be addressed before realistically considering multicell processing for future wireless systems, namely: the need for a high-speed backbone enabling information (data, control/synchronization, and channel state) exchange between the eNodeBs, the requirement of channel information availability for coherent methods, and timing/phase synchronization [21].

Interference Alignment (IA) is another interference management approach. IA promises substantial theoretical gains in cellular networks, however, it comes with challenges in implementation. Authors in [28] address interference cancellation using IA only from one neighboring eNodeB, which performs well in a two-cell layout. Another interference management technique at the eNodeB side is frequency reuse [2]. The concept of frequency reuse is that each cell divides the available bandwidth into a group of frequency bands. Each cell chooses a certain frequency band to use for its cell-edge users, such that there are no neighboring cells using the same frequency band for their cell-edge users. Frequency reuse can effectively reduce interference by spacing the competing transmissions farther away. On the other hand, spectral efficiency is reduced since a portion of the available spectrum is not used by each cell [29]. As the number of data-hungry users in a cellular system increases, the demand for bandwidth increases. Thus, conventional frequency reuse techniques that are based on spectrum partitioning are not promising as a long term solution. Careful management of interference is important in systems such as LTE, which are designed to operate with a frequency reuse factor of one.

Other techniques are focused on the receiver design such as maximum likelihood detection (ML) [4, 5], which is known to minimize the bit error rate (BER) in multiple antenna systems. However, it requires accurate, instantaneous information on the channels of interference which is not possible for LTE. Furthermore, the drastically increased computational complexity of the ML scheme makes it prohibitive in practice, especially for UEs. Reduced-complexity algorithms such as V-

BLAST [30] perform much worse than ML. The sphere decoder (SD) has been proposed as an alternative to ML [31]. SD can provide ML performance with reduced complexity by providing an efficient way for generating all candidate solutions. The main idea of the SD algorithm is to enumerate lattice points that lie inside a sphere defined by the channel matrix and the received signal vector. In the high signal-to-noise ratio (SNR) region, the radius of the sphere can be chosen small enough so that only few candidates are found inside the sphere. This search space is therefore drastically smaller than the ML search space [32].

Although the average complexity of SD algorithms is believed to be polynomial for small array sizes [33], the actual complexity depends on the channel conditions and the noise level, making it difficult to integrate in an actual system, where data needs to be processed at a constant rate (i.e. fixed complexity/throughput). Different methods have been proposed to reduce or limit the complexity of the SD, however, most of them still have a variable complexity depending on the channel conditions. They can be classified in the following categories: 1) Modifications of the algorithm to marginally reduce the complexity requiring additional operations or the calculation of limiting thresholds [34, 35]. 2) Simplifications of the algorithm for specific constellation types [36]. An alternative to SD is the K-best decoders which maintain a fixed throughput, at a performance penalty especially at lower SNRs [37].

Another alternative to the ML receiver is the MMSE receiver [6]. There are two classes of MMSE receivers, a simple one that only needs to know the average interference and an advanced one that needs to know the accurate interferers channels MMSE-SIC [7]. The linear MMSE receiver with only average interference knowledge does not perform well, it has a slight gain compared to the zero forcing receiver that performs poorly due to noise enhancement. In contrast, the advanced MMSE-SIC has relatively good performance since it is able to cancel the interference. However, the performance of the advanced MMSE degrades with the number of interferers. And in LTE, the number of interferers can increase dramatically. For instance, cell-edge UEs significantly suffer from strong interference due to non-negligible neighboring cells. If there are $K$ non-negligible

neighboring BSs, that results in $NK$ interfering signals if each BS has $N$ transmit antenna. Such interference can dramatically affect the performance. Another technique to manage interference is the use of nonlinear decision feedback equalizers (DFE), which performs an iterative interference cancellation as discussed in [8], [9]. Although this technique achieves high performance, the interference is canceled iteratively. Thus, to increase the reliability of the data, a high number of stages/iterations is required.

A promising technique that has been studied in the LTE context is the interference rejection combining receiver (IRC). IRC is a linear combining technique that relies on estimate of the interfering channels to project the received signals on a subspace in which the Mean Square Error (MSE) is minimized [38]. IRC is attractive given that it represents a straightforward add-on to the known Minimum Mean Square Error (MMSE) receiver, which is now considered the baseline receiver in LTE networks [39]. In order to perform near ideal interference suppression, IRC (also known as MMSE-IRC) requires channel estimation and covariance matrix estimation including the interference with high accuracy [39], however, accurate interference knowledge is difficult to get at the UEs due to estimation errors.

To understand this better, we start by making an important observation. The accuracy of the interference covariance matrix estimation in LTE depends on the cross-covariance between the signal of the serving cell and the interfering cells. This is because the interference estimation is done over the data resources.

**Managing non-linearity**

As for managing the interference due to PA nonlinearities, there are some techniques that have been proposed in literature for reducing the PAPR in OFDM systems such as selected mapping, coding techniques, and clipping [40–43]. The first two concepts are not applicable in the context of LTE. Selected mapping requires additional signaling, while coding techniques are not compatible

with the data scrambling used in the LTE downlink. Clipping is a simple technique, where the transmitted signal is clipped to a predefined level to avoid PA distortion. Depending on the linear region, clipping may lead to significant power loss in amplifiers with narrow linear regions.

### 1.2.2 IBFD and MU-MIMO

Both IBFD and MU-MIMO are promising technologies that can provide rate enhancements in wireless LANs. MU-MIMO allows an access point (AP) to send multiple frames to multiple STAs at the same time over the same frequency resources. For correct operation of MU-MIMO, it is essential that the transmissions intended to different STAs be well separated via means of spatial precoding.

#### Self interference

When the UL receiver and the DL transmitter are active in the same AP simultaneously, self-interference is generated. The self-interference problem is out of this thesis scope since it has been extensively studied in the literature as will be discussed.

#### Network Interference

However, when the UL AP is different than the DL AP, network interference is generated. For example, if we have a STA transmitting a packet to an APs as an UL flow, and the AP is transmitting packets to another DL STA, as the DL flow. In this case, the signal transmitted from the UL STA can interfere with the DL STA, which intend to receive the signal from the AP. If the UL STA is located close to the DL STA, and the signal transmitted from the UL STA is very strong, the DL STA will face high interference.

**MU-MIMO Challenges**

The key challenge for MU-MIMO with IBFD is to coordinate multiple downlink (DL) and uplink (UL) simultaneous transmissions which are made possible by the full-duplex capability. This thesis focuses partially on STAs scheduling at both the DL and the UL aiming at improving the sum rate in MU-MIMO wireless LANs with IBFD capability. We consider wireless LANs consisting of APs that are capable of full-duplex communications. We aim at managing interference, including interference due to DL MU-MIMO flows and interference due to the UL flow. To overcome the challenge of interference, we propose a scheduling technique that aims at identifying a group of DL STAs along with an UL STA to be served simultaneously with minimal interference. Furthermore, the UL power is adjusted to maximize the resulting sum throughput.

**Related Work in Literature**

Managing interference due to IBFD has been studied in the existing literature. Recently, several publications [44]-[61] have considered the problem of self-interference cancellation in full-duplex systems by investigating different self-interference cancellation techniques to mitigate the self-interference signal.

Analog cancellation is necessary to obtain preliminary isolation to avoid RF compression and saturation of the analog to digital converters [44]. Analog cancellation uses knowledge of the transmission to cancel self-interference in the RF signal, before it is digitized. One approach to analog cancellation uses a second transmit chain to create an analog cancellation signal from a digital estimate of the self-interference [45]. Another approach is that the transmit signal is tapped at the transmit antenna feed, processed in the analog-circuit domain, and subtracted from the receive-antenna feed in order to cancel self-interference [46]. In [47], authors propose a design that utilizes a copy of the transmitted analog signal and uses a transformer in the analog domain to then create a perfectly inverted copy of the signal. The inverted signal is then connected to a

circuit that adjusts the delay and attenuation of the inverted signal to match the self interference that is being received on the receiver antenna from the transmitter antenna.

On the other hand, digital domain cancellation is based on the subtraction of the interference signal. Digital cancellation techniques aim to cancel self-interference after the analog-to-digital converter (ADC) [48]-[50]. Several experimental and analytical results show that the mitigation capability of digital cancellation techniques is very limited, mainly due to the transmitter and receiver radio circuits' impairments [51]-[54].

The self-interference signal could also be suppressed in the propagation-domain. In propagation-domain suppression techniques [55]-[59], the self-interference signal is suppressed before it is processed by the receiver circuitry. Propagation-domain self-interference suppression mitigates both the self-interference signal and the transmitter noise associated with it. In addition, mitigating the self-interference signal in the propagation domain decreases the effect of the receiver noise and increases the dynamic range allocated for the desired signal. Authors in [60, 61] propose antenna cancellation techniques that, when combined with digital and analog techniques, allow IBFD with negligible self interference. The above studies considered that the STA that is being served on DL is also the STA that is sending UL packets to the AP. In other words, the AP will act as a transmitter to a certain STA and also a receiver to the same STA. Thus, the interference in such situation is purely self interference. Network interference among STAs will occur if different STAs are considered for DL and UL, which may significantly deteriorate the throughput performance of IBFD wireless LANs since multiple STAs are allowed to transmit and/or receive simultaneously. In order to mitigate the interference problem arising in such environment, some studies have been performed to coordinate transmissions with the goal of reducing network interference [62]-[72].

New medium access control (MAC) protocols proposed in [62–70] capture additional transmission opportunities created by full-duplex by modifying contention and back-off mechanisms. In [62], the authors develop a centralized MAC protocol to support asymmetric data traffic where network nodes may transmit data packets of different lengths, and they propose to mitigate the hidden node

problem by employing a busy tone. To overcome this hidden node problem, authors propose to adapt the 802.11 MAC protocol with the RTS/CTS handshake. In [70], authors study the power allocation for IBFD system where STAs operate in the HD mode but the AP communicates by using the FD mode. In [70], the system model considers a single AP and multiple STAs. The UL STA is chosen randomly, then a DL STA with low interference from the UL STA and high received power from the AP is selected. Afterwards, a power control algorithm is used such that the DL SINR and UL SINR satisfies a threshold. [70, see section III]

In contrast to full-duplex MAC protocols, there have been a few efforts to redesign the scheduling algorithms for full-duplex wireless networks while taking network interference into consideration. One approach was studied in [71], such that the AP has a pre-determined DL STA and it aims at scheduling another UL STA simultaneously. The AP randomly picks an UL STA out of several ones that achieve a specific signal to interference (SIR) threshold at the DL STA. Simulations presented show throughput gains.

### 1.2.3 Power Amplifier Nonlinearity

Although there are different multiple access technologies competing for dominance, cellular network standards prefer Orthogonal Frequency-Division Multiple Access (OFDMA) which is the multi-user version of the orthogonal frequency-division multiplexing (OFDM) for its well known advantages. OFDM is highly resistant to frequency selective fading. Moreover, with OFDM technology channel equalization becomes simple. On the other hand, one of the major drawbacks is that the OFDM signal has a high Peak-to-Average Power Ratio (PAPR). Generally, the OFDM transmitter can be seen as a linear transform performed over a large block of independently identically distributed quadrature amplitude modulation (QAM) complex symbols (in the frequency domain). From the central limit theorem [10], the time-domain OFDM symbol may be approximated as a Gaussian waveform. The amplitude variations of the OFDM modulated signal can

therefore be very high. However, practical power amplifiers (PAs) of RF transmitters are linear only within a limited dynamic range. Thus, the OFDM signal is likely to suffer from nonlinear distortion, which results in interference. To avoid such distortion, PAs have to operate with large power back-offs, leading to inefficient amplification and/or expensive transmitters. Typically the power amplifies consumes 50-80% of the power budget of a BS. The PA energy efficiency depends on the frequency band, modulation and operating environment [11]. Modern BSs are highly inefficient because of their need for PA linearity and high PAPR. OFDM schemes commonly used in communication standards such as High Speed Packet Access (HSPA) and LTE are characterized by strongly varying signal envelopes with PAPR that exceeds 10 dB [12]. Along these lines, the width of the backoff region needs to be very wide which reduces the linear region of the PA. Thus PAs operate well below saturation, resulting in poor power efficiency, excessive cost, and size [13]. Otherwise, the use of low-cost non-linear PA can result in the presence of nonlinear interference and spectral spreading of the transmitted signal which can cause adjacent channel interference, and signal constellation deformation and spreading [14].

## 1.3 Practical Adaptive Power Management

### 1.3.1 Adaptive Bit Width Adjustment

Over the last decade, the world has seen a sharp increase in data traffic that necessitates robust, low-power processing cores. However, mobile computing based on traditional architectures is approaching its limits in terms of scalability and power consumption. One means of achieving the desired performance increase is by increasing parallelism rather than depending on transistor feature reduction. This approach also becomes limited if processing elements cannot consume data from memory at the desired processing rate, leading to a significantly degraded overall performance. To address that limitation, new computing paradigms started to emerge that focus more on

the memory bottleneck problem. Theoretically, the most memory efficient paradigm is in-memory computation. This paradigm simply replaces the logic with memory structures, virtually eliminates the need for memory load/store operations during computation.

*Associative processors* (AP) are promising computational platforms for massively in-memory parallel computing. Associative processors can be considered as a type of Single Instruction Multiple Data (SIMD) processors that combine the memory and processor in the same location, so that every row in the memory behaves as an individual processor. Since an operation can be performed on all memory words in parallel, the execution time of an operation does not depend on the vector size. Many parallel systems are uniquely suited to this approach due to the vector based nature of their processing pipelines. This feature largely overcomes the memory-wall problem of traditional Von Neumann architectures since there is no inter-dependance between memory and processor. Associative processing is not a new topic and numerous architectures of associative processors (AP) originated in the 1970's and 1980'; however, in the past, the adoption of APs was limited due to the unmanageable power and area requirements. This reality is changing with the availability of new semiconductor technologies and materials that allow for extremely dense memory structures such as memristor and STT-RAM, leading to a resurrection of the this approach under the name of *Resistive Associative Processor* (RAP).

Another computing paradigm that has become well-known in the recent years is *Approximate Computing*. In approximate computing, the goal is the exploiting the error resiliency by relaxing correctness constraints to achieve the energy efficiency. In a system, approximate computation can be introduced at three different levels: design level, algorithm-architecture level, and logic-circuit level. In the circuit level, the most common method is designing functionally approximate circuits that has lower components than its fully accurate counterpart. Another ways of hardware approximation are overscaling the circuit timing and/or voltage and approximation in memory. At the architecture level, the significant components in the overall system is favored over insignificant ones. In the design level, the approximate computing can be realized by design tools that sup-

ports the approximate computing. For example, a VLSI design software can include approximate versions of some arithmetic circuits and these circuits can be used in error resilient parts of the chip.

Even though RAP architectures promise very efficient parallel computing achievements, there are serious problems of large power density and energy consumption in such architectures mainly due to high switching activity and costly memristor energy. Unless these problems are addressed, it is likely that these architectures cannot be practical. On the other hand, application of approximate computing onto the existing computation systems does not eliminate the aforementioned problems of the traditional computing fully even though it is a rising star in low energy computation. Fortunately, AP architectures inherently facilitates the approximate computing since all computations are performed on per bit basis. Regarding the problems of dark silicon era, combination of associative processing with approximate computing can be a promising approach for the future of computing especially for communication systems. To the best of our knowledge, no prior study has touched on the approximate in-memory computing.

In this study, we introduce the approximate in-memory computation by exploiting the resistive associative processors (RAP) in communication systems. The goal is to replace logic with memory structures, virtually eliminating the need for memory load/store operations during computation together bit dynamic approximate computing in algorithm-architecture level for both energy and performance efficiency. The suitability of resistive associate processors for approximate computing is demonstrated through the implementation of Fast Fourier Transform used in MIMO-based wireless communication system. Results show that approximate in-memory computation in RAPs provides the considerable energy saving by the way of approximation in a reasonable level together with performance gain.

## 1.4 Thesis Contribution

MIMO promises higher spectral efficiency for LTE systems [73]. In order to achieve improved spectral efficiency, channel state information at the transmitter (CSIT) is needed. If CSIT is unavailable, then finding the optimal precoding is not possible, and the transmitter will only have one of two options, either use its total power for a single antenna and turn off the rest of the antennas, or divide the total power equally over the available antennas [74]. In this paper, we assume that CSIT is available and it can be obtained from the UEs UL pilots by operating in Time-Division Duplexing (TDD) mode and exploiting the reciprocity of the radio channels.

We focus on the processing operations at the base station (BS) and at the UE. For interference free transmission, it is well known that singular value decomposition (SVD) beamforming, also known as maximum ratio transmission (MRT)- maximum ratio combining (MRC) provides the optimum performance. We consider practical systems where interference exists and the UE processing stage performs interference estimation (due to neighboring cells) in order to be used effectively in the receiver design. In contrast, the BS processing main role is to provide precoding that maximizes the achievable rate for the UE which is through maximizing its signal-to-interference plus noise-ratio (SINR). Towards that goal, we make the following general assumptions 1) The transmissions intended to UEs need to be channel dependent which requires channel knowledge (TDD reciprocity), and 2) The UEs receiver design need to be able to make sufficient use of the interference knowledge.

In this thesis, we propose a practical method for interference estimation in LTE systems. We study estimating the interference and utilizing this information in the processing operations at both the UE and the eNodeB sides. The UE performs interference estimation in order to be used effectively in the receiver design. In contrast, the eNodeB processing main role is to spatially separate multiple DL transmissions while maximizing the SINR. Our contribution in this topic can be summarized as follows:

- *A practical framework for interference estimation:* We estimate the interference covariance matrix in OFDMA-based LTE context through the use of non-data resources.
- *Utilizing the interference knowledge:* We derive suboptimal simple and non-iterative precoder and receiver aiming at enhancing the system throughput for MU-MIMO operation.
- *Considering the Non-ideality of MU-MIMO precoding:* We tackle the effect of non-ideal precoding for MU-MIMO in LTE networks. The solution aims at updating the interference covariance matrix to add the effect of the MU-MIMO interference.
- *Evaluation of the proposed framework benefits:* We evaluate the proposed framework by comparing different designs. We demonstrate the substantial gains compared to existing approaches.

Moreover, we extend our work to the area on IBFD systems. And with the aim of improving the performance even further, we take use of MU-MIMO on top of IBFD. One of the key shortcomings of the IBFD research noted above is that it does not optimize the STAs selection process. In prior work, any STA that achieve a specific SIR at the DL STA is considered a good candidate. Although, this type of optimization provides a guaranteed minimum throughput, it does not maximize the throughput. Moreover, in such schemes, finding a STA with the satisfying SIR condition is done via exhaustive search over all the STAs, which might be time consuming. Furthermore, none of the prior work discussed scheduling multiple DL transmissions along with an UL transmission, which is needed in practical crowded wireless LANs that serve many STAs simultaneously.

In this thesis, we consider practical wireless LANs in which it is not necessary to perform an exhaustive search over all the STAs to find good candidates to schedule. Our main objective here is to maximize the achievable rate of wireless LANs with full-duplex APs serving multiple DL STAs via MU-MIMO and an UL STA. Also, we consider that the UL STA is not necessarily one of

the DL STAs. In other words, each STA may or may not be served on UL and DL simultaneously. Clearly, MAC protocols will be required to support the required functionality, however these will be the subject of future research.

The contributions of this topic can be summarized as follows:

- *STAs Categorization:* Categorizing STAs aiming at reducing the search space.
- *STAs Scheduling:* Scheduling simultaneous UL and DL flows aiming at maximizing the overall rate via careful STAs selection.
- *Power Adjustment:* Adjusting the UL transmit power that directly affects the DL-STAs, aiming at reducing the interference over DL flows while achieving reasonable UL rate.
- *Evaluation of the proposed techniques:* Evaluate the performance by showing the achievable rate as compared to other IBFD techniques and conventional half-duplex system.

Recently, there has been significant interest in promoting the concept of green communication, where inefficiencies are reduced across the entire network, including, of course, at the BS. Towards that goal, we extend our work and study the PAs behavior in LTE systems, and jointly design the BS transmit vectors and the UEs receiver filters to prevent PA distortion, leading to higher efficiencies. Furthermore, we extend the discussion to include Massive MIMO, where the number of antennas is increased. The contributions in this topics are:

- *PA effects study:* Discuss the impact of PAs on cellular systems.
- *Analyze PAs constraints:*Show the constraints that PAs introduce.
- *Design of transmitter and receiver filters:* Take PA constraints into consideration while searching for the optimum set of transmitter and receiver filters.

- *Analyze Massive MIMO benefits:* Highlight how Massive MIMO cellular networks can relax PAs constraints resulting in low cost PAs, while maintaining high performance.
- *Evaluation of the proposed framework benefits:* Evaluate the performance by showing the probability of error curves and SNR curves for different transmit powers and different number of transmit antennas.

Another promising topic is the approximate in-memory computation concept with the goal enhancing both energy and performance efficiency. The contributions in this topic are:

- *RAP study:* Introducing approximate in-memory computation concept by exploiting the resistive associative processors (RAP) in communication systems.
- *Approximate computing for communication systems:* The suitability of RAPs for approximate computing is demonstrated in the field of communication systems
- *Mathematical analysis:* A novel mathematical model that characterizes system performance of FFT under fractional bits truncation has been derived.
- *Adaptive bit width adjustment:* An adaptive bit width adjustment algorithm has been proposed.
- *Evaluation of the proposed framework benefits:* Simulation results show that by using the proposed adaptive bit width algorithm, we can achieve up to 45% of energy savings with very slight performance degradation.

## 1.5 Thesis Organization

The rest of the thesis is organized as follows. Chapter 2 presents a precoding and receiver design for LTE system with single user MIMO such that the interference in this case is always limited to other cell interference. In chapter 3 we add MU-MIMO operation and study its effects. We explain some practical considerations regarding interference estimation, precoders, and receivers. In chapter 4 we introduce a new look at distributed MU-MIMO with IBFD capability. We discuss the challenges, and provide solutions for users scheduling and power control. Chapter 5 extends the previous results to include PA constraints to avoid PA distortion. And Chapter 6 includes the adaptive bit width adjustment algorithm aiming at enhancing the energy/power efficiency of communication systems.

# Chapter 2

# Managing Other Cell Interference in LTE Networks

## 2.1 Introduction

A major challenge in the development of next generation wireless networks is to design techniques that can provide high throughput over interference limited cellular networks. This chapter focuses on designing a technique that utilizes the interference knowledge to enhance the system performance.

In this chapter, we present a joint precoding and receiver (JPR) design, where each BS does not need to sacrifice a portion of the available bandwidth as the case presented in [2]. Also, unlike the algorithm shown in [3], JPR does not require cooperation between BSs which will definitely reduce the latency. Moreover, the implementation of JPR at the UE side has relatively lower complexity than [4, 5]. Furthermore, JPR performance does not degrade with increasing the number of interferers as the case studied in [6, 7]. Furthermore, JPR provides a closed form solution unlike the iterative solution presented in [8, 9].

### 2.1.1 Contributions

The main results of this chapter are:

1. We develop a model for the interference as correlated noise.
2. We present a jointly optimized solution for designing the precoder and the receiver with given interference covariance matrix.
3. We compare our algorithm with the conventional MRT/MRC technique.
4. We discuss interference covariance matrix estimation in LTE systems.
5. We present a jointly optimized solution for designing the precoder and the receiver for LTE systems
6. We evaluate the designs in terms of throughput for different practical channel models.

The remainder of the chapter is organized as follows: Section 2.2 describes the problem formulation. In Section 2.3, we provide a solution aiming at maximizing the SNR of the received signal. In Section 2.4, we explain rate maximization for multi-carrier systems. In section 2.5, we discuss system assumptions and discuss the LTE context specifications. In Section 2.6, we reformulate the problem according to LTE standard. In Section 2.7, We explain the interference covariance matrix estimation. Section 2.8 provides some remarks and notes about the interference covariance matrix and the channel models. Simulation results are provided in Section 2.9 and we conclude the chapter in Section 2.10.

### 2.1.2 Notation

We use bold lower case for vectors, such as $\boldsymbol{a}$, while bold capital letters are used for matrices such as $\boldsymbol{A}$. Further $||\boldsymbol{A}||$ stands for the norm of the matrix $\boldsymbol{A}$. Further $(\cdot)^H$ stands for Hermitian

transposition. $[\boldsymbol{A}]_{i,j}$ denotes the element in row $i$ and column $j$ of matrix $\boldsymbol{A}$. The cardinality of the set $A$ is denoted by $|A|$. Also $\boldsymbol{E}$ stands for expectation operator.

## 2.2 Problem Formulation

### 2.2.1 System Model

We consider an LTE system where each BS is equipped with *N* antennas and each UE is equipped with *M* antennas, and there are *K* interfering BSs. The system model is shown in Fig. 2.2, where $I_k$ is the interference power of the $k^{th}$ neighbor BS.

In contrast to the prior work on interference management, this paper focuses on using the interference to enhance performance. We show that by taking interference into consideration in jointly optimizing the BS precoding and the UE receiver design, we can actually achieve higher signal-to-noise-ratio (SNR) compared to the case of using the conventional maximum ratio transmission (MRT) and maximum ratio combining (MRC).

In our algorithm, we first model the interference as correlated noise, then the optimization problem for the precoder and the receiver is formulated. The solution to that optimization problem is not well-known since most of the previous studies on MIMO systems assume uncorrelated Gaussian noise. However, noise can be correlated due to the presence of interference [75], as described in details in the following section, which presents the joint design of the precoder and receiver.

### 2.2.2 Noise Correlation

Since the UE suffers from strong interference in certain directions, JPR plays an essential role in enhancing the system performance. The received signal at the UE under consideration can be

denoted as:

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \sum_{j=1}^{K} \boldsymbol{G}_j \boldsymbol{i}_j + \boldsymbol{n}, \tag{2.1}$$

where $\boldsymbol{H}$ is the channel between the home BS and the UE, $\boldsymbol{G}_j$ is the channel between the $j$th interferer and the UE, $\boldsymbol{i}_j$ is the signal from the $j$th interferer, $\boldsymbol{y}$ is the received signal, $\boldsymbol{x}$ is the transmitted signal, and $\boldsymbol{n}$ is the additive noise with a covariance matrix given by $\boldsymbol{R}_n = \boldsymbol{E}[\boldsymbol{n}\boldsymbol{n}^{\boldsymbol{H}}]$. The second term in (2.1) represents the interference from $K$ interferers, and the third term represents the additive noise. Now we will group those two terms together as the following:

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \tilde{\boldsymbol{n}}, \tag{2.2}$$

where, $\tilde{\boldsymbol{n}}$ represents colored noise with the following covariance matrix:

$$\boldsymbol{R}_{\tilde{n}} = \boldsymbol{E}[\tilde{\boldsymbol{n}}\tilde{\boldsymbol{n}}^{\boldsymbol{H}}] = \boldsymbol{R}_n + \sum_{j=1}^{K} \boldsymbol{G}_j \boldsymbol{i}_j {\boldsymbol{i}_j}^H {\boldsymbol{G}_j}^H. \tag{2.3}$$

Although there are many studies that use similar representation to (2.2), most of the work in literature assumes that the noise is uncorrelated at the UE side. However, in cellular networks, noise can be correlated. Noise correlation can arise if noise/interference originates from a common source. For example, as described in LTE systems, UEs are exposed to a set of neighboring interferers (inter-cell interference) due to the broadcasting nature of cellular networks, resulting in correlated noise at the UE as shown in Fig. 2.2.

In other words, a portion of the noise received at the UE antennas have originated from the same sources due to the presence of neighboring cells, which implies that the noise at the UE antennas is correlated. It is important to note that although we assume throughout the paper that the neighboring cells are macro cells (BSs), they can be micro cells, Pico cells, Femto cells, or relay

nodes.

Consequently, it is not practical to assume that noise is uncorrelated. Therefore, in practical cellular networks, conventional techniques such as MRT/MRC fail to perform as expected. In this paper, we aim at exploring the effect of noise correlation in wireless cellular networks. Based on that, we obtain a closed-form solution for the precoding and the receiver optimization problem for cellular networks assuming correlated noise at the UEs. We show that JPR outperform the conventional MRT/MRC especially when the noise is highly correlated.

It is important to note that, here we focus on inter-cell interference as the main reason for noise correlation. However, noise can be correlated due to other factors, such as UE close antenna spacing which results in noise correlation [76]. So our design can be significantly beneficial for communication systems suffering from noise correlation whether it is produced solely by interference or by other factors.

## 2.3 SNR Maximization

We consider the effect of the BS precoding and the receiver filtering (At the receiver, the signals from all receive antenna branches are weighted by the receiver /combining vector). Thus, the detected signal at the UE with the JPR design is:

$$\boldsymbol{y} = \boldsymbol{z}^H \boldsymbol{H} \boldsymbol{v} \boldsymbol{x} + \boldsymbol{z}^H \tilde{\boldsymbol{n}}, \tag{2.4}$$

Where, $\boldsymbol{z}$ and $\boldsymbol{v}$ are the receiver $M \times 1$ and precoding vectors $N \times 1$ respectively. The use of the Hermitian transposition of $\boldsymbol{z}$ is necessary since $\boldsymbol{z}$ is a column vector. The objective is to maximize the SNR of the received signal as:

$$\max_{\boldsymbol{z},\boldsymbol{v}} \frac{|\boldsymbol{z}^H \boldsymbol{H} \boldsymbol{v}|^2}{\boldsymbol{z}^H \boldsymbol{R}_{\tilde{n}} \boldsymbol{z}} \sigma_x^2 \text{ subject to } ||\boldsymbol{v}||^2 = P_t. \tag{2.5}$$

Where $\boldsymbol{E}[|\boldsymbol{x}|^2] = \sigma_x^2$ and $P_t$ is the transmit power. By considering the transmitted symbol power $\sigma_x^2$ constant, the optimization in (2.5) can be expressed as:

$$\max_{\boldsymbol{z},\boldsymbol{v}} \frac{|\boldsymbol{z}^H \boldsymbol{H} \boldsymbol{v}|^2}{\boldsymbol{z}^H \boldsymbol{R}_{\tilde{n}} \boldsymbol{z}} \text{ subject to } ||\boldsymbol{v}||^2 = P_t. \tag{2.6}$$

In order to solve this problem, we will initially assume that $\boldsymbol{v}$ is known, and solve for $\boldsymbol{z}$. The solution will be function of $\boldsymbol{v}$, which is still needed to be designed. Then by substitution in the main optimization problem, we can design $\boldsymbol{v}$.

The optimization in (2.6) assuming that $\boldsymbol{v}$ is known can be expressed as:

$$\max_{\boldsymbol{z}} \frac{|\boldsymbol{z}^H \boldsymbol{H} \boldsymbol{v}|^2}{\boldsymbol{z}^H \boldsymbol{R}_{\tilde{n}} \boldsymbol{z}} = \max_{\boldsymbol{z}} \frac{\boldsymbol{z}^H (\boldsymbol{H} \boldsymbol{v} \boldsymbol{v}^H \boldsymbol{H}^H) \boldsymbol{z}}{\boldsymbol{z}^H \boldsymbol{R}_{\tilde{n}} \boldsymbol{z}} \tag{2.7}$$

According to the generalized eigenvalue problem, the solution to (3.5) is:

$$\begin{aligned} \boldsymbol{z} &= \alpha v_{max}[\boldsymbol{R}_{\tilde{n}}^{-1} \boldsymbol{H} \boldsymbol{v} \boldsymbol{v}^H \boldsymbol{H}^H] \\ &= \alpha \boldsymbol{R}_{\tilde{n}}^{-1} \boldsymbol{H} \boldsymbol{v}, \end{aligned} \tag{2.8}$$

Where $\alpha$ adjusts the power of $\boldsymbol{z}$. Note that the scalar $\alpha$ can be ignored since it has no effect on the original objective function in (2.6). Substituting the above $\boldsymbol{z}$ into the objective function in (2.6), the new objective function will be:

$$\max_{\boldsymbol{v}} \boldsymbol{v}^H \boldsymbol{H}^H \boldsymbol{R}_{\tilde{n}}^{-1} \boldsymbol{H} \boldsymbol{v} \text{ subject to } ||\boldsymbol{v}||^2 = P_t. \tag{2.9}$$

The solution to the above problem is the eigenvector associated with the highest eigenvalue with a unit norm of $\boldsymbol{H}^H \boldsymbol{R}_n^{-1} \boldsymbol{H}$. In other words,

$$\boldsymbol{v} = \beta v_{max}[\boldsymbol{H}^H \boldsymbol{R}_{\tilde{n}}^{-1} \boldsymbol{H}], \tag{2.10}$$

Table 2.1: Comparison of different interference management techniques

| Technique | $Pros$ | $Cons$ |
|---|---|---|
| Frequency reuse [2] | Very simple | Waste of bandwidth, i.e. Low spectral efficiency |
| Cooperative BSs [3] | Higher gain compared to frequency reuse | High latency due to cooperation between BSs |
| ML [4, 5] | Optimal Performance | Very high complexity |
| MMSE-SIC [6, 7] | Much lower complexity than ML | Performance degrades with the number of interferers and requires knowledge of interference |
| Iterative IC [8, 9] | Better performance than MMSE-SIC | Requires large number of iterations and requires knowledge of interference |
| JPR | No waste of bandwidth, lower complexity than ML, No BS coordination required, closed-form, No restrictions on the number of interferers | Requires $2^{nd}$ order statistics of interference |

Where $\beta$ is the scalar that adjusts the transmitted power. We compare JPR to the MRT/MRC solution where:

$$\boldsymbol{v}_{\text{MRT}} = v_{max}[\boldsymbol{H}^H \boldsymbol{H}], \tag{2.11}$$
$$\boldsymbol{z}_{\text{MRC}} = \boldsymbol{H}\boldsymbol{v}_{\text{MRT}}.$$

Moreover, Table 2.1 presents a summary comparison between JPR and different interference management techniques.

## 2.4 Rate Maximization for Multi-Carrier Systems

In the previous section, we explained the SNR maximization solution that is applicable for single carrier systems. Here, we will modify the problem formulation and solution targeting multi-carrier systems.

The resulting received signal by the UE under consideration, is given:

$$\boldsymbol{y}(k) = \boldsymbol{H}_i(k)\boldsymbol{v}_i(k)\boldsymbol{x}_i(k) + \sum_{j=1}^{J} \boldsymbol{H}_j(k)\boldsymbol{v}_j(k)\boldsymbol{x}_j(k) + \boldsymbol{n}(k), \tag{2.12}$$

where $\boldsymbol{H}_i(k)$ is $N_r \times N_t$ matrix that represents the channel between the home BS and the UE on hand over the $k^{th}$ subcarrier, $\boldsymbol{H}_j(k)$ is the channel due to the $j$th interferer, $\boldsymbol{x}_i(k)$ is the signal intended to the UE on hand $\boldsymbol{x}_j(k)$ is the signal from the $j$th interferer, $\boldsymbol{y}(k)$ is the $N_r \times 1$ vector that represents the received signal by the UE on hand, $\boldsymbol{v}_i(k)$ is the $N_t \times 1$ vector that represents the precoding vector ,and $\boldsymbol{n}(k)$ is the additive noise with $\boldsymbol{R}_n(k) = \boldsymbol{E}[\boldsymbol{n}(k)\boldsymbol{n}^{\boldsymbol{H}}(k)]$.
The first term in (2.12) represents the intended signal, the second term represents the interference signal from BSs other than the home BS (other-cell interference), and the final term represents the additive noise.

The received signal with the receiver design is:

$$\begin{aligned} \boldsymbol{y}_R(k) &= \boldsymbol{z}_i(k)^H \boldsymbol{y}(k), \\ &= \boldsymbol{z}_i(k)^H \boldsymbol{H}_i(k)\boldsymbol{v}_i(k)\boldsymbol{x}_i(k) + \boldsymbol{z}_i(k)^H \boldsymbol{w}(k) \end{aligned} \tag{2.13}$$

where, $\boldsymbol{z}_i(k)$ is the $N_r \times 1$ vector that represents the receiver vector, and $\boldsymbol{w}(k)$ is the interference plus noise term.

The objective here is to maximize the rate at the user on hand:

$$R_i(k) = \log_2(1 + \frac{|\boldsymbol{z}_i(k)^H \boldsymbol{H}_i(k)\boldsymbol{v}_i(k)|^2}{\boldsymbol{z}_i(k)^H \boldsymbol{R}_{\boldsymbol{w}} \boldsymbol{z}_i(k)} \sigma^2_{x_i(k)}), \tag{2.14}$$

which results in maximizing the SINR of the received signal:

$$\max_{\boldsymbol{z}_i(k), \boldsymbol{v}_i(k)} \frac{|\boldsymbol{z}_i(k)^H \boldsymbol{H}_i(k)\boldsymbol{v}_i(k)|^2}{\boldsymbol{z}_i(k)^H \boldsymbol{R}_{\boldsymbol{w}} \boldsymbol{z}_i(k)} \sigma^2_{x_i(k)} \text{ subject to } ||\boldsymbol{v}_i(k)||^2 = 1. \tag{2.15}$$

where $\boldsymbol{E}[|\boldsymbol{x}_i(k)|^2] = \sigma^2_{x_i(k)}$. By considering the transmitted symbol power $\sigma^2_{x_i(k)}$ is unity, the optimization in (2.15) can be expressed as:

$$\max_{\boldsymbol{z}_i(k), \boldsymbol{v}_i(k)} \frac{|\boldsymbol{z}_i(k)^H \boldsymbol{H}_i(k)\boldsymbol{v}_i(k)|^2}{\boldsymbol{z}_i(k)^H \boldsymbol{R}_{\boldsymbol{w}} \boldsymbol{z}_i(k)} \text{ subject to } ||\boldsymbol{v}_i(k)||^2 = 1. \tag{2.16}$$

Assuming that $\boldsymbol{v}_i(k)$ is known, the optimization in (2.16) can be expressed as:

$$\max_{\boldsymbol{z}_i(k)} \frac{|\boldsymbol{z}_i(k)^H \boldsymbol{H}_i(k)\boldsymbol{v}_i(k)|^2}{\boldsymbol{z}_i(k)^H \boldsymbol{R}_w \boldsymbol{z}_i(k)} = \max_{\boldsymbol{z}_i(k)} \frac{\boldsymbol{z}_i(k)^H (\boldsymbol{H}_i(k)\boldsymbol{v}_i(k)\boldsymbol{v}_i(k)^H \boldsymbol{H}_i(k)^H)\boldsymbol{z}_i(k)}{\boldsymbol{z}_i(k)^H \boldsymbol{R}_w \boldsymbol{z}_i(k)} \tag{2.17}$$

According to the generalized eigenvalue problem, the solution to (2.17) is [77, 78, 80, 85]:

$$\begin{aligned} \boldsymbol{z}_i(k) &= \alpha v_{max}[\boldsymbol{R}_w^{-1} \boldsymbol{H}_i(k)\boldsymbol{v}_i(k)\boldsymbol{v}_i(k)^H \boldsymbol{H}_i(k)^H] \\ &= \alpha \boldsymbol{R}_w^{-1} \boldsymbol{H}_i(k)\boldsymbol{v}_i(k), \end{aligned} \tag{2.18}$$

where $\alpha$ adjusts the power of $\boldsymbol{z}_i(k)$, and $v_{max}$ is the principal eigenvector (i.e. the eigenvector associated with the highest eigenvalue). Note that the scalar $\alpha$ can be ignored since it has no effect on the original objective function in (2.6). Substituting the above $\boldsymbol{z}_i(k)$ into the objective function in (2.6), the new objective function will be:

$$\max_{\boldsymbol{v}_i(k)} \boldsymbol{v}_i(k)^H \boldsymbol{H}_i(k)^H \boldsymbol{R}_w^{-1} \boldsymbol{H}_i(k)\boldsymbol{v}_i(k) \text{ subject to } ||\boldsymbol{v}_i(k)||^2 = 1. \tag{2.19}$$

The solution to the above problem is the eigenvector associated with the highest eigenvalue of $\boldsymbol{H}_i(k)^H \boldsymbol{R}_w^{-1} \boldsymbol{H}_i(k)$. In other words,

$$\boldsymbol{v}_i(k) = \beta v_{max}[\boldsymbol{H}_i(k)^H \boldsymbol{R}_w^{-1} \boldsymbol{H}_i(k)], \tag{2.20}$$

where $\beta$ is a scalar that is used in the normalization step such that $||\boldsymbol{v}_i(k)||^2 = 1$.

Thus, the precoding design, $\boldsymbol{v}_i(k)$ is the preferred precoding vector requested by the $i^{th}$ UE with the aim of increasing its rate.

## 2.5 LTE Specifications and System Assumptions

Before addressing the challenges in LTE, we review some important elements of LTE (Release 10 and beyond). LTE base station is known as eNodeB, and UEs refer to mobile terminals or user end-devices. In LTE, sub-carriers are grouped in non-overlapping subsets, called Resource Blocks (RBs), and is the smallest allocatable resource in the frequency-time domain. Single subcarrier and singe symbol is called resource element (RE). The unit in time is a 1 msec unit consisting of 14 OFDM symbols. The resource grid shown at Fig 2.2 contains data along with other signals. The Physical Downlink Control Channel (PDCCH) and the cell specific RS (CRS) are used to demodulate the control signaling and perform mobility measurements, the Channel State Information Reference Signal (CSI-RS) are used for raw channel estimation, it is a reference signal used by the UEs to estimate the channel. DeModulation Reference Signals (DM-RS) are used for demodulation purposes. Control signaling, the CRS, and CSI-RS are transmitted without performing precoding, however, DM-RS can be precoded [81]. In some transmission modes DM-RS REs are replaced by data REs. On the other hand, when DM-RS REs are used, the UE is expected to use them to derive the channel estimate for demodulating the data. A typical usage of the DM-RS signal is to enable beamforming of the data transmissions to specific UEs. Such a beam will experience a different channel response between the eNodeB and UE, thus requiring the use of DM-RSs to enable the UE to demodulate the beamformed data coherently [82].

We consider an LTE network formed by UEs with $N_r$ antennas, served in the DL through a multi-path channel by BSs with $N_t$ antennas. We assume that each BS has $j$ adjacent strong neighbors. The system parameters are defined according to the LTE specifications [84]. We further assume that transmission from different cells are not synchronized. In other words, the sub-frames of different cells are not aligned with each other.

## 2.6 Modified Problem for LTE OFDMA Systems

The received signal of the $k^{th}$ subcarrier and the $t^{th}$ OFDM symbol, is given:

$$\begin{aligned}\boldsymbol{y}(k,t) &= \boldsymbol{H}_i(k,t)\boldsymbol{v}_i(k,t)\boldsymbol{x}_i(k,t) \\ &+ \sum_{j=1,j\neq i}^{J} \boldsymbol{H}_j(k,t)\boldsymbol{v}_j(k,t)\boldsymbol{x}_l(k,t) + \boldsymbol{n}(k,t),\end{aligned} \tag{2.21}$$

where $\boldsymbol{H}_i(k,t)$ is $N_r \times N_t$ matrix that represents the channel between the $i^{th}$ eNodeB and the UE under consideration, $\boldsymbol{y}(k,t)$ is the $N_r \times 1$ vector that represents the received signal by the UE on hand, and $\boldsymbol{v}_i(k,t)$ is the $N_t \times N_{stream}$ precoding matrix. Note that, while MU-MIMO in LTE supports only rank-1 transmission [82, Chapter 11], i.e., one stream to each UE, we refer to $\boldsymbol{v}_i(k,t)$ as precoding vector for the purposes of this discussion, although in the LTE specifications the term precoding matrix is used for both SU-MIMO (with rank $\geq 1$) and MU-MIMO (with rank =1). Furthermore, $\boldsymbol{x}_i(k,t)$ is the information signal vector intended to the UE on hand, and $\boldsymbol{n}(k,t)$ is $N_r \times 1$ vector that represents the additive noise with $\boldsymbol{R}_n(k,t) = \boldsymbol{E}[\boldsymbol{n}(k,t)\boldsymbol{n}^{\boldsymbol{H}}(k,t)]$.

The first term in (1) represents the intended signal, the second term represents the interference signal from eNodeBs other than the home eNodeB (other-cell interference), and the final term represents the additive noise. The post processed received signal is:

$$\boldsymbol{y}_R(k,t) = \boldsymbol{z}_i(k,t)^H \boldsymbol{y}(k,t), \tag{2.22}$$

where, $\boldsymbol{z}_i(k,t)$ is the $N_r \times 1$ receiver vector. Our main objective is to design the precoding vector $\boldsymbol{v}_i(k,t)$, and the combining vector $\boldsymbol{z}_i(k,t)$.

## 2.7 Interference Estimation

### 2.7.1 Interference estimation using REs carrying data

In LTE, the covariance matrix $\boldsymbol{R}_{yy}$ is estimated by performing averaging over the received signal at the data signal REs: $\boldsymbol{R}_{yy} = \boldsymbol{E}[\boldsymbol{y}(k,t)\boldsymbol{y}(k,t)^{\boldsymbol{H}}]$. In this case, the estimation error can be neglected when the cross-covariance between the signals of the serving cell and the interfering cell: $\boldsymbol{E}[\boldsymbol{x}_i(k,t)\boldsymbol{x}_l(k,t)^{\boldsymbol{H}}]$, and $\boldsymbol{E}[\boldsymbol{x}_l(k,t)\boldsymbol{x}_i(k,t)^{\boldsymbol{H}}]$ are negligible. However, in a practical situation, the cross-covariance between the signals of the serving cell and the interfering cell is not small. Therefore, the residual cross-covariance incurs performance degradation [83].

### 2.7.2 Interference estimation using REs carrying DM-RS

If the interference plus noise covariance matrix is computed using REs that are not carrying data signals, the cross-covariance between the signals of the serving and interfering cell can be eliminated. In this subsection, we discuss the use of REs carrying DM-RS as a method of estimating the interference. In this case, the UEs can find the interference plus noise covariance matrix $\boldsymbol{R}_{\boldsymbol{w}}$ as follows:

$$\boldsymbol{R}_w = \frac{1}{|N_{DMRS}|} \sum_{\hat{k},\hat{t} \in N_{DMRS}} \boldsymbol{y}_D(\hat{k},\hat{t})\boldsymbol{y}_D(\hat{k},\hat{t})^{\boldsymbol{H}}, \tag{2.23}$$

where, $\boldsymbol{y}_D(\hat{k},\hat{t})$ is the $N_r \times 1$ vector that represents the interference and noise vector on the RE carrying DM-RS. $\boldsymbol{y}_D(\hat{k},\hat{t})$ can be found as follows:

$$\boldsymbol{y}_D(\hat{k},\hat{t}) = \boldsymbol{y}(\hat{k},\hat{t}) - \boldsymbol{G}_i(\hat{k},\hat{t})\boldsymbol{d}(\hat{k},\hat{t}), \quad \hat{k},\hat{t} \in N_{DMRS}, \tag{2.24}$$

where,$\boldsymbol{G}_i(\hat{k},\hat{t})$ is the composite channel at the REs carrying DM-RS signals, $\boldsymbol{d}(\hat{k},\hat{t})$ is the DM-RS known sequence of the serving cell and $N_{DMRS}$ is the set of REs carrying DM-RS.

## 2.8 Remarks

### 2.8.1 Notes on Interference Estimation Process

Although, the estimation of $\boldsymbol{R}_w$ here is suboptimal, we aim at utilizing this information in optimizing the precoder and receiver.

#### Applicability

This interference estimation technique is broad, and can be used in any system where UE-specific REs that are not carrying data are available.

#### Fast-Fading and Frequency Selectivity

The interference covariance matrix is interpreted in time-frequency average sense so that the effect of fast-fading and frequency selectivity is averaged out.

### 2.8.2 Notes on Channel Modeling

We assume different channel models, one of them is a quasi-static flat Rayleigh fading model. We assume perfect channel knowledge at the UE side throughout this thesis. This knowledge can be practically available through various types of channel estimation techniques.

## 2.9 Simulation Results

In this section, the performance of JPR is compared to the conventional method via simulations. First, we assume a quasi-static flat Rayleigh fading model for the channel. Without loss of generality, in the initial set of results, the covariance matrix of the noise is constructed as follows:

$$[\boldsymbol{R}_{\tilde{n}}]_{i,j} = \rho^{|i-j|} \tag{2.25}$$

Where, $\rho < 1$ is the correlation factor. As shown in the previous set of equations, we designed the JPR algorithm to take noise correlation into consideration. The JPR algorithm is also practical since the only overhead relative to the conventional MRT/MRC is that the covariance matrix of the noise need to be available. This assumption is reasonable in both time division duplex (TDD) and frequency division duplex (FDD) LTE. In TDD-LTE, BS can estimate the interference of a certain UE along with its channel using the uplink by means of channel reciprocity. However, in FDD-LTE, UEs can take interference into account in the channel state information (CSI) feedback to the home BS. Thus, UEs can feedback the interference information. In other words, each UE can design its preferable precoding vector and feed it back to the BS. Thus, the proposed design will have a low complexity compared to other interference management techniques and can achieve high performance in terms of both throughput and probability of error. Consequently, JPR can be considered practical, efficient, and has low complexity. The system parameters are defined according to the LTE specifications reported in Table 2.2 [84].

Table 2.2: System Parameters

| Parameter | Value |
|---|---|
| Carrier Frequency | 2 GHz |
| Transmission Bandwidth | 10 MHz |
| Number of subcarriers | 600 |
| Subcarrier spacing | 15 kHz |
| FFT size | 1024 |
| Modulation | QPSK, 16 QAM, and 64 QAM |
| Noise Figure | 5 dB |
| Traffic Model | Full buffer |
| Scheduling | Proportional fair |

Fig. 2.3 shows the output SNR versus the input SNR of the proposed solution for a $3 \times 3$ MIMO system using the QPSK modulation technique. It is clear that the system with 0.5 correlation factor is around 5-dB better than the uncorrelated system. Fig. 2.4 is the same as the previous figure, but for the conventional MRT/MRC solution. It is clear that correlated noise reduces the performance by around 1-dB relative to the uncorrelated system. Fig. 2.5 shows a comparison between the proposed and the conventional solution for different correlation factors. As can be shown, our proposed design results in an increase of around 6-dB for noise correlation of factor 0.5. Figure 2.6 presents the BER performance of the proposed technique versus the conventional approach. It shows a comparison between the proposed and the conventional solution in terms of BER in logarithmic scale for a correlation factor of 0.5, 0.25 and zero (uncorrelated). As shown, our proposed design results in BER reduction as expected from previous figures. It can be noticed that as the correlation factor decreases, the gap between both solutions shrinks until it is eliminated as expected in the case of uncorrelated noise. Which means that our solution reduces to the conventional MRT/MRC in case of uncorrelated noise.

Now, that we have a good understanding and analysis in modeling interference as correlated noise and how to utilize it in both the precoding and receiver, we will do another set of simulations. The previous set of simulation was focused on any system with interferences. Now, we will focus on LTE cellular networks as shown in fig. 2.7.

We assume that each BS has $j$ adjacent strong neighbors. Fig. 2.8 shows a network with 3 strong interference neighboring BSs as an example of the system model under consideration.

We examine the performance degradation due to the MMSE receiver. The motivation of this work is to achieve high data rates and improve the spectral efficiency. In order to focus on the impact of interference errors only, we assume non-ideal interference estimation (by using a reasonable number of DM-RS REs), while we assume ideal TDD reciprocity and noiseless channel estimation. We use the achievable sum-rates as a performance metric.

We provide a comparison between the MMSE with different precoding schemes and our scheme. The simulation setup follows [94]. We perform both a link based simulation, and a system based simulation where, the system consists of seven BSs as shown in Fig. 2.8. Each BS is assumed to have four antennas, while each UE is assumed to have two antennas. Each cell has five UEs. The UEs are independently and randomly located with uniform probability over each cell coverage area. The UEs achievable rates are calculated by averaging over several realizations of the UEs locations. The cell radius is assumed to be 500m. The minimum distance between the BS and the UE is assumed to be 35m.

Fig. 2.9 shows the achievable rates obtained by Monte Carlo simulation where $\boldsymbol{H}_i$ and $\boldsymbol{H}_j$ are assumed to have i.i.d elements $\sim \mathcal{CN}(0,1)$ (normalized independent Rayleigh fading channel) with perfect CSI. The y-axis presents the average throughput from 10,000 random channel realizations. In the throughput calculations, we account for the signaling overhead in terms of the average number of REs that are not used for data. In this figure, we focus on one UE per cell (via link-based simulation). Subsequently, we consider the entire system performance by performing system-based simulation.

Fig. 2.10 is the same as previous figure, but with typical urban (TU) channel. This channel model is the geometry-based stochastic model, which has been used for the IMT-Advanced Self Evaluation

Report. As can be noticed, JPR provides higher gains in channels with high frequency selectivity such as TU. In order to explain these results, we need to describe the interference estimation in more details. Each UE estimates the interference on the REs that are reserved for DM-RS and then these values are interpolated to other REs that contain data symbols. Thus, as the number of REs reserved for DM-RS increases, the interference estimation accuracy increases. On the other hand, increasing the DM-RS in the OFDM frame reduces the resources which can be dedicated to data which implies a capacity loss.

In Rayleigh fading channels, the number of DM-RS seems to be sufficient for the MMSE algorithm to provide good results. However, if the channel suffers from high selectivity as the TU case, the number of REs reserved for DM-RS becomes insufficient for accurate interference estimation. Thus, the MMSE performance is degraded, while the JPR performance is not in the presence of inaccurate interference estimation. Motivated by the LTE standard, we assume that 12 REs are suitable to be used for DM-RS. Based on the simulation results, the proposed scheme has more relaxed requirements for interference knowledge accuracy than MMSE receiver.

In Fig. 2.11, as a comparison, we show the proposed JPR with both the advanced and the simple MMSE with both maximum ratio transmission (MRT) precoding and codebook precoding.

In order to accurately evaluate the system performance, it is important to consider both the cell-edge throughput (5%) and the mean throughput. Fig. 2.12 presents the throughput cumulative distribution function (CDF) for the 6-tap TU channel. As shown, the JPR algorithm enhances the cell-edge UE throughput from 0.84 Mbps to 0.97 Mbps and the mean throughput from 1.31 Mbps to 1.48 Mbps as compared to the advanced MMSE with MRT precoding.

## 2.10 Conclusions

This chapter presents a jointly optimized solution for designing the precoder and the receiver for LTE systems as a key enabling technique to make efficient use of the available interference knowledge. The proposed algorithm can provide high performance in terms of both throughput and probability of error. We first develop a model for the interference, and then we consider it in our joint design. A sufficient condition for the proposed algorithm to work, is that the UE feedback the interference information along with the CSI to the home BS. Since the sufficient condition is practical, our proposed algorithm is efficient and has low complexity. As shown in the simulation section, it can lead to significant increase in throughput, especially for UEs suffering from highly correlated noise. We further compare our algorithm with the conventional MRT/MRC technique that is optimal for uncorrelated noise systems.

Moreover, we have proposed an interference estimation approach aiming at optimizing the performance of LTE network. The proposed scheme can achieve system performance comparable to the interference free transmission but with extra processing at the BS and UE sides. Furthermore, the proposed design is evaluated in terms of throughput for practical channel models. The proposed algorithm is efficient and performs better than currently adopted LTE designs. As shown in the simulation section, it can lead to higher throughput, especially for practical channel models that are highly frequency selective[85, 86].

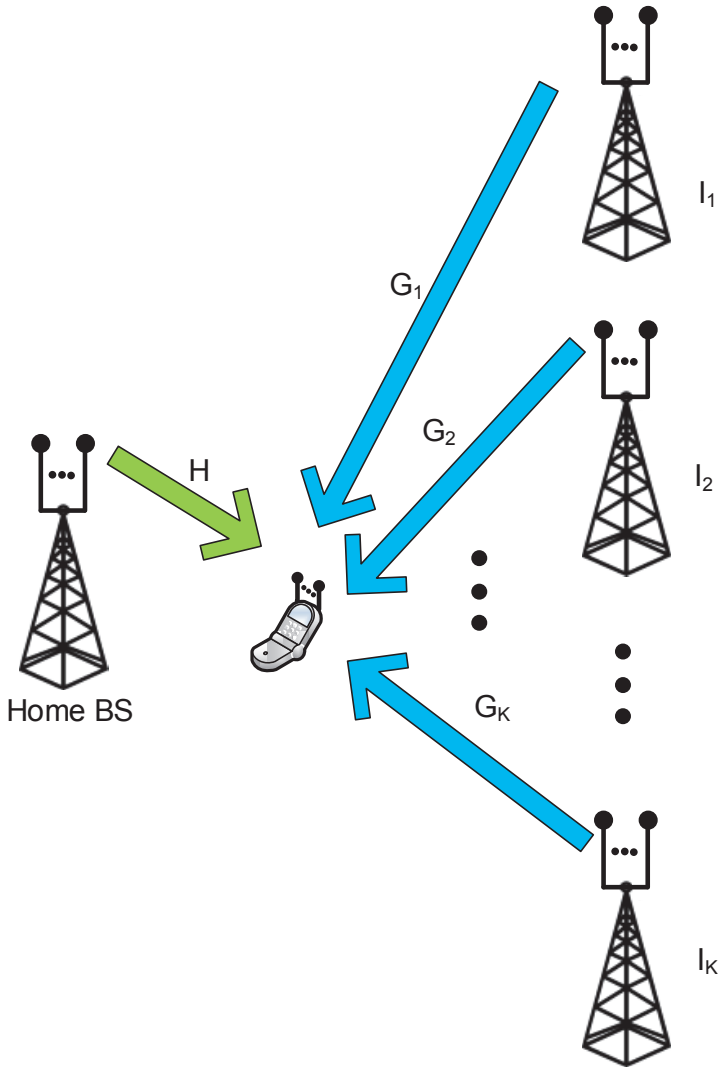



Figure 2.1: System Model

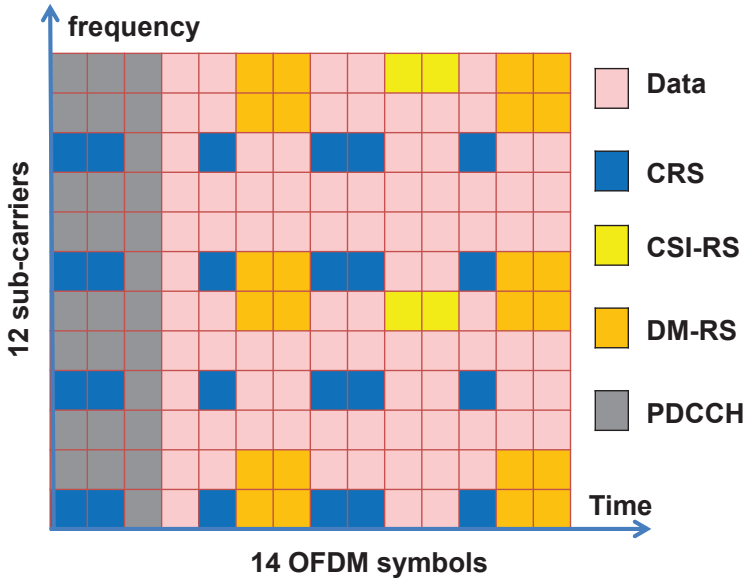



Figure 2.2: LTE Release 10 Resource Grid

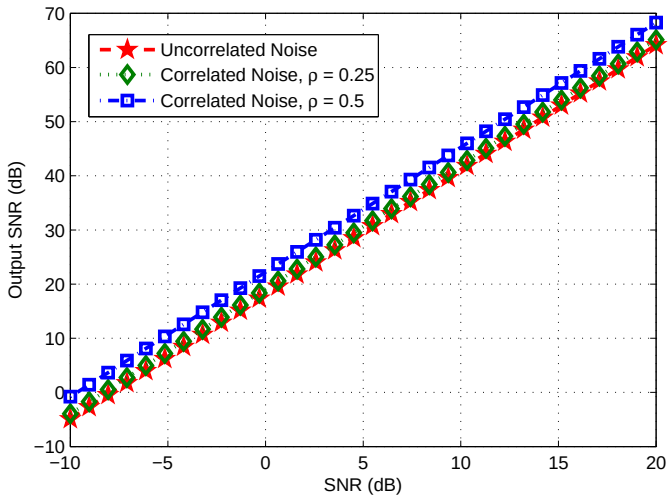



Figure 2.3: Output SNR for JPR design for a 3x3 system with correlated noise

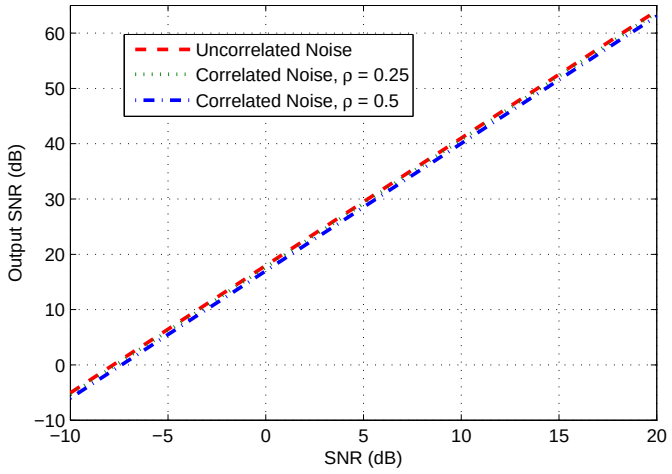



Figure 2.4: Output SNR for Conventional solution for 3x3 system with correlated noise

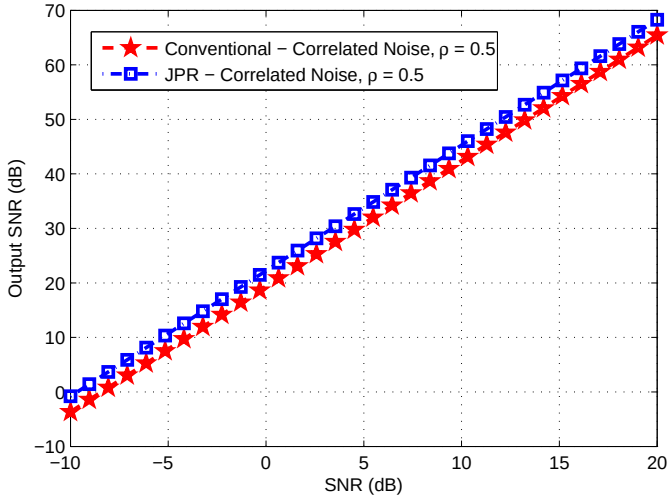



Figure 2.5: Output SNR Comparison for 3x3 system with correlated noise

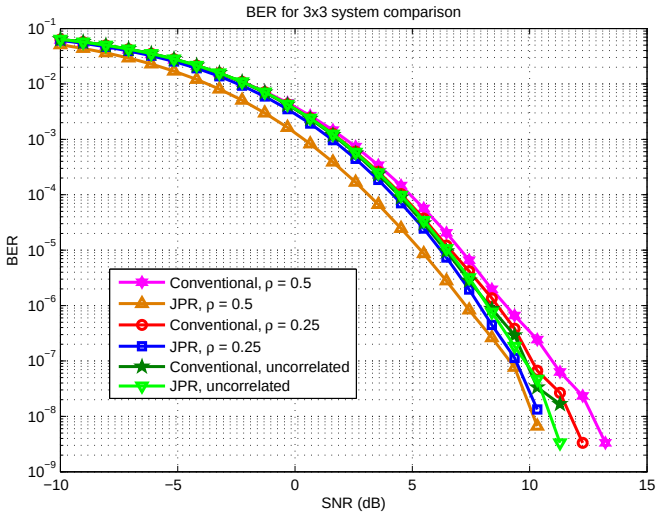



Figure 2.6: Average BER Comparison for 3x3 system with different correlation factors

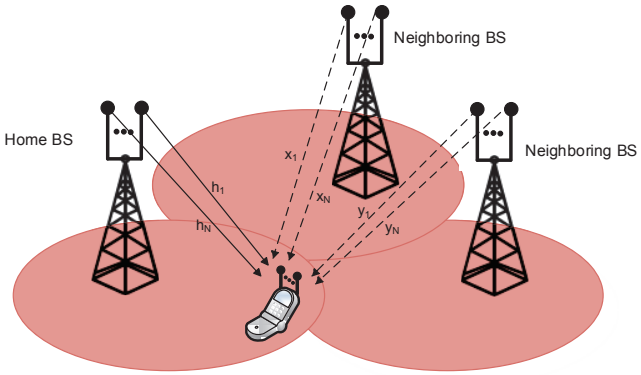



Figure 2.7: Cellular System Example

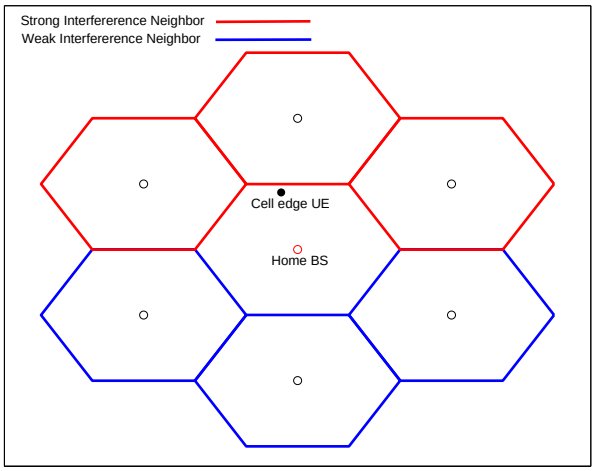



Figure 2.8: Network Example

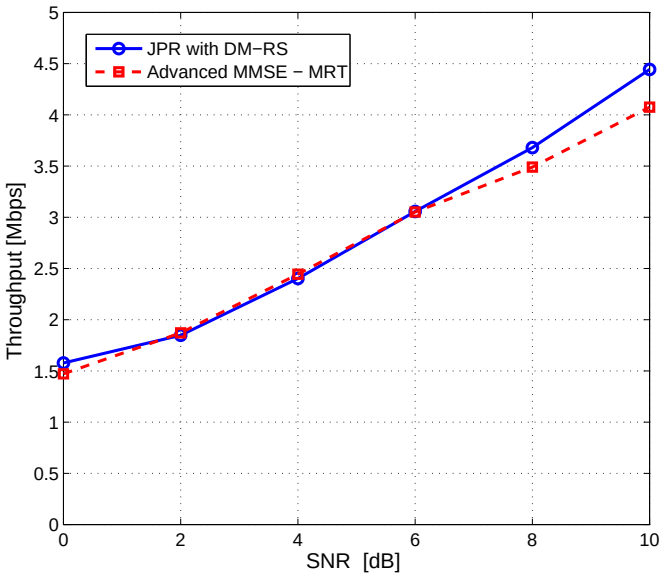



Figure 2.9: Throughput comparison for Rayleigh fading channel

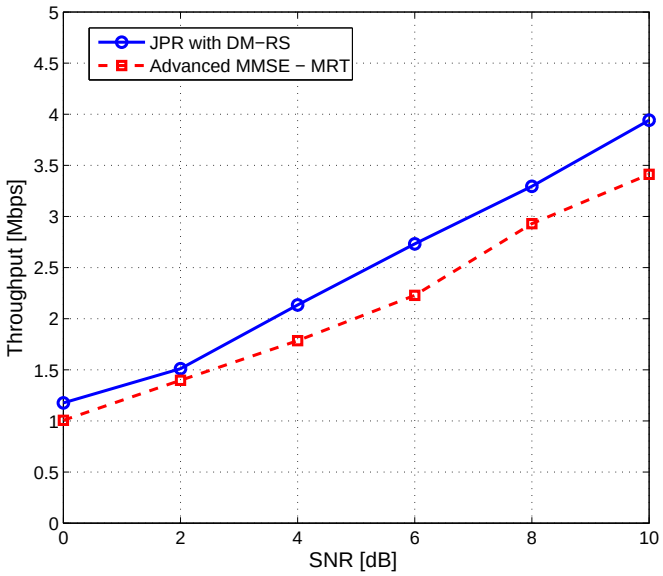



Figure 2.10: Throughput comparison for TU channel

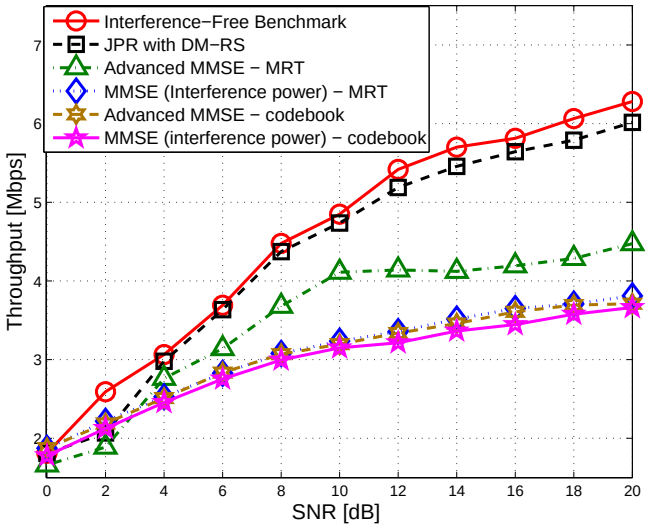



Figure 2.11: Throughput comparison for TU channel using system-level simulation

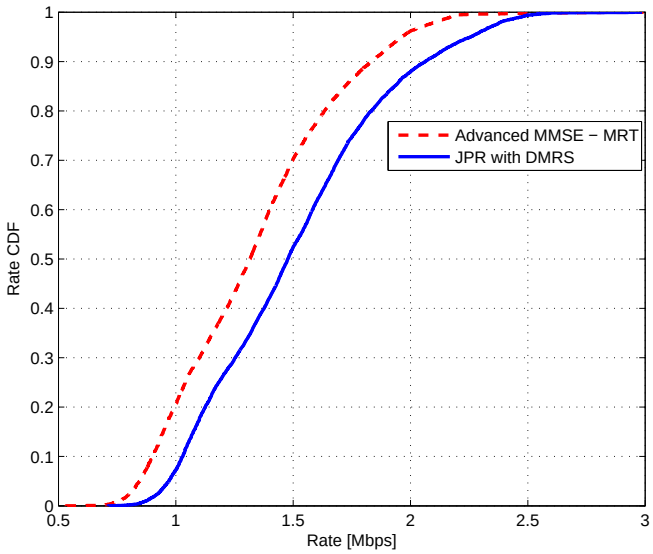



Figure 2.12: Throughput CDF for TU channel using system-level simulation

# Chapter 3

# Practical Considerations in Multi-user LTE Networks

## 3.1 Introduction

MU-MIMO is a promising wireless technique where the same time-frequency channel resources are allowed to be used by multiple UEs simultaneously through spatial precoding. The performance of LTE systems critically depends on how the interference either across different cells or due to MU-MIMO is managed. In this chapter, we focus primarily on utilizing the interference information and considering the MU-MIMO non-ideality. We also study the multi-cell effect. The performance of the proposed approach is benchmarked against the precoding and receiver designs that are currently considered for LTE systems.

### 3.1.1 Contributions

The main results of this chapter are:

1. Based on the interference estimation framework presented in Chapter 2, we derive suboptimal precoder and receiver aiming at enhancing the system throughput for MU-MIMO operation. The precoder and receiver solution is simple (non-iterative) yet enough to account for LTEs interference patterns.

2. We tackle the effect of non-ideal precoding for MU-MIMO in LTE. The solution aims at updating the interference covariance matrix to account for MU-MIMO interference.

The remainder of the chapter is organized as follows: Section 3.2 describes the problem formulation. In Section 3.3, we discuss the effect of ZF beamforming with limited channel knowledge. Simulation results are provided in Section 3.4 and we conclude the chapter in Section 3.5.

### 3.1.2 Notation

We use bold lower case for vectors, such as $\boldsymbol{a}$, while bold capital letters are used for matrices such as $\boldsymbol{A}$. Further $||\boldsymbol{A}||$ stands for the norm of the matrix $\boldsymbol{A}$. Further $(\cdot)^H$ stands for Hermitian transposition. $[\boldsymbol{A}]_{i,j}$ denotes the element in row $i$ and column $j$ of matrix $\boldsymbol{A}$. The cardinality of the set $A$ is denoted by $|A|$. Also $\boldsymbol{E}$ stands for expectation operator.

## 3.2 Problem Formulation

We consider a network formed by UEs with $N_r$ antennas, served by eNodeBs with $N_t$ antennas, using MU-MIMO. We assume that each eNodeB has $J$ adjacent neighbors. The received signal of

the $k^{th}$ subcarrier and the $t^{th}$ OFDM symbol, is given:

$$\begin{aligned}\boldsymbol{y}(k,t) = &\boldsymbol{H}_i(k,t)\boldsymbol{v}_i(k,t)\boldsymbol{x}_i(k,t) \\ &+ \sum_{g=1,g\neq i}^{G} \boldsymbol{H}_i(k,t)\boldsymbol{v}_g(k,t)\boldsymbol{x}_g(k,t) \\ &+ \sum_{j=1,j\neq i}^{J} \sum_{\forall l} \boldsymbol{H}_j(k,t)\boldsymbol{v}_l(k,t)\boldsymbol{x}_l(k,t) + \boldsymbol{n}(k,t),\end{aligned} \tag{3.1}$$

where $\boldsymbol{H}_i(k,t)$ is $N_r \times N_t$ matrix that represents the channel between the $i^{th}$ eNodeB and the UE on hand, $\boldsymbol{y}(k,t)$ is the $N_r \times 1$ vector that represents the received signal by the UE on hand, and $\boldsymbol{v}_i(k,t)$ is the $N_t \times 1$ precoding vector. Furthermore, $\boldsymbol{x}_i(k,t)$ is the information signal vector intended to the UE on hand, $\boldsymbol{n}(k,t)$ is $N_r \times 1$ vector that represents the noise, and $G$ is the number of co-scheduled UEs. The first term in (3.1) represents the intended signal, the second term represents the multiuser interference, the third term represents the neighbouring interference, and the final term represents the noise. The number of co-scheduled UEs is $G \leq N_t$. The post processed received signal is: $\boldsymbol{y}_R(k,t) = \boldsymbol{z}_i^H(k,t)\boldsymbol{y}(k,t)$, where, $\boldsymbol{z}_i(k,t)$ is the $N_r \times 1$ combining vector. Our main objective is to design the precoding vectors $\boldsymbol{v}_i(k,t)$, and the combining vectors $\boldsymbol{z}_i(k,t)$, where, $i = 1 : G$.

Before eNodeBs perform scheduling, each UE will not have knowledge of other UEs that might be scheduled on the same frequency-time resources. Then, the precoding of each UE is designed with the objective of maximizing its local SINR $\boldsymbol{v}_i^l(k,t)$: The superscript $l$ denotes locally optimized precoding, which will be updated later to accommodate for the MU-MIMO effect. Similarly, $\boldsymbol{z}_i^l(k,t)$ is the locally optimized receiver that will be updated later. Therefore, in the initial analysis, the second term in (3.1) will be removed and the other-cell interference plus noise terms will be denoted as $\boldsymbol{w}(k,t)$.

### 3.2.1 Maximization of Local SINR

The received signal for the single user scenario can be expressed as:

$$\boldsymbol{y}(k,t) = \boldsymbol{z}_i^{l^H}(k,t)\left[\boldsymbol{H}_i(k,t)\boldsymbol{v}_i^l(k,t)\boldsymbol{x}_i(k,t) + \boldsymbol{w}(k,t)\right] \tag{3.2}$$

The objective here is to maximize the SINR:

$$\begin{aligned}\max_{\boldsymbol{z}_i^l(k,t),\boldsymbol{v}_i^l(k,t)} & \frac{|\boldsymbol{z}_i^{l^H}(k,t)\boldsymbol{H}_i(k,t)\boldsymbol{v}_i^l(k,t)|^2}{\boldsymbol{z}_i^{l^H}(k,t)\boldsymbol{R_w}\boldsymbol{z}_i^l(k,t)} \\ & \text{subject to } ||\boldsymbol{v}_i^l(k,t)||^2 = 1.\end{aligned} \tag{3.3}$$

The problem in (3.3) is coupled, however, here we present a practical suboptimal non-iterative solution. For a specific $\boldsymbol{v}_i^l(k)$, the optimization in (3.3) can be expressed as:

$$\begin{aligned}&\max_{\boldsymbol{z}_i^l(k,t)} \frac{|\boldsymbol{z}_i^{l^H}(k,t)\boldsymbol{H}_i(k,t)\boldsymbol{v}_i^l(k,t)|^2}{\boldsymbol{z}_i^{l^H}(k,t)\boldsymbol{R}_w\boldsymbol{z}_i^l(k,t)} = \\ &\max_{\boldsymbol{z}_i^l(k,t)} \frac{\boldsymbol{z}_i^{l^H}(k,t)(\boldsymbol{H}_i(k,t)\boldsymbol{v}_i^l(k,t)\boldsymbol{v}_i^{l^H}(k,t)\boldsymbol{H}_i{}^H(k,t))\boldsymbol{z}_i^l(k,t)}{\boldsymbol{z}_i^{l^H}(k,t)\boldsymbol{R}_w\boldsymbol{z}_i^l(k,t)}\end{aligned} \tag{3.4}$$

According to the generalized eigenvalue problem [78], the solution to (3.4) is: $\boldsymbol{z}_i^l(k,t) = \alpha\boldsymbol{R}_w^{-1}\boldsymbol{H}_i(k,t)\boldsymbol{v}_i^l(k,t)$, where $\alpha$ adjusts the power of $\boldsymbol{z}_i^l(k,t)$. Substituting the above $\boldsymbol{z}_i^l(k,t)$ into the objective function in (3.3), the new objective function will be:

$$\begin{aligned}\max_{\boldsymbol{v}_i^l(k,t)} & \boldsymbol{v}_i^{l^H}(k,t)\boldsymbol{H}_i^H(k,t)\boldsymbol{R}_w^{-1}\boldsymbol{H}_i(k,t)\boldsymbol{v}_i^l(k,t) \\ & \text{subject to } ||\boldsymbol{v}_i^l(k,t)||^2 = 1.\end{aligned} \tag{3.5}$$

The solution to the above problem is [78]:

$$\boldsymbol{v}_i^l(k,t) = \beta v_{max}[\boldsymbol{H}_i^H(k,t)\boldsymbol{R}_w^{-1}\boldsymbol{H}_i(k,t)], \tag{3.6}$$

where $\beta$ is a scalar that is used in the normalization step such that $||\boldsymbol{v}_i^l(k,t)||^2 = 1$, , and $v_{max}$ is the principal eigenvector. Thus, the precoding design, $\boldsymbol{v}_i^l(k,t)$ is the preferred precoding vector requested by the $i^{th}$ UE. Thus the solution of the local precoding and combining vectors is as follows:

$$\begin{aligned} \boldsymbol{v}_i^l(k,t) &= \beta v_{max}[\boldsymbol{H}_i^H(k,t)\boldsymbol{R}_w^{-1}\boldsymbol{H}_i(k,t)], \\ \boldsymbol{z}_i^l(k,t) &= \alpha \boldsymbol{R}_w^{-1}\boldsymbol{H}_i(k,t)\boldsymbol{v}_i^l(k,t), \end{aligned} \tag{3.7}$$

We will use the receiver design $\boldsymbol{z}_i^l(k,t)$ that is given in (3.7) and will compare its performance against other receivers. It is important to note that eNBs do not need X2 coordination since the precoder design only needs $\boldsymbol{R}_w$.

### 3.2.2 Maximization of Overall SINR

In the previous section, MU-MIMO effect was not taken into account. However, the eNodeB aims at maximizing the overall sum rate. Thus, the eNodeB will construct $N_t$ x $G$ MU-MIMO precoding matrix with the aim of spatially separating the concurrent transmissions. Zero-forcing (ZF) is considered as an efficient beamforming design for communication systems. In ZF, the weights are selected such that the co-channel interference is canceled. On the other hand, Maximum Ration Transmission (MRT) beamforming maximizes the SNR at each receiver and requires only the knowledge of the direct links. It is worth noting that the MRT does not take into account the simultaneous transmissions and therefore it results in a strong cross-interference. Since, this cross-interference is a bottleneck for multiuser LTE systems, ZF precoding is being studied for current network implementations. We will drop the subcarrier notation $k$ and the OFDM symbol $t$ notation

in the equation below for simplicity:$\boldsymbol{V}^l = [\boldsymbol{v}_1^l ... \boldsymbol{v}_i^l ... \boldsymbol{v}_G^l]$.

$$\begin{aligned} \boldsymbol{V}_{ZF} &= \boldsymbol{V}^l ({\boldsymbol{V}^l}^H \boldsymbol{V}^l)^{-1} \Gamma \\ &= [\boldsymbol{v}_1 ... \boldsymbol{v}_i ... \boldsymbol{v}_G] \end{aligned} \tag{3.8}$$

where, $\Gamma$ is a diagonal matrix that ensure that the columns of $V_{ZF}$ have unit norm. It is important to note, $\boldsymbol{z}_i(k,t)$ can be computed using (3.7) if $\boldsymbol{v}_i(k,t)$ is being used instead of $\boldsymbol{v}_i^l(k,t)$. In other words, $\boldsymbol{z}_i(k,t)$ is a function of $\boldsymbol{v}_i(k,t)$. Initially, the local precoding is designed assuming SU transmission. However, once MU-MIMO precoding matrix is computed, UEs will use a different receiver, not the same as the initial one. A flowchart to the proposed algorithm is shown in Figure 3.1. Typically, the MMSE receiver is a widely used due to its simplicity [96] $\boldsymbol{v}(k,t) = \left\{\boldsymbol{G}_i(k,t)\boldsymbol{G}_i^H(k,t) + \Omega + \sigma_N^2 I\right\}^{-1}\boldsymbol{G}_i(k,t)$, where $\boldsymbol{G}_i(k,t) = \boldsymbol{H}_i(k,t)\boldsymbol{v}_i(k,t)$ is the composite channel, $\Omega$ is $N_r \times N_r$ diagonal matrix with interference powers on the diagonal, and $\sigma_N^2$ is the noise power. Another receiver is the IRC receiver. In IRC, the covariance matrix including interference is obtained by statistical averaging of the received signal [95] $\boldsymbol{v}(k,t) = E[y(k,t)y^H(k,t)]^{-1}\boldsymbol{G}_i(k,t)$.

## 3.3 Effect of ZF Beamforming With Limited Channel Knowledge

If channel knowledge is perfect, the layers will be well-separated using ZF. Thus, the multiuser interference vanishes, and the only interference that the UEs face is the other-cell interference which was taken into account in $\boldsymbol{R}_w$. However, in reality this is not the case, and $\boldsymbol{R}_w$ will be changed to $\tilde{\boldsymbol{R}_w}$ to include the effect of multiuser interference. This is due to the fact that channel knowledge is not perfect due to limited number of resources dedicated to pilots in LTE. Hence, each UE might see interference due to other co-scheduled UEs. Now, we will explain how $\tilde{\boldsymbol{R}_w}$ can

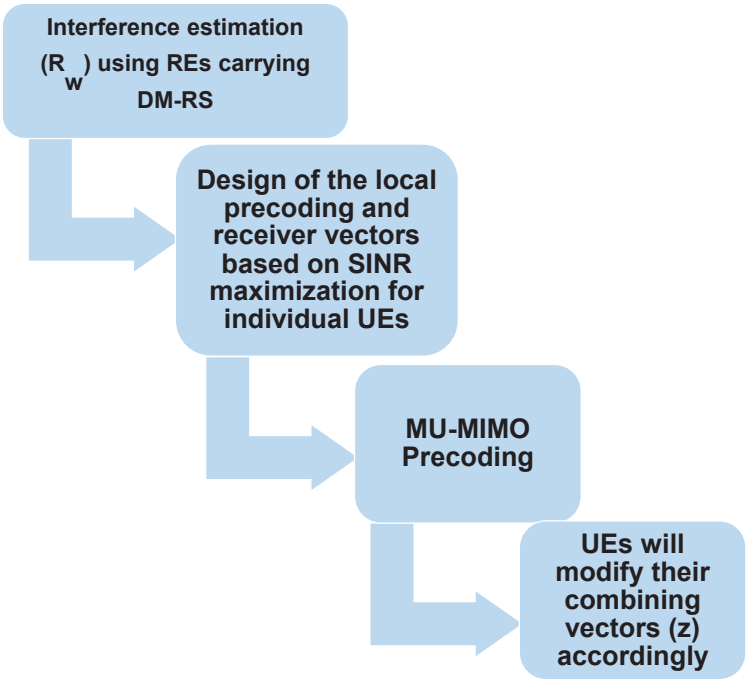



Figure 3.1: Algorithm Block Diagram

be updated to account for such interference. The received signal is:

$$\begin{aligned} \boldsymbol{y}(k,t) &= \boldsymbol{H}_i(k,t)\boldsymbol{v}_i(k,t)\boldsymbol{x}_i(k,t) \\ &+ \sum_{g=1,g\neq i}^{G} \boldsymbol{H}_i(k,t)\boldsymbol{v}_g(k,t)\boldsymbol{x}_g(k,t) + \boldsymbol{w}(k,t), \end{aligned} \tag{3.9}$$

Thus, the UE can update the interference plus noise covariance matrix as follows:

$\tilde{\boldsymbol{R}}_w = \sum_{g=1,g\neq i}^{G} \{\boldsymbol{H}_i(k,t)\boldsymbol{v}_g(k,t)\} \{\boldsymbol{H}_i(k,t)\boldsymbol{v}_g(k,t)\}^H + \boldsymbol{R}_w$

Thus, the final solution to the MU-MIMO receiver filters will be: $\boldsymbol{z}_i(k) = \alpha \tilde{\boldsymbol{R}}_w^{-1} \boldsymbol{H}_i(k)\boldsymbol{v}_i(k)$. The interference at each UE depends on the composite channel formed by the product of its own channel and the precoding vectors of other UEs. A signalling mechanism is needed to allow for successful decoding. An effective means for such signalling is to apply precoding vectors to UE-specific reference signals, allowing for the training of composite channels during data transmission. For more information, we refer readers to [82, Section 8.2.2].

## 3.4 Simulation Results

In this section, we describe the simulations conducted to study the proposed framework. The primary goal of our evaluation is to understand the gains provided by the proposed framework as compared to other approaches. We perform link-level and system-level simulations. In link-level, we assume that all UEs have the same average SNR. We simulate the encoding and decoding at the bit-level, using randomly generated information sequences with the selected modulation and coding scheme. In system-level, a multi-cell simulation is conducted. The interference is generated in the same way as desired signals, and channel propagation (pathloss, shadowing, and multipath) is taken into consideration. Our simulations follow the LTE standard [94]. Each eNodeB has 4 antennas, and 10 UEs (it co-schedules up to 4 UEs simultaneously using MU-MIMO). UEs are independently and randomly located. In summary, we use link-level to get an accurate relationship between SNR and throughput, then we use system-level to get accurate set of SNRs. Finally we get the corresponding accurate system-level throughput. The throughput is calculated by randomizing over several realizations of the UEs locations. Other simulation parameters are summarized in Table 3.1. We compare the performance of the designs shown in Table 3.2 (labeled from 'A' to 'F'). In LTE designs, UEs choose the precoding vector from a priori agreed codebook. Thus, UEs feedback the binary index of the chosen entry. In designs 'D' and 'E', we use the 4-bit LTE codebook. The remaining designs, we use beamforming, outside the LTE unitary precoding matrices known as non-codebook based MIMO approaches. Such non-codebook based MIMO approaches are supported by the structure in LTE Release 10 [8, Chapter 11]. Furthermore, we assume CSI knowledge at the eNodeB via feedback using CSI-RS pilots in Fig. 1, yielding errors in the CSI at the eNodeB.

Table 3.1: Simulation Parameters

| Parameter | Value |
|---|---|
| Frequency Band | 2 GHz |
| Transmission Bandwidth | 10 MHz |
| Number of subcarriers | 600 |
| Subcarrier spacing | 15 kHz |
| FFT size | 1024 |
| Inter-site distance | 500 m |
| Pathloss Model | $34.5 + 35log_{10}(d)$ dB |
| Shadowing SD | 8 dB |
| Channel Model | Typical Urban macro (Uma) |
| Noise Figure | 9 dB |
| Maximum Doppler | 5.55 Hz |

Table 3.2: Different Precoder and receiver designs

| | Precoder | Receiver |
|---|---|---|
| A | ZF over maximized SINRs | Maximized SINR |
| B | ZF Beamforming | IRC [95] |
| C | ZF Beamforming | MMSE [96] |
| D | LTE codebook | IRC [95] |
| E | LTE codebook | MMSE [96] |
| F | Iterative CoMP [97] | treat interference as noise [97] |

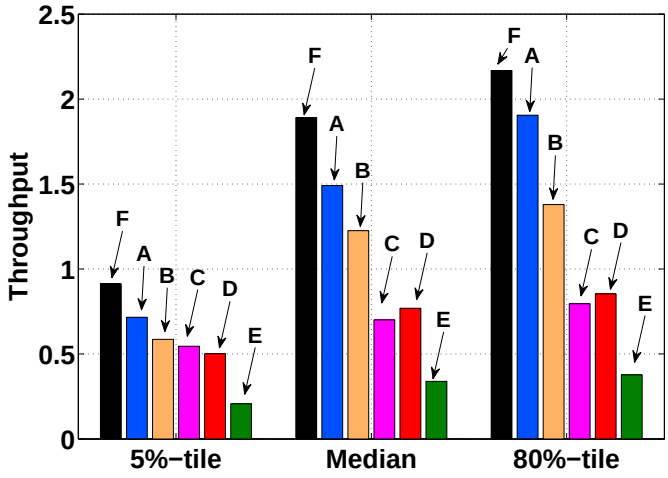



Figure 3.2: Performance comparison in terms of different percentiles

### 3.4.1 Throughput per UE Analysis

In this section, the throughput per UE is evaluated. In Fig. 3.2 we show the performance in terms of different throughput %-tiles. Design 'A' provides throughput improvement over all other designs except 'F', since 'F' uses cooperative resource allocation across eNodeBs. Design 'F' has throughput improvement of 28%, 25%, and 10% for 5 %-tile, 50%-tile, and 80%-tile as compared to 'A'. Note that, the gain is higher at lower %-tiles, which indicates that 'F' is more beneficial to UEs toward the cell-edge. Moreover, Design 'A' that uses DM-RS based covariance matrix estimation and 'B' that uses data based covariance matrix estimation are very close in terms of the 5%-tile throughput. This is because cross covariance is partially small for 5%-tile UEs due to the relatively small received power of the serving eNodeB [83]. However, for median and 80%-tile UEs, the performance of the data signal based covariance matrix is degraded compared to the DM-RS based covariance matrix estimation, 'A' has a gain of 25% and 35% for median and 80%-tile respectively as compared to 'B'.

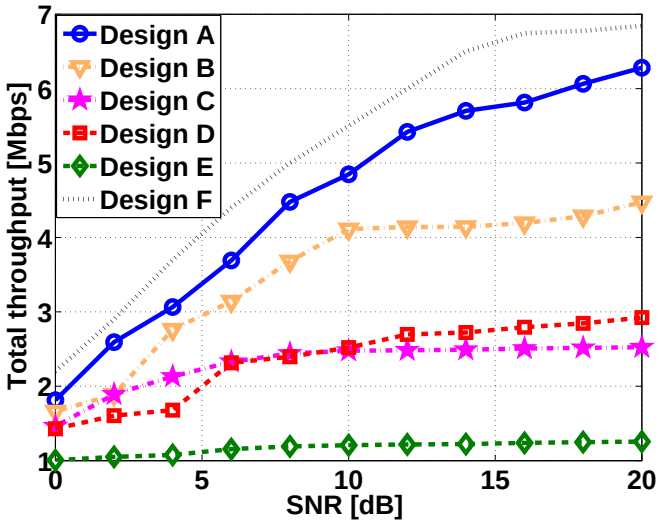



Figure 3.3: Throughput comparison

### 3.4.2 Total Throughput Analysis

In Fig. 4.5 we show the performance of the designs provided in Table 3.2. The performance of Design 'A' is slightly worse than 'F', because Design 'F' uses CoMP to optimize resource allocation which reduces the interference. Although, Design 'F' outperforms the proposed approach, it is an iterative approach and requires signaling overhead. However, Design 'A' provides high gains compared to other non-iterative techniques that can be achieved without cooperation and signaling overhead. Design 'F' aims at reducing the interference caused by neighbouring cells, and treating residual interference as noise [97]. As shown Design 'A' is about 13% worse than Design 'F', however, it outperforms all other designs by approximately 20%, 94%, and 300% with respect to Design 'B', Designs ('C' and 'D'), and Design 'E' respectively.

## 3.5 Conclusions

In this chapter, we present a practical non-iterative method for designing the precoder and the receiver for multi-user LTE systems. By comparing to other designs, we show considerable gains

to be achieved using our approach [98].

# Chapter 4

# Distributed Multi-user MIMO Wireless Networks With Full-Duplex Capability

## 4.1 Introduction

In-band full-duplex (IBFD) communication is very promising in enhancing wireless LANs, where full-duplex APs can support simultaneous UL and DL flows over the same frequency channel. One of the key challenges limiting IBFD benefits is interference. In this chapter, we propose a scheduling technique to manage interference in wireless LANs with full-duplex capability. We focus primarily on scheduling UL and DL clients that can be efficiently served simultaneously.

A common assumption made in prior work is that the client that is being served on DL is also the client that is sending UL packets to the AP. Thus, the interference is purely self interference. Network interference among clients will occur if different clients are considered for DL and UL, which may significantly deteriorate the throughput performance of IBFD wireless LANs.

Furthermore, MU-MIMO has also been studied to follow the trend of faster Wi-Fi. MU-MIMO has been considered in a number of wireless standards such as IEEE 802.11ac [99] and IEEE 802.11ax [100]. In MU-MIMO systems, each client can correctly decode packets simultaneously due to spatial diversity and precoding of channel weights by the transmitter. The total throughput, however, highly depends on the relationship between the channel responses and locations of scheduled clients. None of the prior work discussed scheduling multiple DL transmissions along with an UL transmission.

The problem is further compounded when MU-MIMO is used on the DL [89, 105–107], where APs use beamforming techniques to direct packets simultaneously to spatially diverse clients such as in Figure 4.1. That is, the AP will steer simultaneous beams to different clients, each beam containing specific packets for its target client.

To illustrate the key challenges of IBFD network interference consider Figure 4.2 which shows the interference signals resulting from having simultaneous UL and DL flows. When the UL receiver and the DL transmitter are active at the same AP simultaneously, self-interference is generated (shown as the solid red arrow). However, when the UL AP is different than the DL AP, network interference is generated (shown as the dashed red arrows). The figure assumes that one client is transmitting to one of the APs as an UL flow (shown as the solid blue arrow), and all the APs are transmitting to a set of DL clients, as DL flows (shown as the solid green arrows). The square in Figure 4.2 denotes the set of clients scheduled for DL MU-MIMO. In this case, the signal transmitted from the UL client can interfere with the DL clients. If the UL client is located close to the set of the DL clients, and the signal transmitted from the UL client is very strong, the DL clients will face high interference (shown as the dotted red arrow).

In order to mitigate the interference problem arising in such environment, a number of solutions

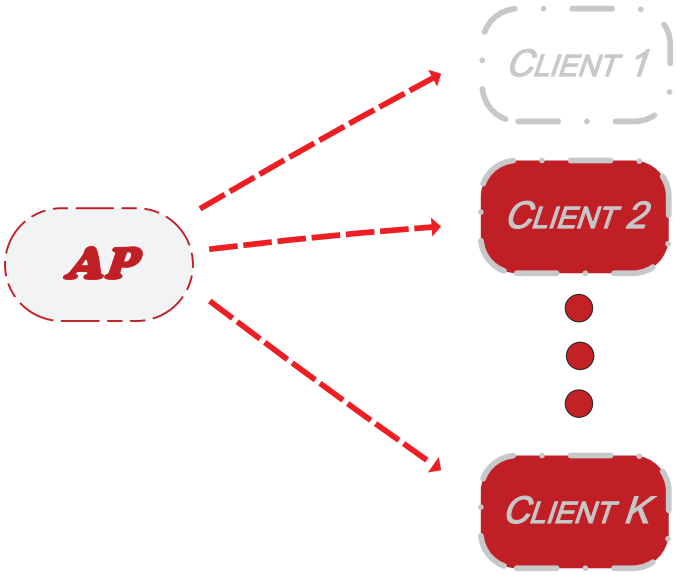



Figure 4.1: AP using MU-MIMO beamforming

have been proposed [62]-[71]. Those solutions capture additional transmission opportunities created by full-duplex by modifying contention and back-off mechanisms. In [62], the authors develop a centralized MAC protocol to support asymmetric data traffic where network nodes may transmit data packets of different lengths, and they propose to mitigate the hidden node problem by employing a busy tone. To overcome this hidden node problem, authors propose to adapt the 802.11 MAC protocol with the RTS/CTS handshake. In [70], authors study the power allocation for IBFD system where clients operate in the HD mode but the AP communicates by using the FD mode. In [70], the system model considers a single AP and multiple clients. The UL STA is chosen randomly, then a DL STA with low interference from the UL STA and high received power from the AP is selected. Afterwards, a power control algorithm is used such that the DL SINR and UL SINR satisfies a threshold [70].

A scheduling approach was studied for full-duplex wireless networks in [71], such that the AP

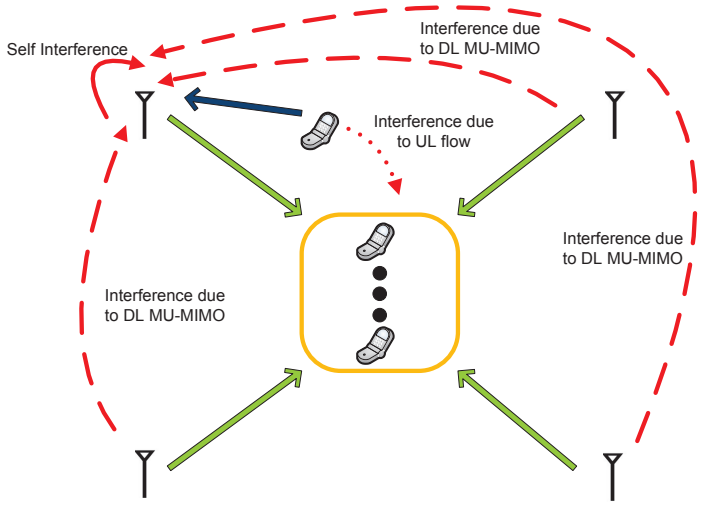



Figure 4.2: Interference in IBFD environment

has a pre-determined DL client and it aims at scheduling another UL client simultaneously. The AP randomly picks an UL client out of several ones that achieve a specific signal to interference (SIR) threshold at the DL client. A key shortcoming of the presented prior work is that any client that achieve a specific SIR at the DL client is considered a good candidate. Although, this type of optimization provides a guaranteed minimum throughput, it does not maximize the throughput. Moreover, in such schemes, finding a client that satisfies the SIR condition is done via exhaustive search over all the clients, which is time consuming.

This chapter focuses on clients scheduling at both the DL and the UL aiming at improving the sum rate in MU-MIMO wireless LANs with IBFD capability. In this chapter, we consider wireless LANs consisting of APs that are capable of full-duplex communications. We aim at managing interference, including interference due to DL MU-MIMO flows and interference due to the UL flow. To overcome the challenge of interference, we propose a scheduling technique that aims at serving a group of DL clients along with an UL client to be served simultaneously with minimal interference. Furthermore, the UL power is adapted to maximize the resulting sum throughput.

### 4.1.1 Contributions

Our main objective in this chapter is to maximize the achievable rate of wireless LANs with full-duplex APs serving multiple DL clients via MU-MIMO and an UL STA. Also, we consider that the UL client is not necessarily one of the DL clients. In other words, each client may or may not be served on UL and DL simultaneously.

The main results of this chapter are:

1. Clients categorization based on received signal strength indicator.
2. Channel access mechanism for clients through contention window adjustment procedure, which results in scheduling a group of DL clients along with an UL client simultaneously with minimal interference. We place no restrictions on the choice of the UL client, i.e. the UL client is not necessarily one of the DL clients
3. Power adaptation algorithm, which adjusts the UL transmit power aiming at maximizing the throughput.

The remainder of the chapter is organized as follows: Section 4.2 describes the problem formulation. In Section 4.3, we provide the scheduling and power adjustment technique. Section 4.4, provides the complexity analysis for the proposed technique. Simulation results are provided in Section 4.5 and we conclude the chapter in Section 4.6.

### 4.1.2 Notation

We use bold lower case for vectors, such as $\boldsymbol{a}$, while bold capital letters are used for matrices such as $\boldsymbol{A}$. Further $||\boldsymbol{A}||$ stands for the norm of the matrix $\boldsymbol{A}$. Further $(\cdot)^H$ stands for Hermitian transposition. $[\boldsymbol{A}]_{i,j}$ denotes the element in row $i$ and column $j$ of matrix $\boldsymbol{A}$. The cardinality of the set $A$ is denoted by $|A|$. Also $\boldsymbol{E}$ stands for expectation operator.

## 4.2 Problem Formulation

We consider an IBFD office wireless LAN scenario that consists of four APs, and comprises 64 cubicles. Each cubicle has four clients [100]. APs are assumed to have full-duplex capability. In other words, we consider that each AP can simultaneously transmit and receive. Throughout the paper, we will refer to the set of clients served on DL MU-MIMO as $S_{DL}$. $P_{UL}$ refers to the UL transmit power.

We assume that each client has $n_s$ antennas, and each AP has $n_a$ antennas. $n_A$ refers to the number of APs to perform MU-MIMO multiplied by the number of antennas per AP. The channel gains are modeled according to TGac channel model D [104] and are assumed to be constant over the duration of each transmission. Since serving different clients results in interference in different directions, Scheduling and Power Adaptation technique (SPA) plays an essential rule in enhancing the system performance.

The resulting received signal $\boldsymbol{y}_i^{dl} \in \mathbb{C}^{n_s}$ by the $i^{th}$ DL client is given by:

$$\boldsymbol{y}_i^{dl} = \boldsymbol{H}_i \boldsymbol{s}_i^{dl} + \sum_{k=1,k\neq i}^{K} \boldsymbol{H}_i \boldsymbol{s}_k^{dl} + \boldsymbol{F}_{j,i} \boldsymbol{s}_j^{ul} + \boldsymbol{n}_i, \tag{4.1}$$

where,

$$\boldsymbol{H}_i = \begin{bmatrix} \boldsymbol{H}_{1i} \\ \vdots \\ \boldsymbol{H}_{ai} \\ \vdots \\ \boldsymbol{H}_{Ai} \end{bmatrix}^T \tag{4.2}$$

$\boldsymbol{H}_i$ is $n_s \times n_A$ matrix that represents the channel between the $i^{th}$ client and all APs, $\boldsymbol{H}_{ai}$ is $n_a \times n_s$

sub matrix that represents the channel between the $a^{th}$ AP and the $i^{th}$ client, $\boldsymbol{F}_{j,i}$ is $n_s \times n_s$ matrix that represents the interference channel from the UL client (served by the $j^{th}$ AP) to the DL client $i$ due to the UL flow, $\boldsymbol{s}_j^{ul} \in \mathbb{C}^{n_s}$ is the transmit signal of the UL client, and $\boldsymbol{n}_i$ is the noise vector at the $i^{th}$ client.

The resulting received signal by the $j^{th}$ AP that is serving the UL client $\boldsymbol{y}_j^{ul} \in \mathbb{C}^{n_a}$, is given by:

$$\boldsymbol{y}_j^{ul} = \boldsymbol{H}_{ju}\boldsymbol{s}_j^{ul} + \sum_{a=1,a\neq j}^{A}\sum_{k=1}^{K}\boldsymbol{E}_{a,j}\boldsymbol{s}_k^{dl} + \boldsymbol{z}_j + \boldsymbol{n}_j, \tag{4.3}$$

where,

$$\boldsymbol{z}_j = \beta\sum_{k=1}^{K}\boldsymbol{E}_{j,j}\boldsymbol{s}_k^{dl} \tag{4.4}$$

$\boldsymbol{H}_{ju}$ is $n_a \times n_s$ sub matrix that represents the channel between the $j^{th}$ AP and the scheduled UL client, $\boldsymbol{E}_{a,j}$ is the $n_a \times n_a$ matrix that represents the channel between the $a^{th}$ AP and the $j^{th}$ AP, $\boldsymbol{s}_k^{dl} \in \mathbb{C}^{n_a}$ is the transmit signal of the $k^{th}$ DL client. $\boldsymbol{z}_j$ is the self-interference, $\beta$ is the self interference cancellation coefficient, $\boldsymbol{n}_j$ is the noise vector, and $K$ is the number of co-scheduled clients in DL MU-MIMO.

The first term in (4.1) represents the intended signal, the second term represents the co-layer interference, the third term represents the IBFD network interference, and the final term represents the additive noise. In (4.3), the first term is the intended signal in the UL direction, the second term is the interference resulting from serving the DL clients, the third term is the self interference, and the final term is noise. We define the SINR of an UL and DL flow as follows:

$$SINR_{UL} = \frac{||\boldsymbol{H}_{ju}\boldsymbol{s}_j^{ul}||^2}{\sum_{a=1,a\neq j}^{A}\sum_{k=1}^{K}||\boldsymbol{E}_{a,j}\boldsymbol{s}_k^{dl}||^2 + \sum_{k=1}^{K}||\beta\boldsymbol{E}_{j,j}\boldsymbol{s}_k^{dl}||^2 + \eta_j} \tag{4.5}$$

$$SINR_{DL} = \frac{||\boldsymbol{H}_i \boldsymbol{s}_i^{dl}||^2}{\sum_{k=1,k\neq i}^{K} ||\boldsymbol{H}_i \boldsymbol{s}_k^{dl}||^2 + ||\boldsymbol{F}_{j,i} \boldsymbol{s}_j^{ul}||^2 + \eta_i} \tag{4.6}$$

We define the achievable total sum-rate as follows:

$$R_{tot} = \log_2(1 + SINR_{UL}) + \sum_{i=1}^{K} \log_2(1 + SINR_{DL}) \tag{4.7}$$

where $\eta_j$ is the noise power at the $j^{th}$ AP and $\eta_i$ is the noise power at the $i^{th}$ client. The first and second terms denote the UL and DL rates, $R_{ul}$ and $R_{dl}$ respectively.

## 4.3 SPA: Scheduling and Power Adaptation

We propose a scheduling and power aadaptation technique (SPA) for IBFD wireless LANs. Traditionally, an AP is solely using an exclusive RF channel to limit interference via frequency reuse [100]. Theoretically, IBFD can be applied at each AP, thus an AP would support an UL and DL. However, viable IBFD choices will be limited due to the proximity of clients resulting in significant network interference. To solve the network interference problem, we propose that all four APs in the example scenario presented perform distributed MU-MIMO utilizing the aggregated bandwidth. Thus, the network serves multiple clients in the DL at a higher capacity via MU-MIMO while supporting an UL link via IBFD. The main benefit that can be earned when this model is considered in practical environment, is that there is a better chance of finding clients eligible for IBFD as the physical space that all APs are covering is larger than each AP alone.

The following general system considerations are presented:

**`System Consideration 1:`** The selected UL client should be spatially separated from the DL clients to reduce co-channel interference.

**`System Consideration 2:`** DL clients should be spatially separated to maximize MU-

MIMO DL rates [89, 105, 106].

Figure 4.3 shows the importance of the system considerations discussed above. The y-axis represents the sum rate, where 4 APs are located on the vertices of a square with a side length of 10m. An UL client is chosen randomly and is considered as a center of a circle where 4 DL clients are equally-spaced on its circumference. By increasing the diameter of the circle, the DL clients gets further away from the UL client. Besides, the DL clients get further from each other. An example for the setup is shown in figure 4.4, where clients on the same circle are scheduled for DL simultaneously using MU-MIMO, while the client in the center of the circle is scheduled on the UL.

In figure 4.3, the sum rate is computed with respect to different circle diameter values. As shown, when the circle diameter is higher, i.e when the DL clients are far from the UL client and are far from each other, the inter-client interference from the UL client is weak and the MU-MIMO gain is higher. Thus the sum rate becomes high, as shown in Figure 4.3. On the other hand, a small circle diameter means a strong inter-client interference from the UL client towards the DL clients and also, DL clients are very close to each other, as a result, the sum rate is reduced.

### 4.3.1 Clients Categorization

In order to categorize clients we propose the use of controller unit (CU). One of the APs can act as the CU. The CU will be responsible of all aspects of MU operation. The CU will store sorted vectors of the received signal strength indicator (RSSI) indices of the APs as measured by the clients. i.e., a client with: $AP_a, AP_b, AP_c, AP_d$, has high RSSI from the $a^{th}$ AP, and low RSSI from the $d^{th}$ AP. Since in the office scenario we assume 4 APs, the outcome will be a lookup table with 24 (4!) categories.

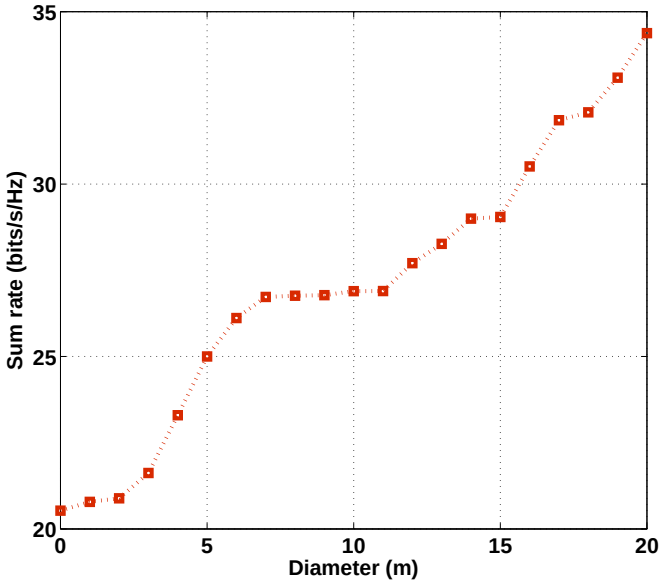



Figure 4.3: Sum rate with DL clients on a circle circumference and an UL client on its center

### 4.3.2 Contention Window Adjustment Procedure

The 802.11 protocol uses a carrier sense multiple access (CSMA) scheme, where channel needs to be idle for any transmission or reception. When channel is idle, a backoff timer is randomly chosen over the interval of $[0, CW]$, where $CW$ stands for contention window size. In this paper, we propose $CW$ adjustment mechanism, the proposed mechanism maintains backward compatibility. The legacy clients will still be able to demodulate and decode packet headers, and backoff when the medium is busy.

Initially, an UL client is selected based on CSMA. Depending on the category (RSSI vector) of this UL client, it is better to schedule DL clients belonging to categories far from the UL client. In other words, to reduce interference with the UL client, it is better to schedule DL clients with RSSI vector with least significant digit equal to the most significant digit of the UL client. i.e., if the UL client has $AP_1, AP_2, AP_3, AP_4$, DL clients is preferred to belong to the following:$(AP_4, AP_3, AP_2, AP_1)$,$(AP_4, AP_2, AP_3, AP_1)$,$(AP_3, AP_4, AP_2, AP_1)$

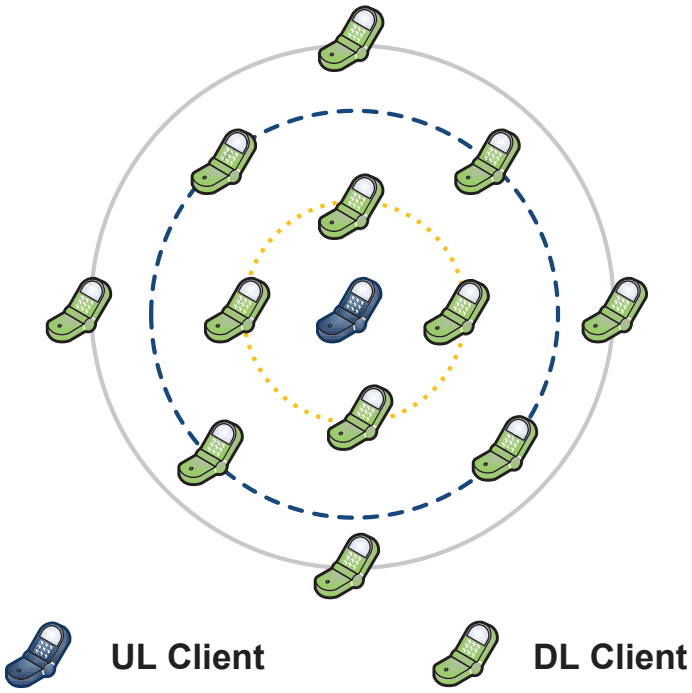



Figure 4.4: Example with DL clients on circle's circumference and an UL client on its center

,$(AP_3, AP_2, AP_4, AP_1)$,$(AP_2, AP_4, AP_3, AP_1)$,$(AP_2, AP_3, AP_4, AP_1)$.

Thus, using clients categorization, the $(CW)$ size needs to be designed which controls the backoff counters, such that clients belonging to the above 6 categories, get the smallest $CW$ size. However, in some cases, based on the relative differences of signal strengths from APs, this potential client may not be a good candidate in terms of increasing the DL MU-MIMO rate benefits. Thus, the potential DL client will only be added to the set of scheduled DL clients $S_{DL}$, if the condition below is satisfied:

$$R_{ul}^{p+1} + R_{dl}^{p+1} + R_{potential} \geq R_{ul}^{p} + R_{dl}^{p}, \tag{4.8}$$

where, $R_{dl}^{p}$ is the rate of the scheduled DL clients at the $p^{th}$ iteration, and the $CW$ of all clients belonging to the same category will increase. However, if the rate condition is not satisfied, that client will solely increase its $CW$. The $CW$ adjustment procedure is explained in Table 4.1, and

channel access mechanism is explained in the flow chart in Fig 4.5:

Table 4.1: $CW$ Adjustment Procedure

| |
|---|
| 1: Initially, all clients have same $CW = CW_{int}$ |
| 2: After UL is selected, $CW = \alpha CW_{int}$,where, $\alpha = \frac{1}{\boldsymbol{a}_i^u}$ ,$\boldsymbol{a}_i^u$ is the index of the UL AP within the client's sorted RSSI vector |
| 3: If a client fails the rate condition, its $CW$ is increased. |
| 4: If a client passes the rate condition, the $CW$ of other clients belonging to same category is increased. |

### 4.3.3 Power Adaptation

To improve the performance, the UL power $P_{UL}$ needs to be adjusted. Initially, the UL client uses full power. If the rate condition is not satisfied, $P_{UL}$ is reduced, and same steps are repeated. If the rate condition fails again, $P_{UL}$ is reduced until reaching a minimum power $P_{min}$ that satisfies an UL SINR threshold. SPA algorithm is explained in Table 4.2. It is important to note that, the selected $P_{UL}$ is based on the rate, however, in wireless networks, it is important to enhance the throughput, which takes into account both rate and packets errors. Therefore it has become rather important decision to update the $P_{UL}$ adaptively based on throughput.

Thus, the first transmission/reception event for a set of clients will be based on the algorithm discussed in Table 4.2. However, upon completing each transmission/reception event, the status will be checked. The goal is to use the results of every transmission/reception event (i.e packets are acked or not, etc.) to increase or decrease $P_{UL}$ accordingly. After each event, the throughput can be computed as follows:

$$T = (1 - PER_{UL}) * R_{UL} + (1 - PER_{DL}) * R_{DL} \tag{4.9}$$

Then, the algorithm needs to decide whether to reduce or increase $P_{UL}$. Our ultimate goal is to be able to estimate $T_l$ and $T_h$, which is the throughput at lower and higher $P_{UL}$ respectively. Then,

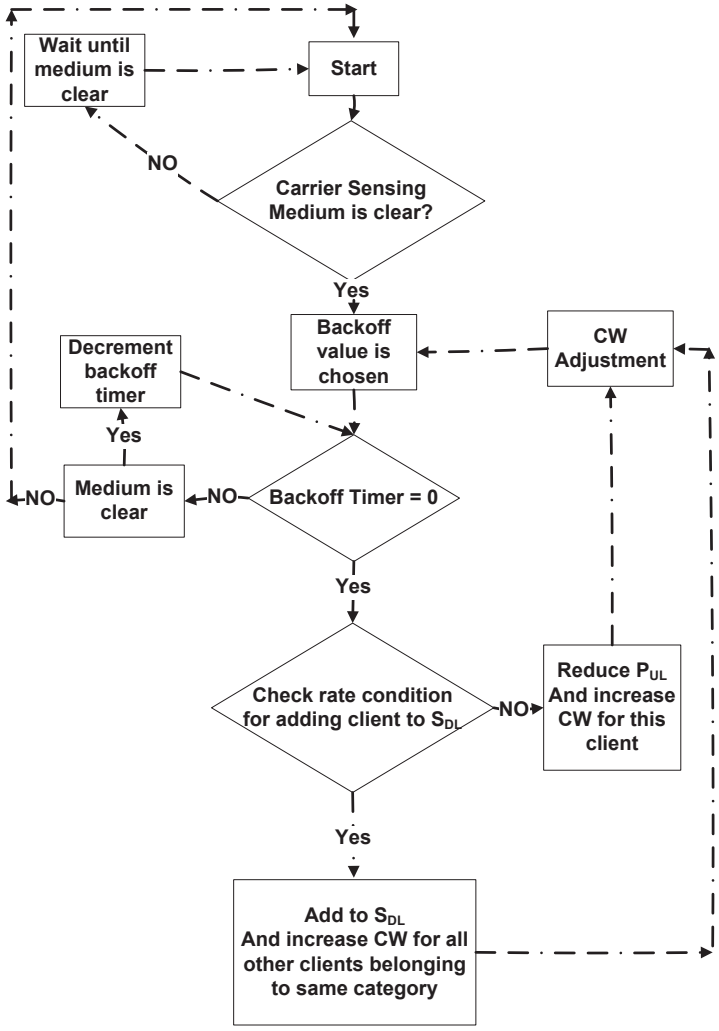



Figure 4.5: Channel Access Algorithm

Table 4.2: SPA Algorithm

1: Categorize clients based on sorted RSSI indices
2: UL client is selected
3: Update CW of clients based on the UL client and step 1
4: Initialize: $S_{DL} = 0$ and $P_{UL} = P_{max}$
5: **while** $P_{UL} > P_{min}$
6: **while** $|S_{DL}| < n_A$
7: Select a potential DL client
8: **if** $R_{ul}^{p+1} + R_{dl}^{p+1} + R_{potential} \geq R_{ul}^{p} + R_{dl}^{p}$
9: Add potential client and update $S_{DL}$ accordingly
10: Increase CW of all clients belonging to the same category
11: Select a new potential client
12: **else**
13: Increase the contention window of this potential client
14: break from while loop
15: **end if**
16: **end while**
17: $P_{UL} = P_{UL} - \Delta$
18: **end while**

the algorithm can select $P_{UL}$ accordingly.

However, there is a challenge on computing $T_h$ and $T_l$ before the transmission/reception, since the $PERs$ are measured after the event completion. We propose estimating $T_h$ and $T_l$ and using them in the power adaptation algorithm as explained below.

**1. Primary transmission/reception event:**

When the link is just established, use the primary $P_{UL}^{p}$ selected by SPA algorithm in Table 4.2. Upon the completion of the event, use the information such as: number of packets that have been successfully received, total number of packets transmitted to compute packet success ratio $(PSR = 1 - PER)$. Then compute the primary throughput $T_p$

**2. Secondary transmissio/reception event:**

Calculate $PER_{UL}$ and $PER_{DL}$ and do the following:

if $PER_{DL} > PER_{UL}$

$$P^s_{UL} = P^p_{UL} - \Delta \tag{4.10}$$

else

$$P^s_{UL} = P^p_{UL} + \Delta \tag{4.11}$$

Similar to step 1, compute the secondary throughput $T_s$, then the event that leads to higher throughput will be used as a current initial throughput $T_c$ as follows:

$$T_c = \max(T_p, T_s) \tag{4.12}$$

$P^c_{UL}$ is either the primary or the secondary $P_{UL}$ based on the selection of $T_c$.

**3. Following events:**

At this step, we have valuable information from primary and secondary events. We have rate, PSR, and throughput for primary and secondary events. An example is shown in Figure 6. Note that, $P_{UL}$ affects throughput, by affecting both rate and PSR. The effect on rate is known before transmission/reception. However, the effect on PSR is only known after the completion of the event. In this step, our target is to tune $P_{UL}$ with a small tunable $\delta$, such that:

$$\begin{aligned} P^n_{UL} &= P^c_{UL} - \delta, \quad \text{if} \quad T_c < T_l \\ &= P^c_{UL} + \delta, \quad \text{if} \quad T_c < T_h \end{aligned} \tag{4.13}$$

where, $P^n_{UL}$ is the new UL power.

In order to estimate $T_l$ and $T_h$, we need to estimate $PSRs$ at both points. For that purpose, we use the primary and secondary points as shown in figure 6, and perform interpolation/extrapolation to

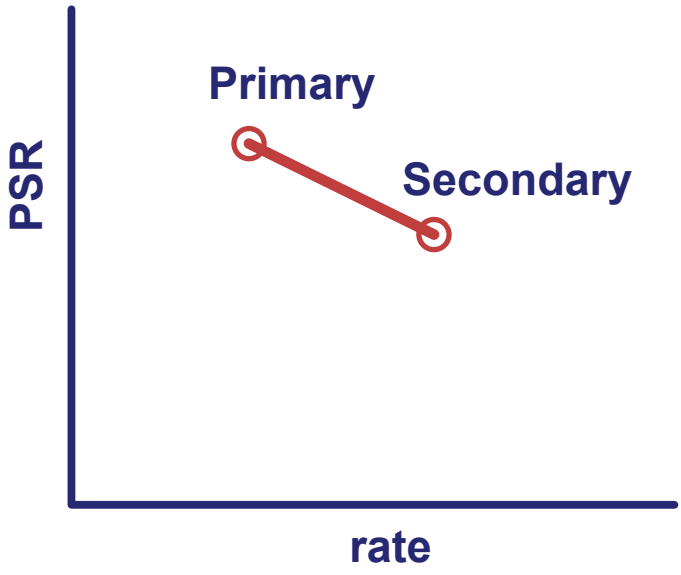



Figure 4.6: Primary and Secondary information

find $PSR_l$ and $PSR_h$. After doing so, we can get $T_h$ and $T_l$ and select the one that maximizes the throughput.

So, in summary, we keep the PSR estimates at four points. We update those points upon each transmission/reception event, we need to update the PSR estimates according to the exponential moving average as follows:

$$PSR^n = \gamma PSR^{n-1} + (1-\gamma) * \frac{n_{suc}}{n_{tot}}, \tag{4.14}$$

where, $PSR^n$ is the new estimate, $PSR^{n-1}$ is the previous estimate, $\gamma \in [0,1]$ is the aging factor, and $n_{suc}$ is the number of successful packets, and $n_{tot}$ is the total number of packets.

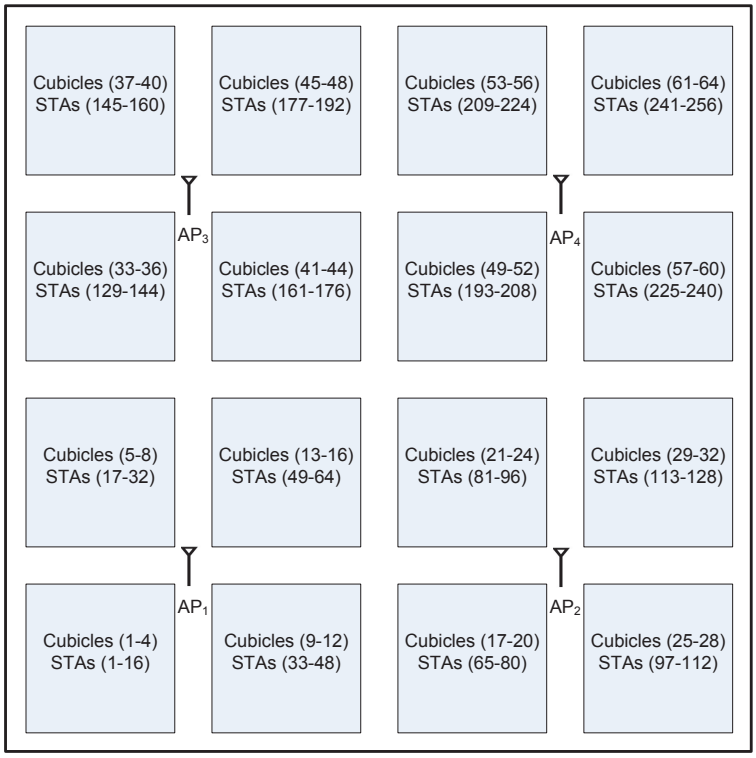



Figure 4.7: Office wireless LANs scenario

## 4.4 Simulation Results

Our simulation follows the office environment described in Section 4.2 and shown in figure 4.7. The position of the APs is fixed, and clients are randomly distributed inside each cubicle. The main simulation parameters are summarized in Table 4.3 [108, 109]. We compare the performance of SPA with that of IBFD with power control that is presented in [70], IBFD without power control and also half-duplex conventional scenario. It is important to note that [70] is only applicable for a single AP, thus, IBFD in [70] is implemented for each AP separately.

### 4.4.1 Rate Comparisons

Figure 4.8 shows the sum rate for different algorithms. The rate in the y-axis is a sum rate of co-scheduled clients. As shown, the rate of IBFD without power control is worse than HD, because

Table 4.3: Simulation Parameters

| Parameter | Value |
|---|---|
| Office Area | 20 m $\times$ 20 m |
| clients Locations | randomly distributed within each cubicle |
| DL Power | Different across APs based on MU-MIMO |
| UL Power | satisfy the lowest UL MCS Level 2 |
| Frequency Band | 5 GHz |
| Channel Bandwidth | 80 MHz |

the DL rate will be affected by high interference generated by the UL client. In contrast, the rate of IBFD system increases when power control is added. However, the high gains of IBFD can not be achieved using the power control algorithm in [70]. As shown, HD and IBFD with power control [70] are close to each other, which is expected since network interference is limiting the benefits of IBFD. Thus, the power control algorithm in [70] cannot utilize IBFD capability in the office scenario. This is due to the fact that, choices are limited due to the proximity of clients. i.e. the network interference caused by the UL will significantly reduce the SINR at the DL clients. However, SPA can overcome this problem, since SPA has a better chance of finding clients that are eligible for IBFD. i.e. SPA benefits from spatial separation. As shown SPA algorithm outperforms all other algorithms by approximately 150%, 268%, and 101% with respect to HD, IBFD without power control, IBFD with power control [70] respectively. It is important to note that more than twice the rate is achieved by SPA algorithm compared with traditional HD, due to the MU-MIMO gains.

### 4.4.2 Fairness Index

Figure 4.9 shows the fairness index for different IBFD algorithms. IBFD with SPA achieves comparable fairness index to the algorithm in [70]. That is, the clients under SPA can be provided with fair scheduling opportunities. Note that, SPA is adaptively making sure that UL and DL flows are achieving comparable good throughput.

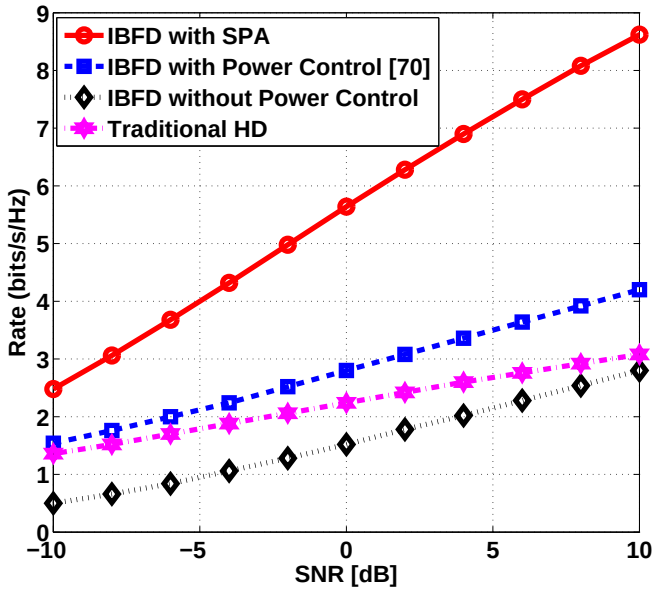



Figure 4.8: Rate comparison for office scenario

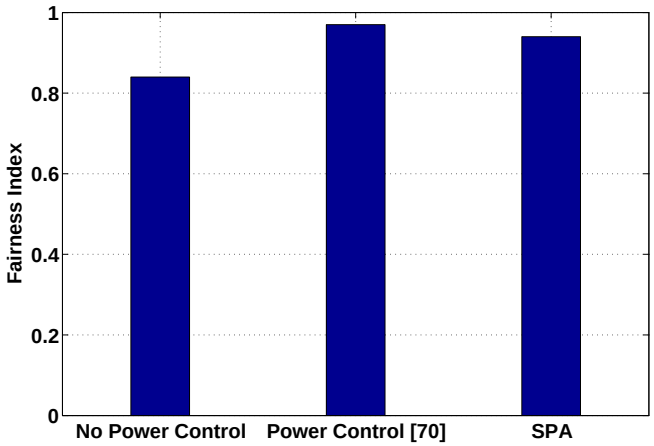



Figure 4.9: Fairness index comparison

### 4.4.3 Impact of Self Interference

In this paper, APs are equipped with elaborate antenna techniques and signal processing modules for self interference cancellation. In previous simulations, we assumed perfect self interference cancellation. Here, we show the impact of imperfect self-interference on different algorithms. Figure 4.10 shows average SINR for UL an DL clients with respect to self interference cancellation. The SINR of both UL and DL of IBFD increase as the self interference cancellation increases, since self interference cancelation directly benefits the UL client, and indirectly benefits the DL clients due to the power adaptation scheme. Also, IBFD with power control in [70] can benefit from self interference cancellation in both UL and DL directions. However, it cannot sufficiently overcome the problem caused by the proximity of clients resulting in significant network interference, and the SINR performance is then deteriorated. On the other hand, in the case of the IBFD without power control, UL SINR increases as self interference cancellation increases, while DL-SINR does not change, since the DL flow will suffer from same interference regardless of self interference cancellation. Figure 4.11 shows the difference between IBFD with SPA and IBFD in [70] in terms

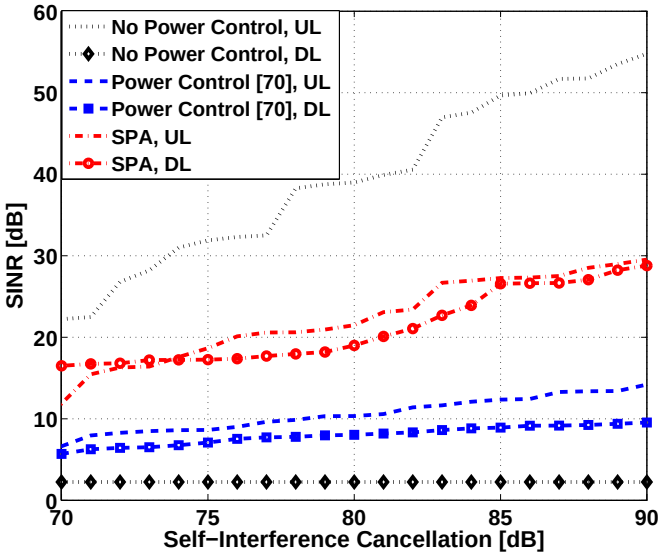



Figure 4.10: SINR comparison for different self interference cancellation

of total sum rate. In the office scenario, IBFD in [70] can serve up to 8 clients, on the other

hand, IBFD with SPA can only serve up to 5 clients simultaneously. However, the sum rate of SPA exceeds the algorithm in [70] as shown in 4.11. Note that, in [70] the average inter-client interference between clients increases because the distance between clients shorten, hence the rate is degraded. Moreover, due to the distributed MU-MIMO model that is utilized in SPA, clients can get higher throughput opportunities.

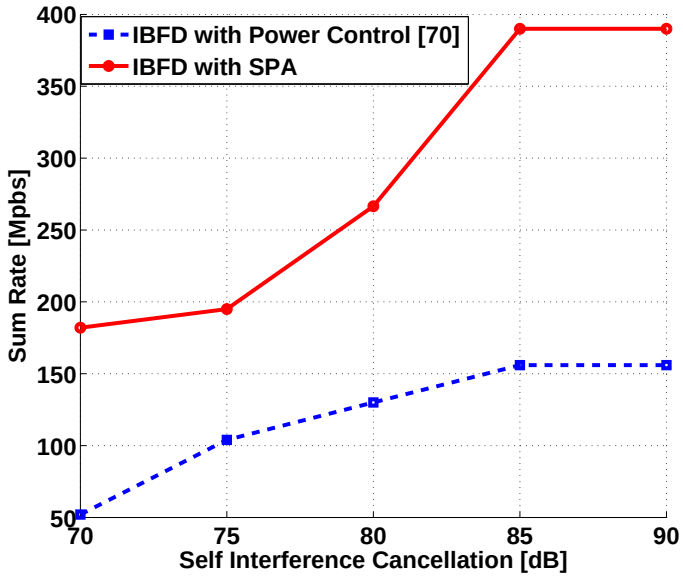



Figure 4.11: Sum rate comparison for different self interference cancellation

### 4.4.4 MCS Levels Comparison

Table IV shows the MCS levels of active links. Active link is any scheduled DL or UL flow. Since SPA utilizes spatial separation, it provides high operation percentage on high MCS levels(7 and 8) can be achieved approximately with 38.28%, 0.78%, and 48.59% using IBFD with SPA, with power control [70], and without power control respectively. The IBFD without power control achieves higher percentage than SPA, because without power control, UL clients gets high SINRs on the expense of DL clients getting very low SINRs. As can be noticed, SPA provides the lowest percentage of low MCS levels.

Table 4.4: MCS levels of active links for IBFD

| | SPA | power control [70] | no power control |
|---|---|---|---|
| MCS 7-8 | 38.28% | 0.78% | 48.59% |
| MCS 5-6 | 30.80% | 0.79% | 0.65% |
| MCS 3-4 | 30.60% | 20.04% | 0.48% |
| Lower | 0.32% | 78.39% | 50.28% |

## 4.5 Conclusions

In this chapter, we present scheduling and power adaptation technique to provide higher performance in the IBFD environment for office wireless LANs. The proposed approach can provide good IBFD opportunities. Our proposed algorithm aims at selecting clients that can efficiently be served simultaneously with low interference between UL and DL transmissions. At a given time, an UL client is scheduled and its power is adapted while selecting multiple DL clients taking the IBFD interference into account. Simulation results to evaluate the system performance is presented, which show significant increase in rate compared to recent proposed scheduling and power control algorithms for IBFD [111].

# Chapter 5

# MIMO Cellular Systems With Power Amplifiers

## 5.1 Introduction

OFDMA is the modulation of choice due to its robustness to time-dispersive radio channels, low-complexity receivers, and simple combining of signals from multiple transmitters in broadcast networks. However, the transmitter design for OFDMA is more costly, as the PAPR of an OFDMA signal is relatively high, resulting in the need for highly linear RF power amplifiers (PA). This problem becomes more compounded when a large number of PAs is required, as in Massive MIMO. In this chapter, we discuss the impact of PAs on cellular systems. We show the constraints that PAs introduce, and we take these constraints into consideration while searching for the optimum set of transmitter and receiver filters. Moreover, we highlight how Massive MIMO cellular networks can relax PAs constraints resulting in low cost PAs, while maintaining high performance. The performance is evaluated by showing the probability of error curves and signal-to-noise-ratio curves for different transmit powers and different number of transmit antennas.

### 5.1.1 Power Amplifiers

An ideal PA would produce as its output a perfect replica of the input multiplied by a scalar value. However, practical PAs exhibit various nonlinearities. Numerous practical PA models exist, a selection of which can be found in [101]. Here we describe the solid state power amplifier (SSPA) model presented at [102], where the output signal can be written as:

$$f(r) = \frac{r}{[1 + (\frac{r}{O_s})^{2\beta}]^{\frac{1}{2\beta}}}, \tag{5.1}$$

where $r$ is the input signal, $O_s$ is the saturation output, and $\beta$ is a parameter that controls the smoothness of the transition from linear to nonlinear operation. Conventionally, a PA with a wide linear region is preferable, to reduce the effect of distortion. To ensure that, a safety region between the linear and nonlinear region is required, and is known as the ”backoff region”. The width of the backoff region depends on the expected PAPR in the communication system. It is important to note that there is a tradeoff between the width of the linear region and the cost of the PA. In other words, PAs with relatively low cost have narrow linear region, while PAs with wide linear region have a high cost. Both the linear region and backoff region define the operating point of the PA either by its input or its output.

### 5.1.2 Contributions

In this chapter, we aim at reducing the power consumption at the BS and optimizing both the transmitter and receiver filter design to avoid the occurrence of nonlinear distortion. First, we aim at operating within the reduced linear region that will be constrained by the PAs cost. Second, we aim at improving the efficiency of the PAs by relaxing the signal total power constraints.

The main results of this chapter are:

1. Discuss the impact of PAs on cellular systems.

2. Show the constraints that PAs introduce.

3. Take PA constraints into consideration while searching for the optimum set of transmitter and receiver filters.

4. Highlight how Massive MIMO cellular networks can relax PAs constraints resulting in low cost PAs, while maintaining high performance.

The remainder of this chapter is organized as follows: Section 4.2 describes the problem formulation. In Section 4.3, we explain the transmitter and receiver filter design in order to account for PA non-ideality. Section 4.4, we discuss the relationship between the PA linear region operation and MIMO order. In Section 4.5 we provide remarks and notes on the proposed design. Simulation results are provided in Section 4.6 and we conclude the chapter in Section 4.7.

### 5.1.3 Notation

We use bold lower case for vectors, such as $\boldsymbol{a}$, while bold capital letters are used for matrices such as $\boldsymbol{A}$. Further $||\boldsymbol{A}||$ stands for the norm of the matrix $\boldsymbol{A}$. Further $(\cdot)^H$ stands for Hermitian transposition. $[\boldsymbol{A}]_{i,j}$ denotes the element in row $i$ and column $j$ of matrix $\boldsymbol{A}$. The cardinality of the set $A$ is denoted by $|A|$. Also $\boldsymbol{E}$ stands for expectation operator.

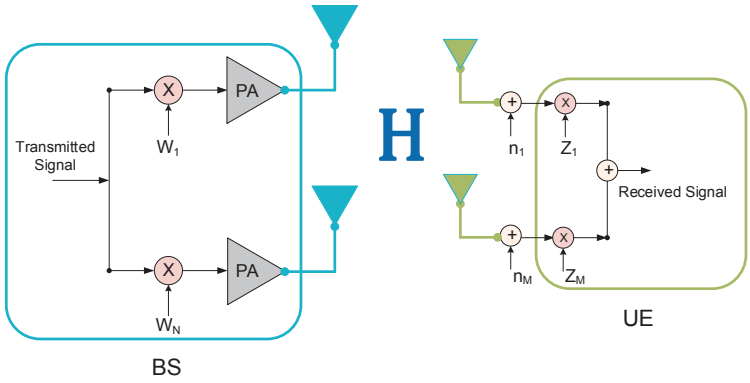



Figure 5.1: System Model

## 5.2 Problem Formulation

### 5.2.1 System Model

We consider a cellular network, where each cell consists of a home BS with $N$ antennas and several UEs with $M$ antennas each. Hence, the link between the BS and each UE can be represented as $N$ X $M$ MIMO system. An example of such a model is represented in Fig. 5.1. In this paper, we focus on the effect of PAs on the transmitted signal. As shown in Fig. 5.1, $\boldsymbol{w}$ is the transmit vector with $N$ elements and $\boldsymbol{z}$ is the receiver combining vector with $M$ elements.

### 5.2.2 Maximizing SNR

We design the transmit vector $\boldsymbol{w}$ indirectly through designing the vector at the output of the PA $\boldsymbol{v}$. Moreover, we design the receiver filter $\boldsymbol{z}$ so as to maximize the signal to noise ratio (SNR). The received signal by the UE will be:

$$\boldsymbol{y} = \boldsymbol{z}^H \boldsymbol{H} \boldsymbol{v} \boldsymbol{x} + \boldsymbol{z}^H \boldsymbol{n}, \tag{5.2}$$

where $\boldsymbol{H}$ is the channel between the BS and the UE, $\boldsymbol{x}$ is the transmitted signal, and $\boldsymbol{n}$ is the additive noise with covariance matrix $\boldsymbol{R}_n = \boldsymbol{E}[\boldsymbol{n}\boldsymbol{n}^{\boldsymbol{H}}]$, $\boldsymbol{z}$ and $\boldsymbol{v}$ are the receiver and the vector at the output of the PA respectively. It is important to note that noise is colored due to interference from other transmissions. Now, we aim at maximizing the SNR of the received signal as:

$$\max_{\boldsymbol{z},\boldsymbol{v}} \frac{|\boldsymbol{z}^H \boldsymbol{H} \boldsymbol{v}|^2}{\boldsymbol{z}^H \boldsymbol{R}_n \boldsymbol{z}} \sigma_x^2 \text{ subject to } ||\boldsymbol{v}||^2 = P_T. \tag{5.3}$$

Where $\boldsymbol{E}[|\boldsymbol{x}|^2] = \sigma_x^2$ and $P_T$ is the transmit power.

## 5.3 Transmitter and Receiver Filters

According to the generalized eigenvalue problem [78], the solution to (5.3) is:

$$\begin{aligned} \boldsymbol{z} &= \alpha\ \boldsymbol{R}_n^{-1} \boldsymbol{H} \boldsymbol{v} \\ \boldsymbol{v} &= \gamma\ v_{max}[\boldsymbol{H}^H \boldsymbol{R}_n^{-1} \boldsymbol{H}], \end{aligned} \tag{5.4}$$

where $\alpha$ and $\gamma$ adjust the power of $\boldsymbol{z}$ and $\boldsymbol{v}$ respectively. Then, by using the PA characteristics, we can find the vector $\boldsymbol{w}$ at the input of the PA. An example to the characteristics of such PA is shown in Fig. 5.2, where $O_s = \beta = 1$ is assumed. In order to guarantee that each output has a corresponding input, each element in the designed vector $\boldsymbol{v}$ has to be less than the PA operating point.

## 5.4 Linear Region Operation

it is important to answer the following question: How likely will the designed transmit vector (PA output) be below the operating point? In order to answer the question above, cumulative

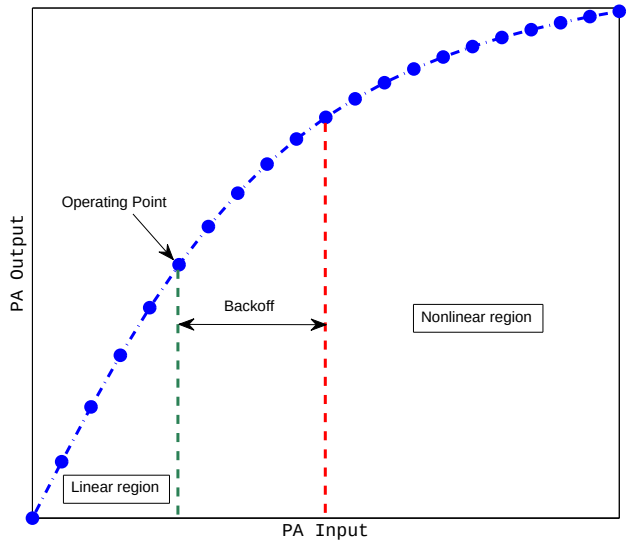



Figure 5.2: SSPA Input-Output Characteristics

distribution function (CDF) simulations of the PA output have been conducted. The results can be summarized as in Table 5.1. Without loss of generality, the operating point here is assumed to be unity for simplicity.

According to the table, as the number of antennas at the BS $N$ increase, the transmit power limits can also be increased while having a relatively high percentage of PAs operating in the linear region. It is worth mentioning that also increasing the number of antennas at the receiver $M$ can lead to higher system performance due to diversity gain.

It is obvious from Table 5.1 that there are scenarios where the occurrence of nonlinear operation is extremely rare for example, when the BS, has more than 8, 16 antennas for transmit power of 2 and 4 respectively. For massive MIMO, where the number of antennas at the BS can reach several hundreds, the probability of operating in the nonlinear region is almost zero even if the transmit power is relatively high. As will be shown in section **??**, for BSs with 100 antennas, the transmit power can reach up to around 20 with 100% linear region operation.

Table 5.1: linear region operation percentages

| Number of BS antennas | $P_T = 2$ | $P_T = 4$ | $P_T = 8$ |
|---|---|---|---|
| 2 | 50% | 25.1% | 12.6% |
| 4 | 86.9% | 54.4% | 33.3% |
| 8 | 99.2% | 86.8% | 60.5% |
| 16 | 100% | 98.8% | 86.5% |
| 32 | 100% | 99% | 86.7% |

## 5.5 Remarks and Notes

The main advantage of JTR is that the transmitted signal will not be distorted by the PA non-linearity since the number of BS antennas and the total transmitted power shall be designed to avoid non linear operation. Hence, JTR can be used with the aim of saving energy, where low cost PAs with smaller linear region are used to reduce the power consumed to operate the PAs in the BSs. Also JTR can be used with the aim of enhancing the system performance, since the total transmitted power can be increased leading to highly reliable communication. It is important to note that JTR puts some constraints on the maximum power to be used for signal transmission, otherwise nonlinearies may exist. These constraints fade in case of massive MIMO since the maximum power is very high compared to what is needed for current cellular networks.

## 5.6 Simulation Results

In fig. 5.3, we present the CDF of the PA output, assuming 16 antennas at the BS and 2 antennas at the UE. As shown in the figure, although the transmit power is increased to 2 and 4, the PAs are guaranteed to operate in the linear region for 98.8% and 100% respectively. However, increasing the transmit power to 8 may lead to a slight degradation as will be shown later. In fig. 5.4, we present the CDF of the PA output by fixing the transmit power to 2 while using a different number of antennas at the BS. As shown in the figure, increasing the number of antennas increases the

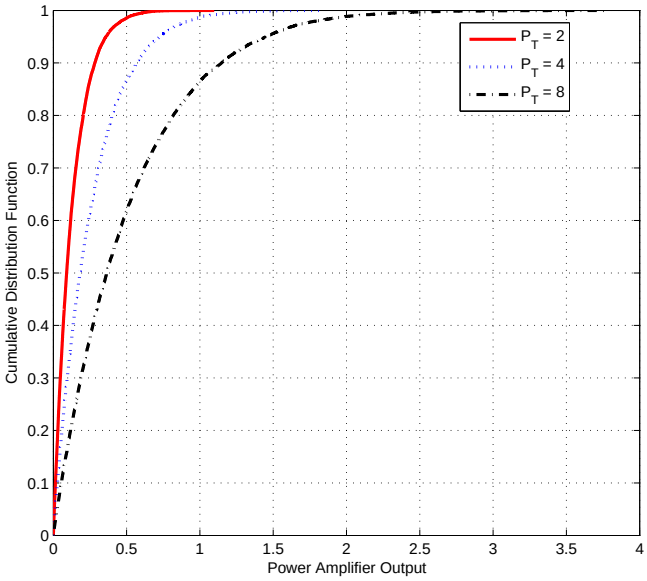



Figure 5.3: CDF of PA output for 16x2 MIMO system

probability of operating in the linear region. Hence, having 8 or 16 antennas at the BS will guarantee that PAs operate in the linear region for 99.2% and 100% respectively. However, reducing the BS antennas to 4 may lead to nonlinear operation.

Fig. 5.5 is the same as fig. 5.4 except that the transmit power is fixed to 4 instead of 2. As shown, since the transmit power is increased, then the probability of operating in the linear region will decrease. Thus, having 16 antennas at the BS will guarantee that PAs operate in the linear region for 98.8%. However, reducing the BS antennas to 8 or 4 may lead to nonlinear operation. Based on the previous discussion, it is important to see the behavior of massive MIMO BSs with a large number of antennas. In fig. 5.6, we show that having 100 antennas at the BS can allow for increasing the transmit power significantly while operating in the linear region. In that figure, using transmit power of 20, 30, and 40 will guarantee operating in the linear region with probability 100%, 97.3%, and 93.7% respectively.

To summarize, massive MIMO can benefit the most from the presented joint transmit and receive design. Since the number of BS antennas is relatively high, then the probability of operating below

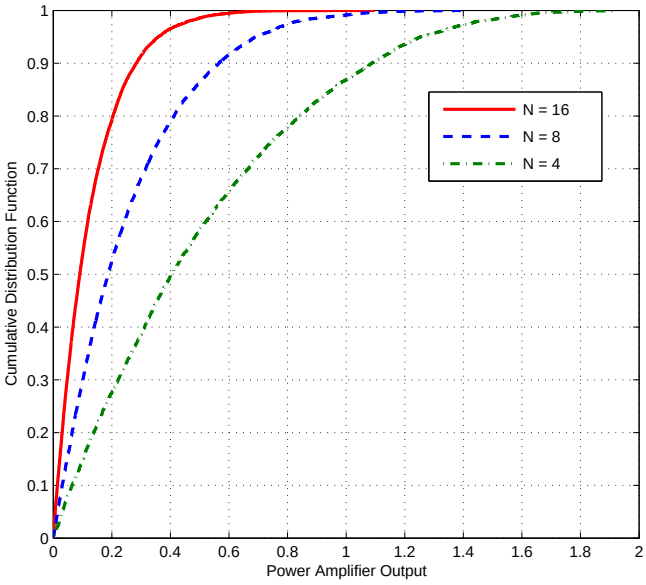



Figure 5.4: CDF of PA output with $P_T = 2$

the operating point is close to one. Along these lines, massive MIMO BS can be supported with low-cost PAs instead of the current LTE adopted high-cost PAs.

Figure 5.7 shows the output SNR with $P_T = 3$ assuming 2 antennas at the UE and 16 antennas at the BS for the proposed JTR method as compared to an ideal PA with an infinite linear region. As shown, the performance of the JTR matches exactly the performance of an ideal PA, which clearly indicates that there is no nonlinear distortion. As shown in Table 5.1, JTR in this scenario will lead to about 99.5% linear operation, thus the performance is only affected by very slight nonlinear distortion that can be neglected.

As mentioned earlier reducing the number of antennas at the BS will also require reducing the transmitted power, so fig. 5.8 is the same as fig 5.7 except that there is only 8 antennas at the BS and the transmitted power is reduced to 2. The same observation still holds, that nonlinear distortion using JTR is minimal at about 99.2% linear operation. It is worth mentioning that, increasing the transmit power beyond certain limit that depends on the number of BS antennas may introduce nonlinear distortion such as in fig. 5.9 where as shown that $P_T = 2$ provides

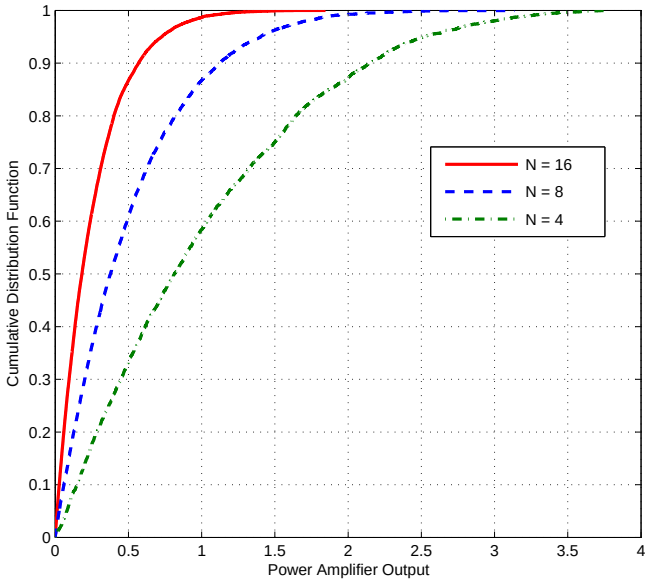



Figure 5.5: CDF of PA output for $P_T = 4$

very close performance to the ideal case. However, increasing the transmit power to 3 leads to an obvious gap between JTR and the ideal response.

## 5.7 Conclusions

We present joint transmitter and receiver filter design taking PAs behavior into account. Furthermore, we show that by using a high number of antennas at the BS such as in massive MIMO, the presented design can eliminate the effect of nonlinearity. The proposed model can be used either to introduce energy savings in the BSs by using low-cost PAs, or to achieve higher performance in terms of rate and/or reliability by relaxing the total transmit power constraints [103].

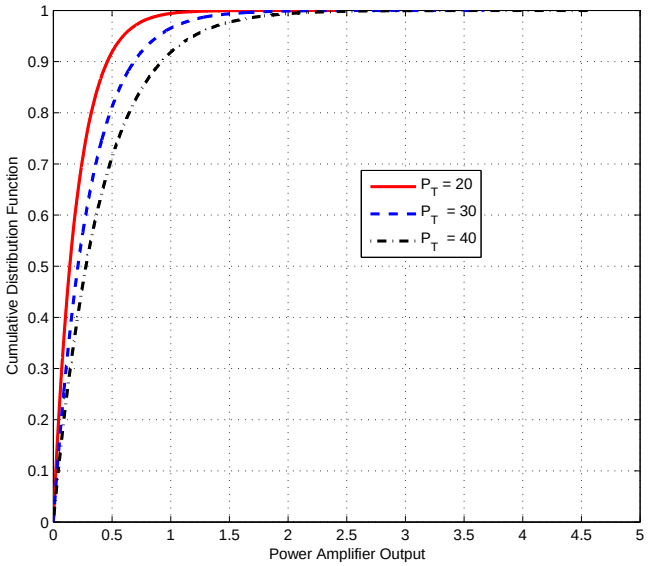



Figure 5.6: CDF of PA output for massive MIMO BS with 100 antennas

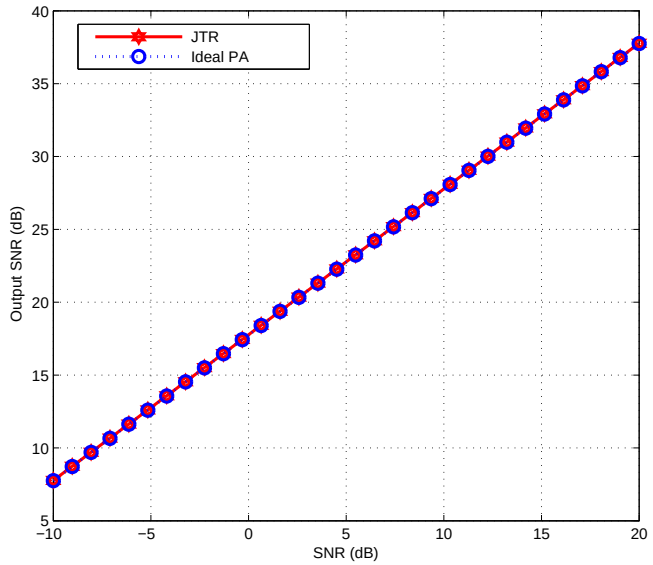



Figure 5.7: Output SNR for 16x2 system with $P_T = 3$

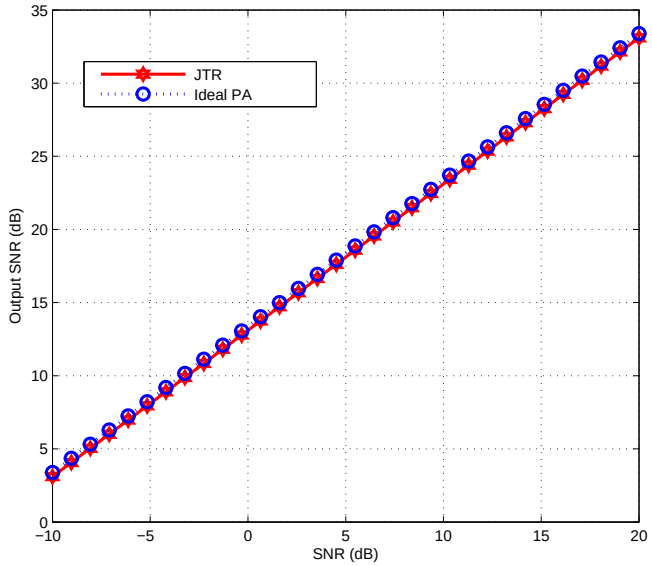



Figure 5.8: Output SNR for 8x2 system with $P_T = 2$

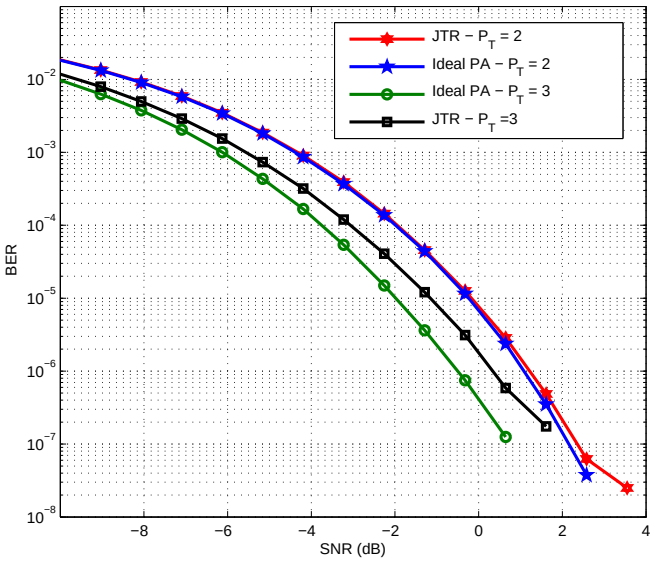



Figure 5.9: Average BER for 8x2 system

# Chapter 6

# Optimizing Energy Through Adaptive Bit Width Adjustment

## 6.1 Introduction

Over the last decade, the world has seen a sharp increase in data traffic that necessitates robust, low-power processing cores. However, mobile computing based on traditional architectures is approaching its limits in terms of scalability and power consumption. One means of achieving the desired performance increase is by increasing parallelism rather than depending on transistor feature reduction [112]. This approach also becomes limited if processing elements cannot consume data from memory at the desired processing rate, leading to a significantly degraded overall performance. To address that limitation, new computing paradigms started to emerge that focus more on the memory bottleneck problem. Theoretically, the most memory efficient paradigm is in-memory computation. This paradigm simply replaces the logic with memory structures, virtually eliminates the need for memory load/store operations during computation.

*Associative processors* (APs) are promising computational platforms for massively in-memory

parallel computing [113]. APs can be considered as a type of Single Instruction Multiple Data (SIMD) processors that combine the memory and processor in the same location, so that every row in the memory behaves as an individual processor. Since an operation can be performed on all memory words in parallel, the execution time of an operation does not depend on the vector size [114]. Many parallel systems are uniquely suited to this approach due to the vector based nature of their processing pipelines. This feature largely overcomes the memory-wall problem of traditional Von Neumann architectures since there is no inter-dependance between memory and processor. Associative processing is not a new topic and numerous architectures of APs originated in the 1970's and 1980's [113] [115]; however, in the past, the adoption of APs was limited due to the unmanageable power and area requirements. This reality is changing with the availability of new semiconductor technologies and materials that allow for extremely dense memory structures such as memristor [116] and STT-RAM [117], leading to a resurrection of this approach under the name of *Resistive Associative Processor* (RAP) [118].

Another computing paradigm that has become well-known in the recent years is *Approximate Computing*. In approximate computing, the goal is exploiting the error resiliency by relaxing correctness constraints to achieve the energy efficiency. In a system, approximate computation can be introduced at three different levels: design level, algorithm-architecture level, and logic-circuit level [119]. In the circuit level, the most common method is designing functionally approximate circuits that has lower components than its fully accurate counterpart. Other ways of hardware approximations are overscaling the circuit timing and/or voltage [120] and approximation in memory [121] [122]. At the architecture level, the significant components in the overall system is favored over insignificant ones. In the design level, the approximate computing can be realized by design tools that supports the approximate computing [123]. For example, a VLSI design software can include approximate versions of some arithmetic circuits and these circuits can be used in error resilient parts of the chip.

Even though RAP architectures promise very efficient parallel computing achievements, there are

serious problems of large power density and energy consumption in such architectures mainly due to high switching activity and costly memristor energy [118]. Unless these problems are addressed, it is likely that these architectures cannot be practical. On the other hand, applying approximate computing onto the existing computation systems does not fully eliminate the aforementioned problems of the traditional computing, even though it is a rising star in low energy computation. Fortunately, AP architectures inherently facilitates the approximate computing since all computations are performed on per bit basis. Regarding the problems of dark silicon era, combination of associative processing with approximate computing can be a promising approach for the future of computing especially for communication systems. To the best of our knowledge, no prior study has touched on the approximate in-memory computing.

In this study, we introduce the approximate in-memory computation by exploiting RAP in communication systems. The goal is to replace logic with memory structures, virtually eliminating the need for memory load/store operations during computation together bit dynamic approximate computing in algorithm-architecture level for both energy and performance efficiency. The suitability of RAP for approximate computing is demonstrated through the implementation of FFT used in wireless communication system. Results show that approximate in-memory computation in RAPs provides considerable energy saving by means of approximation in a reasonable level together with performance gain.

## 6.2 System Architecture

### 6.2.1 Associative Processor (AP)

The detailed architecture of the AP is shown in Figure 6.1. The processor comprises a content addressable memory (CAM), a controller, an instruction cache, an interconnection circuit, and registers. Inside the AP, a CAM stores the data on which operations are performed in parallel. The

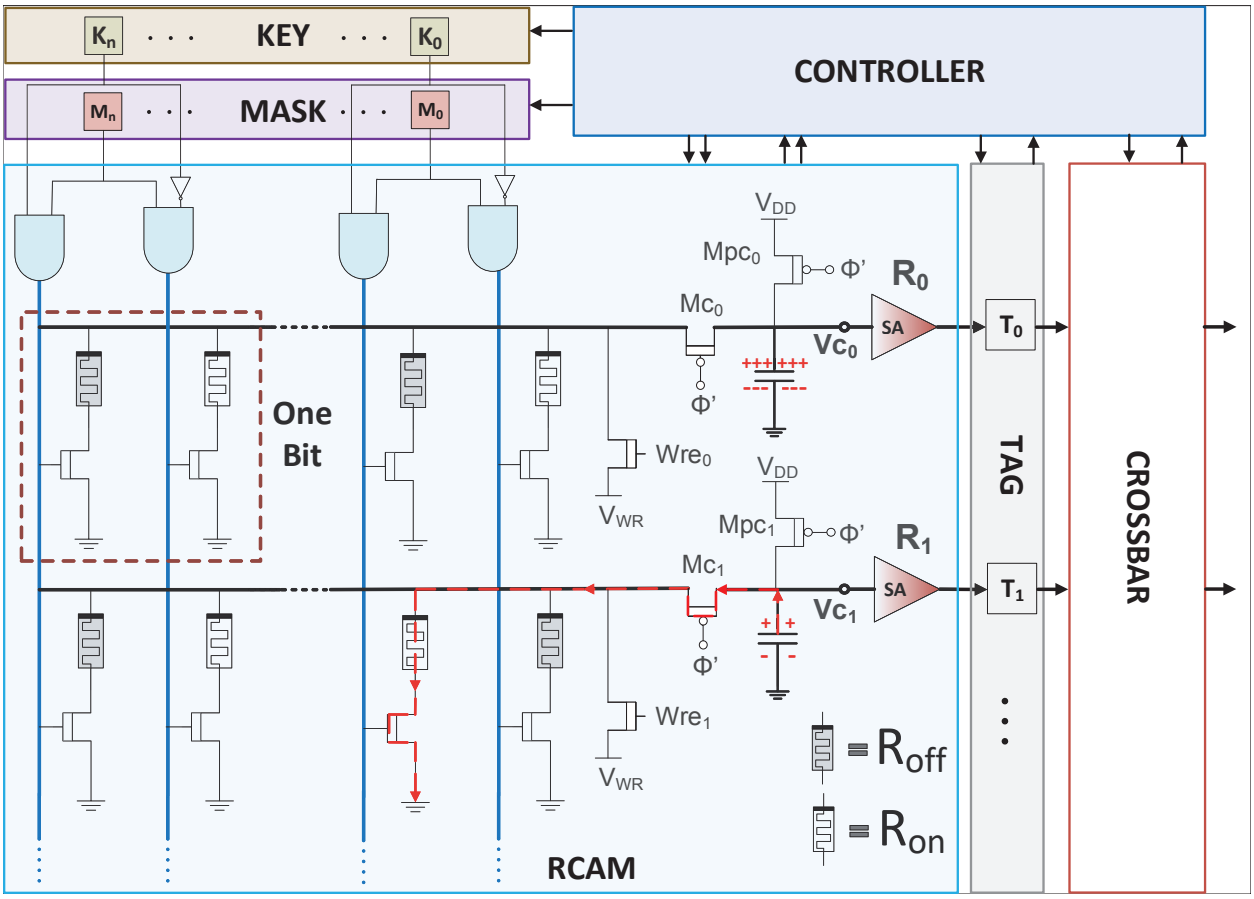



Figure 6.1: Architecture of an Resistive Associative Processor (RAP)

key register is used to present the value that is written to the CAM or compared against in the CAM. The mask register indicates which bits are activated during a comparison or a write. The rows matched by the compare operation are marked in the tag field where the rows tagged with logic-1 means that the corresponding CAM row has been matched with the given key and mask value. For example, if the key is set as 110 and the mask as 101, the tag bits of the corresponding rows whose first and third bits are logic-0 and logic-1 respectively becomes logic-1. The interconnection matrix (i.e. crossbar) is a basic circuit switched matrix that allows rows of the AP to communicate bitwise in parallel. The controller generates the required mask and key values for each corresponding instruction and manages the data interchange between the surrounding units (either CPU or another RCAM).

As a central part of the AP, the main requirement of a CAM array is to identify the row location

of matches against a search word. Such requirements can be achieved using various CAM cells. However, compact and energy efficient implementation is the key point that becomes feasible with the emerge of new semiconductor technologies. One of the most promising candidates for a CAM basic cell is described in [124], and is composed of two memristors and two transistors (2T2M). A memristor device is a nonlinear passive device that changes its state according to the net charge passing through its two terminals, and maintains its state after the electrical bias is removed. Binary data is stored in the memristor device is the form of ”High” and ”Low” resistances. The device can therefore work as a storage element and a switch at the same time. As pointed in Figure 6.1, the cells of our AP implementation based on the memristor. This type of CAM implementation is called *Resistive CAM* (RCAM) and correspondingly AP implementation is called *Resistive AP* (RAP).

In the figure, gray memristor corresponds to the memristor with high resistance state ($\mathrm{R_{off}}$) and white one corresponds to the one with low resistance state ($\mathrm{R_{on}}$). A search operation in RCAM is carried out in two sequential phases: pre-charge and evaluation. In the pre-charge phase, all the rows of the array which forms a parasitic capacitance are pre-charged concurrently. During the evaluation phase, a search word is applied to the columns, enabling one of the pass transistors in each CAM cell. A CAM cell should connect a path to the ground in case of a mismatch between the data it is holding and the data assigned to its column. The charges on a row capacitance leak from the mismatched cell, where the memristor and the series transistor are of low resistance creating a path to the ground. Since the data is stored in this ”2T2M” cell in a complimentary mode, the high resistance device will not leak charges to the ground even in case of mismatch, however its complement device will do so. For example, the first row in the figure shows an RCAM row in case of a match, where no low-resistance path to the ground is available. On the other hand, the second row leaks the charge since there is a path to ground through a memristor whose state is low, so causing mismatch. Writing to the RCAM in an RAP is performed using a one column at a time scheme. However, this is translated into two writing steps, since a complimentary data-column is electively made of two columns of the CAM array. The bits to write are loaded to the match lines

of the rows, with a key value of logic-1 to activate the column of interest. This eliminates the need for any modification to the column driving circuitry used for reading.

An operation on AP consists of consecutive $compare$ and $write$ phases. During the compare phase, the matched rows are selected and in write phase, the corresponding masked key values are written onto tagged CAM words. Depending on the desired arithmetic operation, the controller sets the mask and key values by referencing a lookup table (LUT). In the compare phase, the key and mask fields are set and compared with CAM content, while in the write phase, tagged rows are changed with the key. In other words, the truth table of the function is applied (in an ordered sequence) to the CAM to implement the required function. Utilizing consecutive compare and write cycles with a corresponding truth table, any function with corresponding truth table can be performed on RAPs. As an example, to perform an XOR operation on two input columns and then write the result on another column initialized as all 0s, the LUT of XOR operation ($\mathrm{R} = \mathrm{A}\&\mathrm{B}$) is applied to the RCAM where RCAM is searched for ”10” in the input columns (A and B) and ”1” is written to the result column (R) of the tagged rows, and then same operation can be done for ”01” as well, then the operation is completed with an R column stores the XOR of A and B. Similar to XOR, any function that can be performed on a sequential processor (i.e. including but not limited to addition, subtraction, multiplication, absolute value, 2’s & 1’s complement, logical operations, etc.) can also be performed on RAPs in parallel by utilizing consecutive compare and write cycles in their corresponding LUTs.

### 6.2.2 Fast Fourier Transform

Equation 6.1 shows the formula of Discrete Fourier Transform (DFT). This transform can be implemented faster by interleaving it into butterfly steps, which is known as FFT [125]. FFT consists of the butterfly operations in successive stages. Each stage includes a number of butterfly operations depending on input size. Figure 6.3a shows the simplest butterfly diagram consisting of two

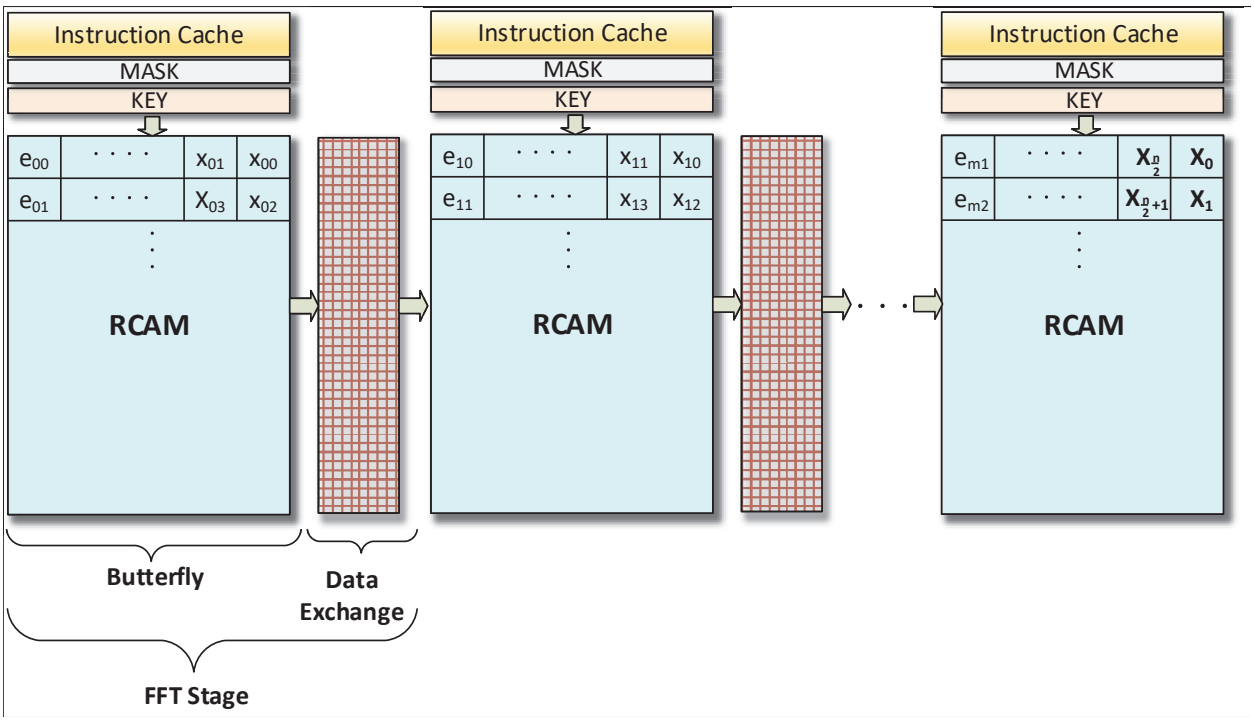



Figure 6.2: Implementation of FFT on pipelined RAPs

inputs, two outputs and one exponential coefficient (twiddle factor). Figure 6.3b shows an example 4-point, radix-2 FFT operation in two stages. As shown in the figure, after each stage, the partial outputs of previous stages are re-arranged as an input of the next butterfly stage. From RAP-based point of view, each row can be regarded as a different processor with their own registers to perform a butterfly operation, so two input and one exponential factor must be stored in the same row. After completion of a butterfly stage, the output of the previous stage is rearranged for the next stage.

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi k \frac{n}{N}} \tag{6.1}$$

In the RAP, all butterfly operations on a CAM are performed simultaneously, so the running time of one stage does not depend on the number of samples.

For FFT implementation on the RAP, the architecture described in Figure 6.2 is used. The architec-

ture consists of the pipelined RAPs. In the RAP, FFT operation consist of the consecutive butterfly and data exchange phases. Figure 6.3c shows the butterfly operation on a RCAM row step by step. The correspondence of each step is explained in the algorithm showed in Figure6.3d. In the algorithm, each operation is performed on complex numbers, that is, performed on real and imaginary parts separately. In the RAP, all butterfly operations on a RCAM are performed simultaneously, so the running time of one stage does not depend on the number of samples. After each stage, the partial outputs are directed to the corresponding places in the following RCAM by interconnection matrix. The interconnection matrix can be implemented as hardwired connection since communication pattern is know in advance. Combination of one butterfly and its data exchange phase (interconnection matrix) constitutes one FFT stage. In Figure 6.2, $e_{sr}$ stands for the twiddle factor of corresponding stage $s$ and row $r$, whereas $x_{si}$ corresponds to the input of a butterfly operation where $s$ is the stage number and $i$ is the input number. For example, $x_{00}$ corresponds to the first input of the first butterfly stage. For an n-point FFT operation, the overall system requires $log_2(n)$ APs and each AP requires $n/2$ rows. For example, the system requires 3 RAPs and 4-rows in each AP for 8-point FFT operation. The exponential coefficients ($e_{xy}$) are assumed to be placed to the CAM arrays before the operations. It is worth noting that order of $x_{0i}$ values is *reverse bit* order of the real input samples ($x_0, x_1, ..., x_n$).

In associative computing, an arithmetic operation can be started with any of the bits by disregarding its remaining rightmost bits and go through the most significant bits since all operations are performed as bitwise. For this reason, the associative computing provides a natural support for approximate computing. As an example, the figure 6.4a shows a fixed point representation of a real number in the RAP. If the last 4-bits of the number is trimmed, 0.2% percent accuracy of the number is lost (Figure 6.4b). However, this lost in precision provides both run time and energy savings in performing operations on this number. For the addition of two 16-bit fixed point numbers, this approximation provides 25% decrease in run time and about 25% decrease in energy consumption.

By carefully setting the degree of approximation seriously regarding the needs of specific applica-

tions, approximate in-memory computing can provide both considerable energy and performance improvements. In the following section, approximate in-memory computing in OFDM-based wireless communication systems is presented.

### 6.2.3 Error Analysis

The output of a butterfly in the $k^{th}$ stage can be given as follows:

$$\boldsymbol{a}_k = \boldsymbol{a}_{k-1} + \boldsymbol{e}_k \boldsymbol{b}_{k-1}, \tag{6.2}$$

$$\boldsymbol{b}_k = \boldsymbol{a}_{k-1} - \boldsymbol{e}_k \boldsymbol{b}_{k-1}$$

Where, $\boldsymbol{a}_k, \boldsymbol{b}_k$ are the two outputs of the $k^{th} stage$, and $\boldsymbol{e}_k$ is the coefficient of the $k^{th}$ stage. The error variance, $\sigma_k^2$ at the output of stage $k$ is given by the simple recursion

$$\sigma_k^2 = \gamma_k \sigma_{k-1}^2 + \beta_k, \tag{6.3}$$

where $\gamma_k$ is the propagation scaling factor of the stage due to inputs bit truncation , and $\beta_k$ is the stage specific error variance due to truncating bits. We need to take a closer look at each operation in the AP aiming at getting $\gamma_k$ and $\beta_k$. We will focus on a butterfly of one stage.

The first operation in the AP is the absolute operation, and this operation error will be taken into consideration along with the multiplication operation.

**Multiplication Operation**

In the multiplication operation, the variance of the error depends on the precision of the multiplication. Multiplying two inputs with $B_{in}$ bits, results in $B_{out} = 2B_{in}$. Thus, there will be an additional error due to the finite precision at the output. The multiplication operation is performed on complex numbers, with the real part of the output expressed using Euler's formula as follows:

$$\boldsymbol{y}_m = (\boldsymbol{b}_R + \boldsymbol{e}_{b_R})(\cos\theta_k + \boldsymbol{e}_{e_R}) + (\boldsymbol{b}_I + \boldsymbol{e}_{b_I})(\sin\theta_k + \boldsymbol{e}_{e_I}) + \boldsymbol{e}_{k_{tm}} \tag{6.4}$$

where, $\boldsymbol{b}_R$ and $\boldsymbol{b}_I$ are the real and imaginary parts of $|b|$ respectively, $\boldsymbol{e}_{b_R}$ and $\boldsymbol{e}_{b_I}$ are the errors due to the partial truncation of fractional bits from $\boldsymbol{b}_R$ and $b_I$ respectively, $\boldsymbol{e}_{e_R}$ and $\boldsymbol{e}_{e_I}$ are the errors due to truncating fractional bits from $\cos\theta_k$ and $\sin\theta_k$ respectively, and $\boldsymbol{e}_{k_{tm}}$ is the error due to truncating the output of the multiplier. Thus, the output of the multiplication can be expressed as:

$$\begin{aligned} \boldsymbol{y}_m =& \boldsymbol{b}_R\cos\theta_k + \boldsymbol{b}_I\sin\theta_k + \boldsymbol{e}_m \\ =& \widehat{\boldsymbol{y}}_m + \boldsymbol{e}_m \end{aligned} \tag{6.5}$$

where, $\boldsymbol{e}_m$ is the cumulative error at the output of the multiplier, and can be expressed as:

$$\boldsymbol{e}_m = \boldsymbol{e}_{b_R}\cos\theta_k + \boldsymbol{e}_{b_I}\sin\theta_k + \boldsymbol{e}_{e_R}(\boldsymbol{b}_R + \boldsymbol{e}_{b_R}) + \boldsymbol{e}_{e_I}(\boldsymbol{b}_I + \boldsymbol{e}_{b_I}) + \boldsymbol{e}_{k_{tm}} \tag{6.6}$$

The variance of the error at the output of this sub-stage is

$$\sigma^2_{k_m} = \sigma^2_{k-1} + 2 * \sigma^2_{k_e}(\boldsymbol{P}_{k-1} + \sigma^2_{k-1}) + \sigma^2_{k_{tm}} \tag{6.7}$$

where, $\sigma^2_{k_m}$ is the variance of the error at the output of the multiplier of the $k^{th}$ stage. $\sigma^2_{k-1}$ is the propagated error, which is the input to the $k^{th}$ stage and the output of stage $k-1$. $\sigma^2_{k_e}$ is the error in the FFT coefficient, $\boldsymbol{P}_{k-1}$ is the signal power of stage $k-1$, and finally, $\sigma^2_{k_{tm}}$ is the error due to bit truncation.

After the multiplication, there is an XOR operation followed by an absolute operation, both of these operations are error free. Finally, there is either an addition or subtraction operation. The details of the error analysis is discussed in the next subsection.

**Addition/Subtraction Operation**

Addition and subtraction operations have the same error analysis as multiplication. We focus on the addition operation. The variance of the error depends on the bit width of the operation. There are two cases, 1) The output has the same bit width as the input , 2) An extra bit of precision could be added at the output of the addition operation. We consider both cases. The output of the adder is as follows:

$$\boldsymbol{y}_a = \boldsymbol{a} + \boldsymbol{e}_a + \widehat{\boldsymbol{y}}_m + \boldsymbol{e}_m + \boldsymbol{e}_{k_{ta}} \tag{6.8}$$

where, $\boldsymbol{e}_a$ is the error in the input $\boldsymbol{a}$, and $\boldsymbol{e}_{k_{ta}}$ is the error due to truncation of the addition operation.

$$\boldsymbol{y}_a = \boldsymbol{a} + \widehat{\boldsymbol{y}}_m + \boldsymbol{e}_{ad} \tag{6.9}$$

where, $\boldsymbol{e}_{ad} = \boldsymbol{e}_a + \boldsymbol{e}_m + \boldsymbol{e}_{k_{ta}}$.

**Error Propagation**

Using a uniform random error model:

$$\sigma_k^2 = \frac{2^{-2(B_k-1)}}{12} \tag{6.10}$$

where $B_k$ is the bit width of the FFT coefficient of stage $k$. Similarly, we can get the other error variance of multiplier and adder in terms of the bit width used for the multiplier and the adder respectively. The variance of the error at the output of $k^{th}$ stage is

$$\sigma_k^2 = \sigma_{k-1}^2 + \sigma_{k_m}^2 + \sigma_{k_{ta}}^2 \tag{6.11}$$

$$\sigma_k^2 = 2\sigma_{k-1}^2 + 2 * \sigma_{k_e}^2 \boldsymbol{P}_{k-1} + 2 * \sigma_{k_e}^2 \sigma_{k-1}^2 + \sigma_{k_{tm}}^2 + \sigma_{k_{ta}}^2 \tag{6.12}$$

$$\sigma_k^2 = 2\sigma_{k-1}^2(1 + \sigma_{k_e}^2) + 2 * \sigma_{k_e}^2 \boldsymbol{P}_{k-1} + \sigma_{k_{tm}}^2 + \sigma_{k_{ta}}^2 \tag{6.13}$$

By using (6.13) and (6.3), we can get the following:

$$\gamma_k = 2(1 + \sigma_{k_e}^2) \tag{6.14}$$

$$\beta_k = \begin{cases} 2 * \sigma_{k_e}^2 \boldsymbol{P}_{k-1} + \sigma_{k_{tm}}^2 + \sigma_{k_{ta}}^2, & B_{out} \leq B_{in} \\ 2 * \sigma_{k_e}^2 \boldsymbol{P}_{k-1} + \sigma_{k_{tm}}^2, & B_{out} > B_{in} \end{cases} \tag{6.15}$$

where, $B_{out}$ and $B_{in}$ is the bit width at the output and the input of the addition operation respectively. The the signal-to-quantization noise ratio (SQNR) of stage $k$ is computed as follows:

$$\rho_k = \frac{\boldsymbol{P}_k}{\sigma_k^2} = \frac{\boldsymbol{P}_k}{\gamma_k \sigma_{k-1}^2 + \beta_k} \tag{6.16}$$

At the beginning of each stage, the SQNR at the output of the stage can be estimated as follows:

$$\rho_k = \frac{\boldsymbol{P}_{k-1}}{(1 + \sigma_{k_e}^2)\sigma_{k-1}^2 + \sigma_{k_e}^2 \boldsymbol{P}_{k-1} + 0.5(\sigma_{k_{tm}}^2 + \sigma_{k_{ta}}^2)} \tag{6.17}$$

Note that:

$$\rho_k \geq \rho_{k+1} \tag{6.18}$$

Simulation results confirm that the numerical and theoretical models match as shown in figure 6.5.

## 6.3 Adaptive bit width adjustment

The objective of the adaptive bit width adjustment is to utilize the knowledge of the channel state information to minimize the energy consumption by adjusting the bit width of the AP stages such that a certain performance is achieved. The error variance of truncating bits either at the input or at subsequent operations will be as follows:

$$\sigma^2(b_i) = \frac{2^{-2(b_i-1)}}{12} \tag{6.19}$$

We can derive the error variance of the real part at the output of the FFT as the following:

$$\sigma_n^2(b_i) = \gamma^n \sigma^2(b_i) + \sum_{j=0}^{n-1} \gamma^j \beta_{n-j+2} \tag{6.20}$$

where,

$$n = \log_2 N_{FFT}, \tag{6.21}$$

$$\gamma = 2(1 + \sigma^2(b_i)),$$

$$\beta_k = 2 * \sigma^2(b_i)(\boldsymbol{P}_{k-1} + 1).$$

Now, by central limit theorem, the real and imaginary parts of the error distribution at the output of last stage can be modeled as Gaussian, such that:

$$e_{real} \sim \mathcal{N}(0, \sigma_n^2(b_i)), e_{imag} \sim \mathcal{N}(0, \sigma_n^2(b_i)) \tag{6.22}$$

The received signal can be expressed as:

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}_{tot} \tag{6.23}$$

where, $\boldsymbol{H}$ is the wireless channel, $\boldsymbol{x}$ is the transmitted signal, and $\boldsymbol{n}_{tot}$ is the total noise.

$$\boldsymbol{n}_{tot} = \boldsymbol{w} + \boldsymbol{e}, \tag{6.24}$$

where, $\boldsymbol{w}$ is the zero mean AWGN noise, and $\boldsymbol{e}$ is the noise due to errors in the AP stages.

$$\sigma_{tot}^2(b_i) = \sigma_w^2 + \sigma_e^2(b_i)$$

where $\sigma_e^2 = 2 * \sigma_n^2(b_i)$. The BER can be expressed in terms of the SNR for M QAM as the following:

$$BER(\zeta, b_i) = \frac{2\left(1 - \frac{1}{\sqrt{M}}\right)}{\log_2 M} \text{erfc}\left(\sqrt{\frac{1.5\zeta}{\sigma_{tot}^2(b_i)(M-1)}}\right) \tag{6.25}$$

where, the received SNR has an exponential distribution as derived in [126].

$$P(\zeta) = \frac{\boldsymbol{e}^{\frac{-\zeta}{\bar{\zeta}}}}{\bar{\zeta}}, \zeta \geq 0 \tag{6.26}$$

The Rayleigh channel is modeled as finite state markov channel as derived in [127]. In that model, the range of the channel SNR is portioned into $J$ non-overlapping intervals denoted by $[\Gamma_j \Gamma_{j+1})$, where $j = 0, 1, ..., J-1$. Thus, the channel is said to be in state $\boldsymbol{H}_j$ if the received SNR is in the

interval $[\Gamma_j \Gamma_{j+1})$. The average BER given the channel state $\boldsymbol{H}_j$ is expressed as follows:

$$BER(\boldsymbol{H}_j, b_i) = \frac{2\left(1-\frac{1}{\sqrt{M}}\right)}{\delta_j * \log_2 M} * \int_{\Gamma_j}^{\Gamma_{j+1}} \text{erfc}\left(\sqrt{\frac{1.5\zeta}{\sigma_{tot}^2(b_i)(M-1)}}\right) \frac{\boldsymbol{e}^{\frac{-\zeta}{\bar{\zeta}}}}{\bar{\zeta}} d\zeta \tag{6.27}$$

where, $\delta_j$ is the steady state probability of being in state $j$, such that:

$$\delta_j = e^{\frac{-\Gamma_j}{\bar{\zeta}}} - e^{\frac{-\Gamma_{j+1}}{\bar{\zeta}}} \tag{6.28}$$

The solution to (6.27) can be derived as in [128] as the following:

$$BER(\boldsymbol{H}_j, b_i) = \frac{\boldsymbol{g}_j - \boldsymbol{g}_{j+1}}{e^{\frac{-\Gamma_j}{\bar{\zeta}}} - e^{\frac{-\Gamma_{j+1}}{\bar{\zeta}}}} \tag{6.29}$$

where,

$$\boldsymbol{g}_j = f(M) * e^{\frac{-\Gamma_j}{\bar{\zeta}}} * \text{erfc}\left(\sqrt{\frac{1.5 * \Gamma_j}{\sigma_{tot}^2(b_i)(M-1)}}\right) - f(M) * \sqrt{\frac{1.5\bar{} * \zeta}{(M-1)\sigma_{tot}^2(b_i) + 1.5 * \bar{\zeta}}} * \text{erfc}\left(\sqrt{\frac{1.5 * \bar{\zeta} * \Gamma_j}{(M-1)\sigma_{tot}^2(b_i) + 1.5 * \bar{\zeta}}}\right) \tag{6.30}$$

where,

$$f(M) = \frac{2\left(1-\frac{1}{\sqrt{M}}\right)}{\log_2 M} \tag{6.31}$$

The problem of adjusting the bit width is modeled as Markov decision process (MDP), in which $\boldsymbol{S} = \{\boldsymbol{s}_1, ..., \boldsymbol{s}_N\}$ are all the possible states. Also, $B = \{b_1, ..., b_M\}$ are all the possible actions, an action means a certain bit width. Since there is two noise contributions (channel and bit width), a

state is defined by these two elements, thus each state will have an SNR slack. The objective is to find the optimal action (policy) $\pi^*(\boldsymbol{s})$ that minimizes the utility:

$$\pi^*(\boldsymbol{s}) = \underset{b}{\operatorname{argmin}}\, U(\boldsymbol{s}, b) \tag{6.32}$$

where, $U(\boldsymbol{s}, b)$ denotes the utility, which is the value of taking action $b$ in state $s$.

$$U(\boldsymbol{s}, b) = C(\boldsymbol{s}, b) + \sum_{\hat{\boldsymbol{s}} \epsilon S} T(\boldsymbol{s}, b, \hat{\boldsymbol{s}}) V(\hat{\boldsymbol{s}}) \tag{6.33}$$

where, $V(\boldsymbol{s})$ is the value of the state:

$$V(\boldsymbol{s}) = U(\boldsymbol{s}, \pi^*(\boldsymbol{s})), \tag{6.34}$$

and $C(\boldsymbol{s}, b)$ is the cost, which is defined using Karush Kuhn Tucker (KKT) conditions as follows:

$$C(\boldsymbol{s}, b) = \boldsymbol{E}(b) + \mu * \left( BER(\boldsymbol{s}, b) - \frac{BER_{opt}(s)}{\lambda} \right) \tag{6.35}$$

where, $\boldsymbol{E}(b)$ is the energy consumption if bit width $b$ is used. The transition model $T(\boldsymbol{s}, b, \hat{\boldsymbol{s}})$ specifies the probability of transition from state $\boldsymbol{s}$ to state $\hat{\boldsymbol{s}}$ on taking bit width $b$. The channel transition probability is derived in [127]. $\mu$ is the KKT multiplier, $BER_{opt}(s)$ is the optimal BER using maximum bitwidth, and $\lambda$ is the performance threshold.

The adaptive bit width algorithm selects the appropriate bit width of the next transmitted packet. Thus, the objective is to save energy consumption by lowering the bit width of the AP stages.

The solution to MDP can be found using the value iteration algorithm as shown in Table 6.1. And the adaptive bit width algorithm can be summarized as in Table 6.2. The adaptive bit width algorithm can be summarized as in Fig. 6.6.

Table 6.1: Value Iteration Algorithm

For each state $\boldsymbol{s}$, initialize the value of the state to zero: $V_0(\boldsymbol{s}) = 0$
For $l = 1 : L$
  For each state $\boldsymbol{s}$
    For each bit width $b$
      Compute the value of taking action $b$ in state $\boldsymbol{s}$
    $U_l(\boldsymbol{s}, b) = C(\boldsymbol{s}, b) + \sum_{\hat{\boldsymbol{s}}} T(\boldsymbol{s}, b, \hat{\boldsymbol{s}}) V_{l-1}(\hat{\boldsymbol{s}})$
    end
   Compute the optimal policy for state $\boldsymbol{s}$: $\pi_l^*(\boldsymbol{s}) = \text{argmin}_b\, U_l(\boldsymbol{s}, b)$
   Update the value of each state $V_l(\boldsymbol{s}) = U_l(\boldsymbol{s}, \pi_l^*(\boldsymbol{s}))$
  end
end
Return $\pi_L^*(\boldsymbol{s})$

Table 6.2: Adaptive Bit Width Algorithm

| |
|---|
| Step 1: Initial Learning Phase:<br>Define and solve the MDP using Table 6.1 |
| Step 2: Populating LUT Phase:<br>Store the optimal bit width for each state in a LUT |
| Step 3: Runtime Phase:<br>Identify the current state and find the optimal action |

## 6.4 Performance Evaluation

### 6.4.1 Communication system model

In this paper, we use LTE communication system setup. In LTE, the unit in time is a 1 msec unit consisting of 14 OFDM symbols. Thus, the processing time requirement for FFT = $\frac{1}{14}$ msec = 71 $\mu$sec. The wireless channel is assumed to be Rayleigh fading and is modeled as a finite state Markov channel (FSMC). We aim at tracking the received SNR and utilizing this information to

Table 6.3: Simulation Parameters

| Parameter | Value |
|---|---|
| Frequency Band | 2 GHz |
| Transmission Bandwidth | 10 MHz |
| Number of subcarriers | 600 |
| Subcarrier spacing | 15 kHz |
| FFT size | 1024 |
| Channel Model | Rayleigh |

reduce the energy consumption of the system by reducing the bit width when the errors due to channel noise dominate. The simulation parameters are summarized in Table 6.3.

### 6.4.2 CAM/AP

For the evaluation of approximate in-memory computing in RAPs, a Matlab-based RAP simulator is implemented along with a Spice-based cycle-accurate circuit simulator. The simulator is capable of performing pipelined RAP simulations on different features. All truth tables for required arithmetic operations to perform FFT (addition, subtraction, absolute value, 2's complement, multiplication) are generated inside the simulator. Outputs of the functions were compared with their corresponding Matlab functions to verify their correctness.

For circuit implementation, the platform allows plugging in any memristor model for any two terminal resistive devices and we adopt the device model presented in [129]. The existing FFT block is replaced with its RAP-based counterpart in an OFDM based MIMO system. This 1024-point FFT unit consists of 10 pipelined RAPs that can efficiently simulate realistic RAPs. For the numbers, fixed point representation is used with separate integer and fractional parts. The design is implemented to allow the largest precision during the operation (i.e. 12 bits for integer and fractional parts). The reported operating frequency of RAP is $500\,\mathrm{MHz}$ .

In terms of precision, based on system level simulation, data values are stored as 4 bits for the

integer part and 4-8 bits (variable) for fractional part. For the complex coefficients (eg. twiddle factors in FFT), we used 2 bit for the integer part and 6-10 bits (variable) for the fractional part. Therefore, the RAP is capable of processing up to 12-bit precision for data and complex coefficient values. The operating frequency of RAP is $500\,\mathrm{MHz}$ .

Figures 6.7 show the change of energy reduction and throughput per 1K-FFT with respect to the number of fractional bits. As the precision decreases, the energy saving and the throughput increase. On the other hand, increasing precision results in more run-time to complete one FFT, and correspondingly a higher energy consumption. As shown in the throughput results, the system throughput is about 80 MS/s which means that the system can perform 1K-FFT within 12.92 $\mu$s, which is well within the 71 $\mu$s for FFT required by LTE frames.

### 6.4.3 FFT

In this subsection, we justify the benefits and compare the performance of the proposed adaptive bit width algorithm. We assume Rayleigh channel model. The fractional bit width is selected adaptively among five levels shown in Table 6.4. Although bit width of 14 bits is widely used, we performed simulations to select a tighter bit width aiming at having a meaningful comparison. In this paper, we use a metric ”normalized performance”, defined as the ratio between the minimum BER (assuming highest bit width) and the achieved BER. So, a normalized performance close to unity is desired. As shown in Fig. 6.8, 12 bits provides the same performance as 14 bits. Thus we select 12 bits as our performance reference.

An important observation is that at low SNRs, decreasing the bit width leads to very slight performance degradation. However, at high SNRs, reducing the bit width results in large degradation in the performance. This observation clearly indicates the need of SNR dependant adaptive bit width adjustment. Figure 6.9 shows the performance degradation versus energy consumption for different SNRs. A good solution would be closer to the left vertical line (consumes low energy)

and the top horizontal line (closer to no performance degradation), as shown, the proposed adaptive bit width algorithm results in low energy consumption while maintaining slight performance degradation.

Figure 6.10 shows the performance of the proposed algorithm utilizing different bit width levels. As shown, the algorithm adjusts the bit width adaptively so that it select the min bit width that satisfies the performance constraint, which will reduce the energy consumption.

Table 6.4: Energy consumption for different bit width levels

| Bit width | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|
| Energy [%] | 55 | 65 | 76 | 87 | 100 |

The comparison of RAP-based implementation of FFT processors with traditional ones is shown in Table 6.5. The table includes two version of RAP implementation of 1K FFT which represent the fixed 12-bit precision and adaptive precision between 8-12 bits respectively. The table shows the normalized area, power efficiency numbers and a figure of merit in terms of throughput over power density . For a fair comparison, the respective numbers are normalized according to the equations 6.36, 6.37, and 6.38 where N corresponds to the FFT size. As shown in the table, adaptive bit-width methodology provides a considerable gain in both energy and run time, thus positively influences the efficiency of RAP based FFT implementation within CMOS-based counterparts in which bit-scale computing is not possible architecturally.

Table 6.5: Comparison with other ASIC implementations of FFT

| | **Technology** (nm) | **Size** (points) | **Word Width** (bits) | **Area** ($mm^2$) | **Throughput** (MS/s) | **Normalized Area Efficiency** ($GS/s/mm^2$) | **Normalized Power Efficiency** ($GS/s/W$) | **Normalized FOM** ($GS/s/W/mm^2$) |
|---|---|---|---|---|---|---|---|---|
| RAP | 16 & RRAM | 1K | 12 | 0.002 | 79.28 | 39.84 | 0.94 | 470.41 |
| RAP | 16 & RRAM | 1K | Adaptive (12-8) | 0.002 | 128.80 | 62.21 | 1.52 | 761.74 |
| [130] | 65 | 1K | 16 | 8.29 | 240 | 0.16 | 252.54 | 670.34 |
| [131] | 90 | 256 | 10 | 5.1 | 2,400 | 2.21 | 652.25 | 674.43 |
| [132] | 45 | 2K | 32 | 0.973 | 0.222 | 0.0002 | 0.44 | 20.80 |

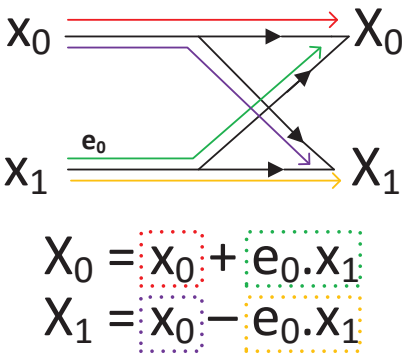
$$\underset{Area}{Normalized} = \frac{Area}{\left(\frac{Tech}{16\ nm}\right)^2 \cdot \left(\frac{Wordlength}{12}\right) \cdot \left(\frac{N \cdot log_2 N}{10240}\right)} \tag{6.36}$$

$$\underset{Power}{Normalized} = \frac{Power}{\left(\frac{Tech}{16\ nm}\right) \cdot \left(\frac{Wordlength}{12}\right) \cdot \left(\frac{V_{DD}}{0.7\ V}\right)^2 \cdot \left(\frac{N \cdot log_2 N}{10240}\right)} \tag{6.37}$$

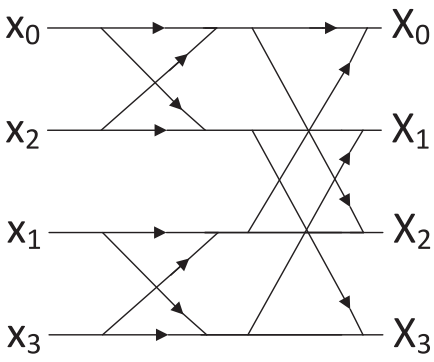
$$\underset{Throughput}{Normalized} = \frac{Throughput}{\left(\frac{16\ nm}{Tech}\right) \cdot \left(\frac{12}{Wordlength}\right) \cdot \left(\frac{N \cdot log_2 N}{10240}\right)} \tag{6.38}$$

## 6.5 Conclusion

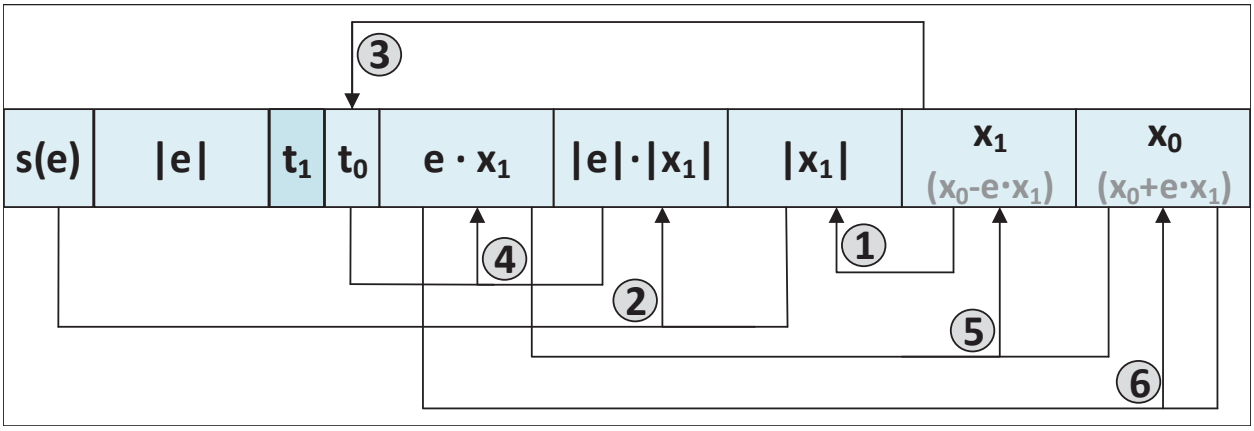
In this study, the approximate in-memory computation concept is introduced by exploiting the resistive associative processors in communication systems. The goal is to replace logic with memory structures together bit dynamic approximate computing for both energy and performance efficiency. The suitability of resistive associate processors for approximate computing is demonstrated. As an application, a novel mathematical model that characterizes system performance of FFT under fractional bits truncation has been derived. Based on that model, an adaptive bit width adjustment algorithm has been proposed. Simulation results show that by using the proposed adaptive bit width algorithm, we can achieve up to 45% of energy savings with very slight performance degradation [133].




(a) Simple Butterfly




(b) 4-point FFT

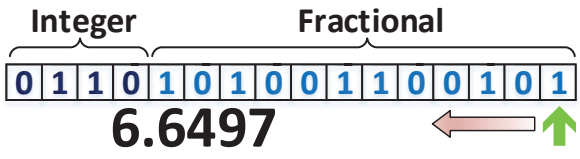



(c) Butterfly on a RAP row

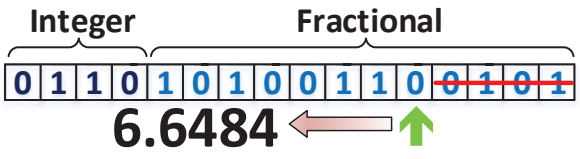
1: **procedure** BUTTERFLY$(x_0, x_1)$
2: $|x_1| \leftarrow Abs(x_1)$
3: $|e| \cdot |x_1| \leftarrow Multiply(|e|, |x_1|)$
4: $t1 \leftarrow XOR(s(e), s(x_1))$ ▷ s(x) = sign of x
5: $e \cdot b \leftarrow Abs(|e| \cdot |x_1|, t1)$
6: $X_1 \leftarrow SubtractOOP(e \cdot x_1, x_0)$
7: $X_0 \leftarrow AddIP(e \cdot x_1, x_0)$
8: **end procedure**

(d) Butterfly algorithm for RAP

Figure 6.3: FFT and Butterfly Operation in the AP




(a) Actual value




(b) Approximate value

Figure 6.4: Approximation in the RAP where some least signification bits are trimmed

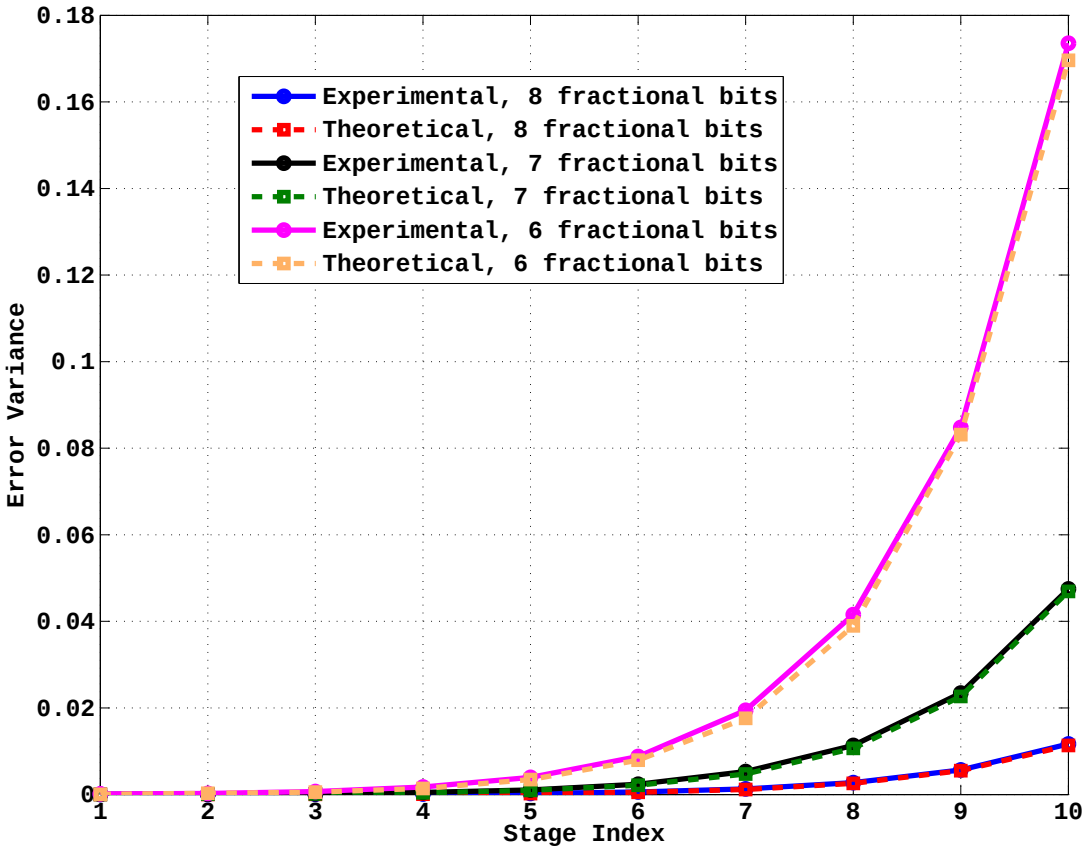



Figure 6.5: Error variance of different fractional bits

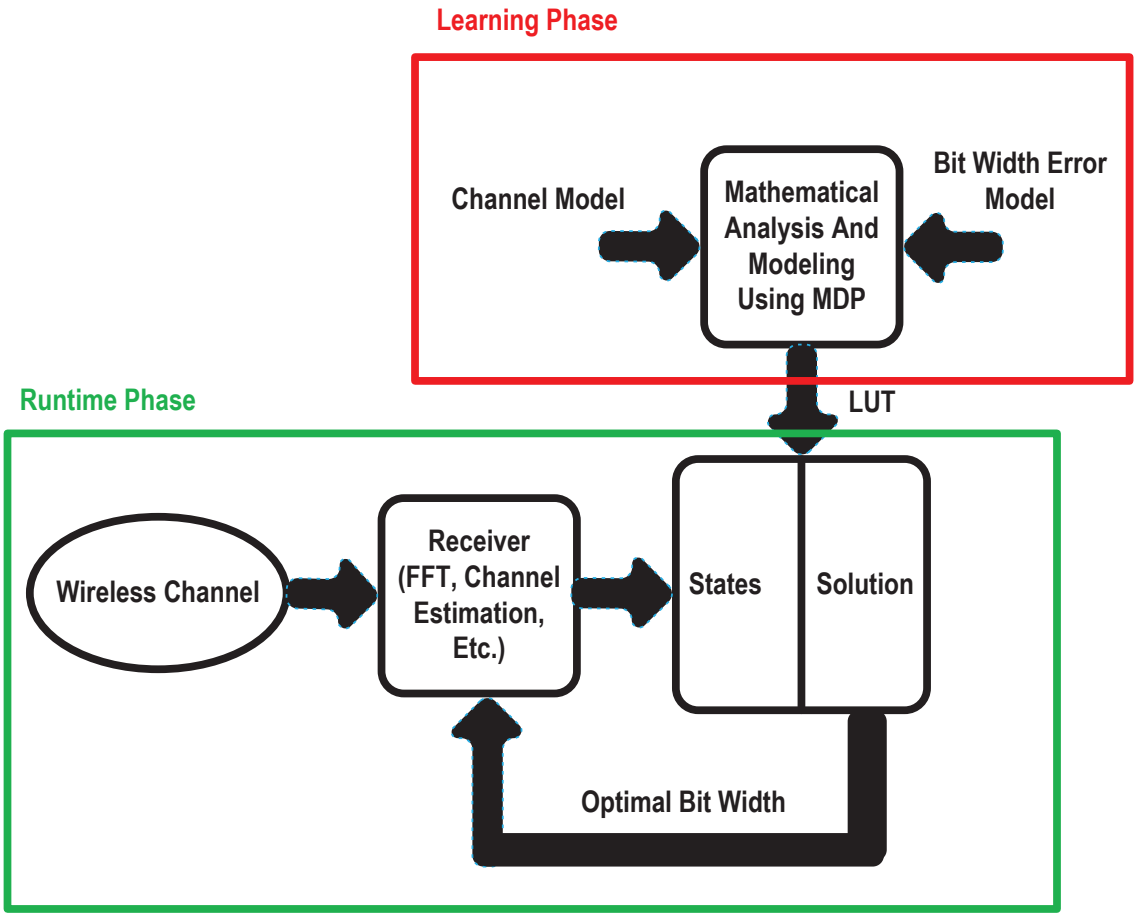



Figure 6.6: Adaptive Bit width Algorithm

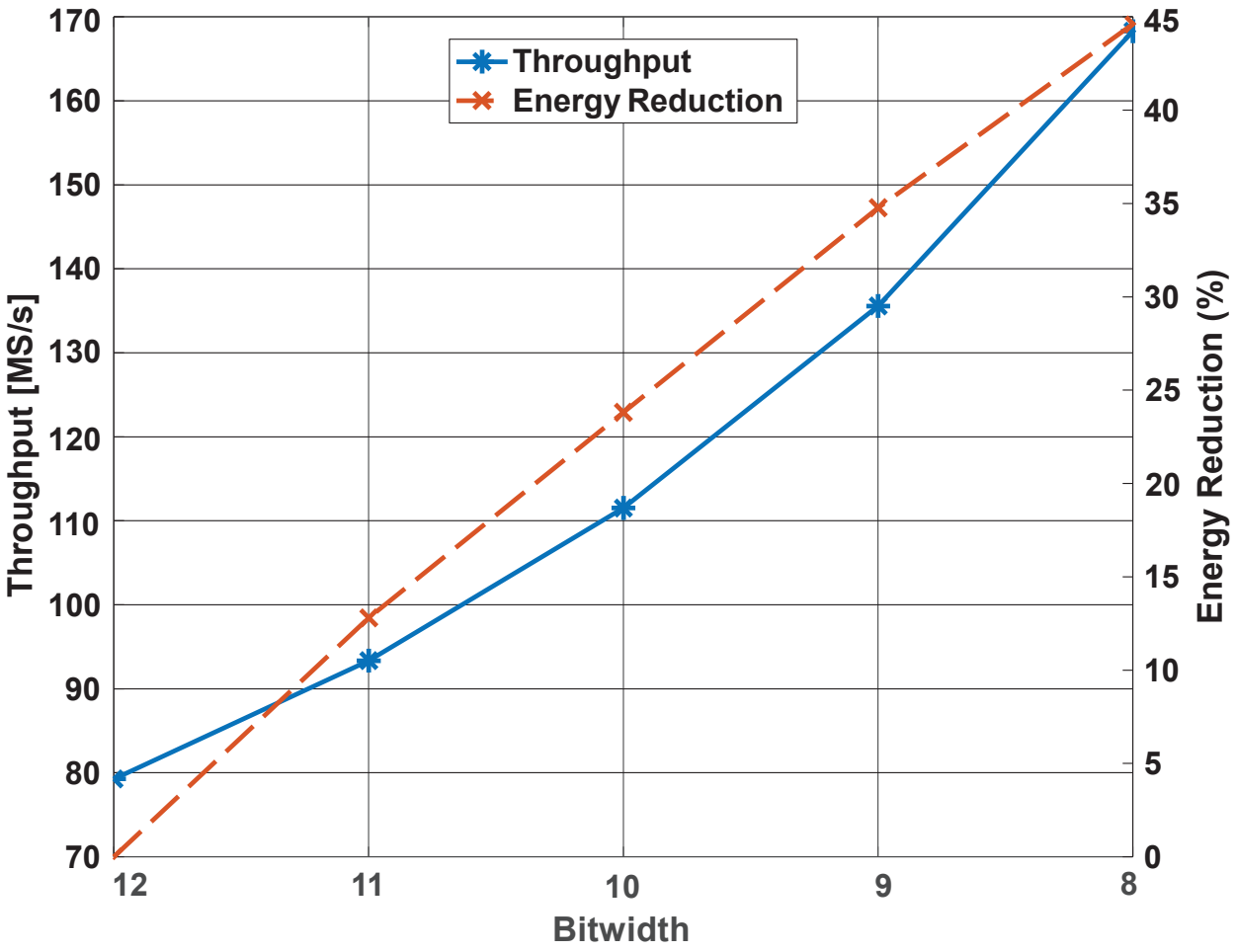



Figure 6.7: Bit width vs run-time and normalized energy

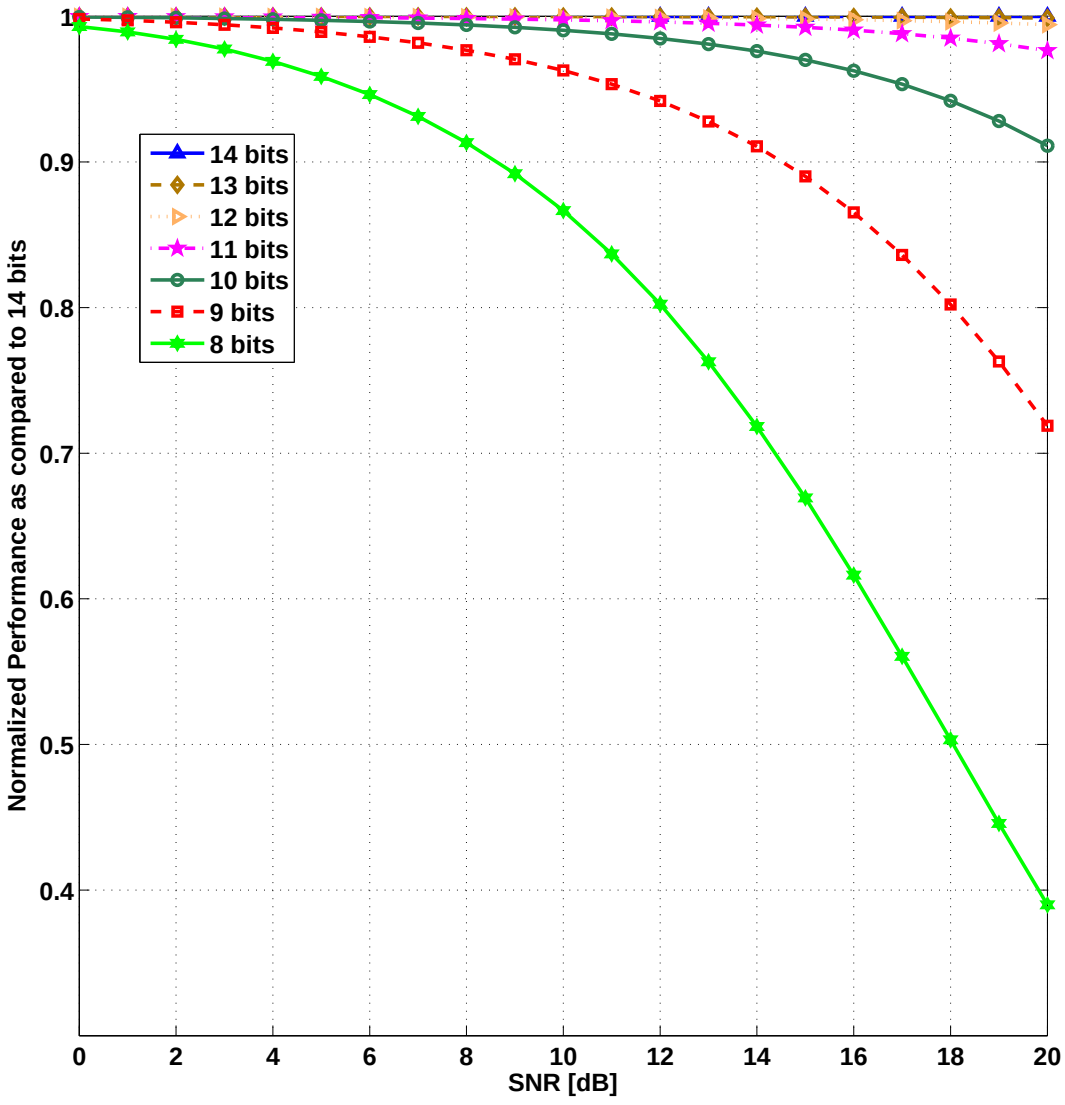



Figure 6.8: Normalized performance for different bit width levels

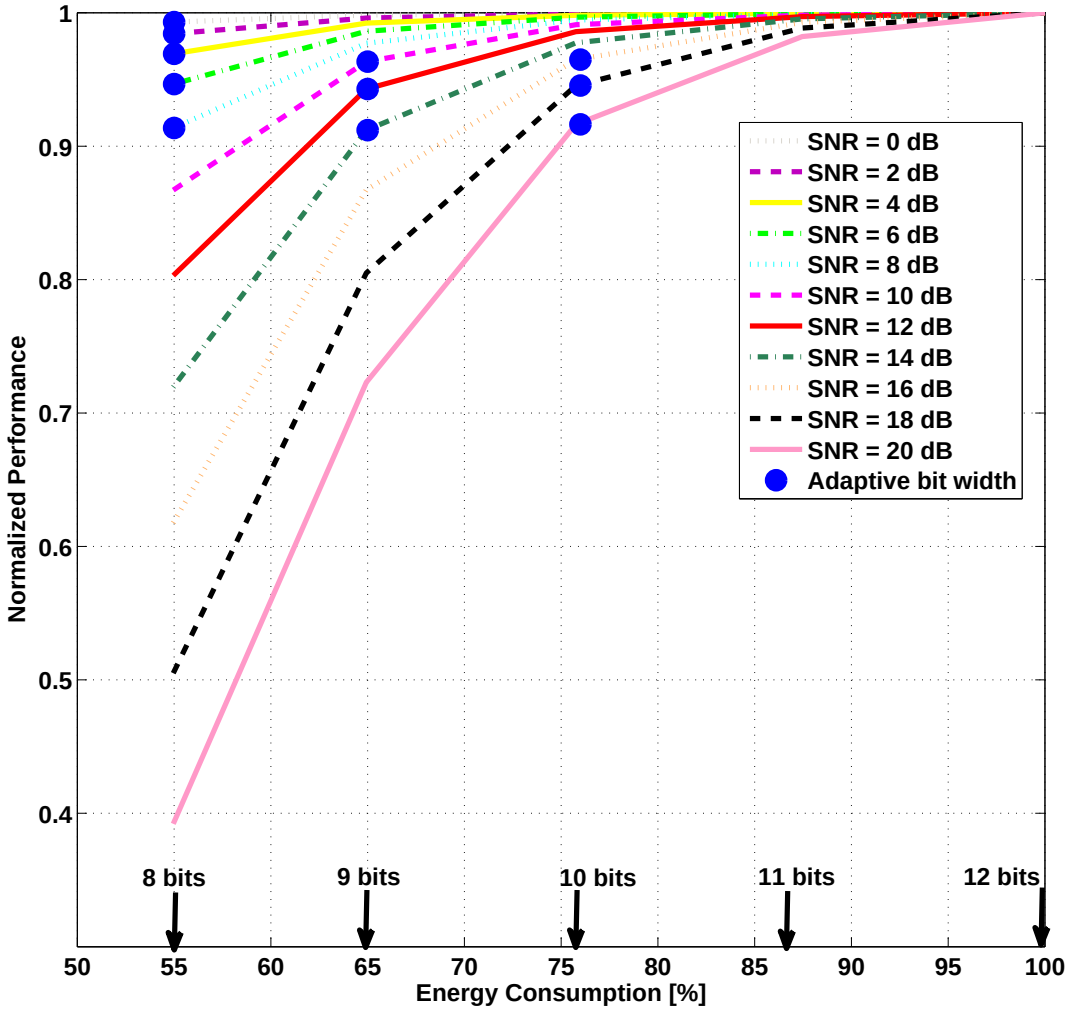



Figure 6.9: Normalized performance versus energy consumption for different SNRs

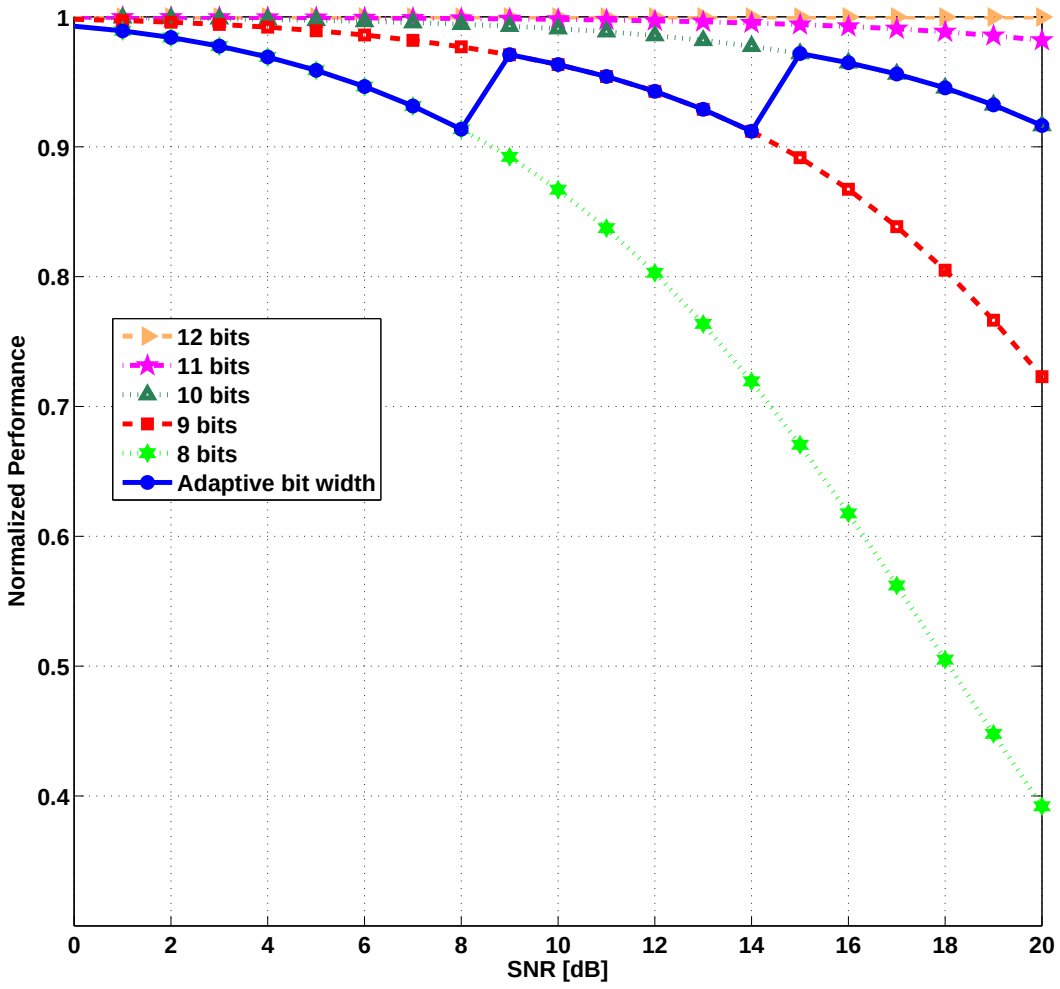



Figure 6.10: Normalized performance as compared to different bit width levels

# Bibliography

[1] A. Goldsmith, “Wireless Communications,” *Cambridge University Press*, 2005.

[2] Y. Choi, C. Kim, and S. Bahk, “Flexible design of frequency reuse factor in OFDMA cellular networks,” *in Proc. IEEE ICC 06*, vol. 4, pp. 1784-1788, 2006.

[3] A. Tolli, H. Pennanen, and P. Komulainen, “On the value of coherent and coordinated multi-cell transmission,” *in Proc. IEEE ICC 09*, 2009.

[4] W. V. Etten, “Maximum Likelihood Receiver for Multiple Channel Transmission Systems,” *IEEE Trans. Commun.*, vol. 24, no. 2, pp. 276-283, 1976.

[5] H. Dai, A. Molisch, and H. Poor, “Downlink Capacity of Interference-Limited MIMO Systems with Joint Detection,” *IEEE Trans. Wireless Commun.*, vol. 3, no. 2, pp. 442-453, 2004.

[6] U. Madhow and M. L. Honig, “MMSE Interference Suppression for Direct-Sequence Spread-Spectrum CDMA,” *IEEE Trans. Commun.*, vol. 42, no. 12, pp. 3178-3188, 1994.

[7] P. Patel and J. Holtzman, “Analysis of a Simple Successive Interference Cancellation Scheme in a DS/CDMA System,” *IEEE JSAC*, vol. 12, no. 5, pp. 796-807, 1994.

[8] S. Tomasin, A. Gorokhov, H. Yang, and J. P. Linnartz, “Iterative Interference Cancellation and Channel Estimation for Mobile OFDM,” *IEEE Trans. Wireless Commun.*, vol. 4, no. 1, pp. 238-245, 2005.

[9] M. Sawahashi et al., “Expirements on Pilot-Assisted Coherence Multistage Interference Canceller for DSCDMA Mobile Radio,” *IEEE JSAC*, vol. 20, no. 2, pp. 433-449, 2002.

[10] T. Henk, “Understanding Probability: Chance Rules in Everyday Life,” *Cambridge University Press*, 2004.

[11] J. T. Louhi, “Energy efciency of modern cellular base stations,” *International Telecommunications Energy Conference (INTELEC)*, pp. 475-476, 2007.

[12] 3GPP Technical Specification 36.321, “Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation (Release 8),” *www.3gpp.org*.

[13] L. M. Correia, et. al, “Challenges and enabling technologies for energy aware mobile radio networks,” *IEEE Communications Magazine*, vol. 48, no. 11, pp. 66-72, 2010.

[14] E. Costa, M. Midrio, and S. Pupolin, “Impact of amplifier nonlinearities on OFDM transmission system performance,” *IEEE Commun. Lett.*, vol. 3, no. 6, pp. 37-39, 1999.

[15] M. Jain, J. I. Choi, T. Kim, D. Bharadia, K. Srinivasan, S. Seth, P. Levis, S. Katti, and P. Sinha, “Practical, Real-time, Full Duplex Wireless,” *in Proceeding of the ACM Mobicom*, Sept. 2011.

[16] B. Radunovic, D. Gunawardena, P. Key, A. P. N. Singh, V. Balan, and G. Dejean, “Rethinking indoor wireless Mesh Design: Low power, low frequency, full duplex,” *Wireless Mesh Networks (WIMESH 2010), 2010 Fifth IEEE Workshop on*, pp.1-6, June 2010.

[17] M.E. Knox, “Single antenna full duplex communications using a common carrier,” *Wireless and Microwave Technology Conference (WAMICON), 2012 IEEE 13th Annual*, 15-17 April 2012.

[18] A. Balatsoukas-Stimming, P. Belanovic, K. Alexandris, and A. Burg, “On Self-interference Suppression Methods for Low-complexity Full-duplex MIMO,” *Asilomar Conference on Signals, Systems and Computers*, November 2013.

[19] E. Ahmed and A. M. Eltawil, “Full-Duplex Systems Using Multi-Reconfigurable Antennas, ” *Wireless Communications, IEEE Transactions on*, Early access, July 2015.

[20] O. Somekh, B. M. Zaidel, and S. Shamai (Shitz), “Sum rate characterization of joint multiple cell-site processing,” *IEEE Transactions on Information Theory*, vol. 53, no. 12,pp. 44734497, 2007.

[21] M. Karakayali, G. Foschini, and R. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Wireless Commun.*, vol. 13, no. 4, pp. 56-61, 2006.

[22] O. Somekh, O. Simeone, Y. Bar-Ness, and A. M. Haimovich, "Distributed multi-cell zero-forcing beamforming in cellular downlink channels," *in Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM 06)*, pp. 16, 2006.

[23] H. Dahrouj, W. Yu, "Coordinated beamforming for the multicell multi-antenna wireless system," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 5, pp. 1748-1759, 2010.

[24] 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, pp. 714725, 2008.

[25] R. Zhang, and S. Cui, "Cooperative interference management with MISO beamforming," *Signal Processing, IEEE Transactions*, vol. 58, no. 10, pp. 5450-5458, 2010.

[26] S. Jing, D. Tse, J. Soriaga, J. Hou, J. Smee, and R. Padovani, "Multicell downlink capacity with coordinated processing," *EURASIP J. Wireless Commun. and Networking*, vol. 2008, no. 5, 2008.

[27] R. Mudumbai, et al. "Distributed transmit beamforming: challenges and recent progress," *Communications Magazine, IEEE*, vol. 47, no. 2, pp. 102-110, 2009.

[28] C. Suh, M. Ho, and D. Tse, "Downlink interference alignment," *IEEE Trans. Commun.*, vol. 59, no. 9, pp. 2616-2626, 2011.

[29] G. Andrews,W. Choi, and R. Heath Jr, "Overcoming interference in spatial multiplexing MIMO cellular networks," *IEEE Wireless Communications Magazine*, vol. 14, no. 6, pp. 95-104, Dec. 2007.

[30] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using the V-BLAST space-time communication architecture," *Electronics Lett.*, vol. 35, pp. 14-15, 1999.

[31] M. Damen, H. El Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2389-2402, 2003.

[32] B. Hassibi and H. Vikalo, "On the sphere-decoding algorithm I. Expected complexity," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp.2806-2818, 2005.

[33] B. Hassibi and H. Vikalo, "On the expected complexity of sphere decoding," *in Proc. 35th Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp.10511055, 2001.

[34] L. Brunel, "Multiuser detection techniques using maximum likelihood sphere decoding in multicarrier cdma systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 3, pp. 949957, 2004.

[35] W. Zhao and G. B. Giannakis, "Sphere decoding algorithms with improved radius search," *IEEE Trans. Commun.*, vol. 53, no. 7, pp. 11041109, 2005.

[36] Z. Safar, W. Su, and K. R. Liu, "Fast sphere decoding of space-frequency block codes via nearest neighbor signal point search," *in Proc. 5th European Wireless Conference (EW 04)*, vol. 1, 2004.

[37] Z. Guo, and P. Nilsson, "Algorithm and implementation of the K-best sphere decoding for MIMO detection," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 491-503, 2006.

[38] J. Winters, "Optimum combining in digital mobile radio with cochannel interference," *IEEE Journal on Selected Areas in Communications*, vol. 2, no. 4, pp. 528539, 1984.

[39] 3GPP TR 36.829 V11.1.0, "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Enhanced performance requirement for LTE User Equipment (UE),", 2013. [Online]. Available:http://www.3gpp.org.

[40] A. Jones, T. Wilkinson, and S. Barton, "Block Coding Scheme for Reduction of Peak to Mean Envelope Power Ratio of Multicarrier Transmission Schemes," *Electronics Letters*, vol. 30, pp. 2098-2099, 1994.

[41] C. Tellambura, "Use of M-sequence for OFDM Peak-to-Average Power Ratio Reduction," *Electronics Letters*, vol. 33, pp. 1300-1301, 1997.

[42] X. Li, and L. Cimini, "Effects of Clipping and Filtering on the Performance of OFDM," *IEEE Vehicular Technology Conference*, vol. 3, pp. 1634-1638, 1997.

[43] J. Armstrong, "Peak to Average Power Reduction for OFDM by Repeated Clipping and Frequency Domain Filtering," *Electronics Letters*, vol. 38, pp. 246-247, 2002.

[44] T. Riihonen, and R. Wichman "Analog and digital self-interference cancellation in full-duplex MIMO-OFDM transceivers with limited resolution in A/D conversion," *In Proceedings of Signals, Systems and Computers (ASILOMAR)*, pp. 45-49, 2012.

[45] M. Duarte, and A. Sabharwal, "Full-duplex wireless communications using off-the-shelf radios: Feasibility and first results," *In Proceedings of Signals, Systems and Computers (ASILOMAR)*, pp. 1558-1562, 2010.

[46] T. Riihonen, S. Werner, and R. Wichman, "Mitigation of loopback self-interference in full-duplex MIMO relays," *Signal Processing, IEEE Transactions on*, pp. 5983-5993, 2011.

[47] M. Jain, J. Choi, T. Kim , D. Bharadia, S. Seth, K. Srinivasan, ... and P. Sinha "Practical, real-time, full duplex wireless," *In Proceedings of the 17th annual international conference on Mobile computing and networking (ACM)*, pp. 301-312, 2011.

[48] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-Driven Characterization of Full-Duplex Wireless Systems," *Wireless Communications, IEEE Transactions on*, vol.11, no.12, pp.4296,4307, December 2012.

[49] E. Ahmed, A. M. Eltawil, and A. Sabharwal, "Self-Interference Cancellation with Nonlinear Distortion Suppression for Full-Duplex Systems," *Asilomar Conference on Signals, Systems and Computers*, November 2013.

[50] E. Ahmed, A. M. Eltawil, and A. Sabharwal, "Self-Interference Cancellation with Phase Noise Induced ICI Suppression for Full-Duplex Systems," *Global Telecommunications Conference (GLOBECOM 2013)*, December 2013.

[51] E. Ahmed, A. M. Eltawil, and A. Sabharwal, "Rate Gain Region and Design Tradeoffs for Full-Duplex Wireless Communications, " *Wireless Communications, IEEE Transactions on* , vol.12, no.7, pp.3556,3565, July 2013.

[52] E. Ahmed, and A. M. Eltawil, "On Phase Noise Suppression in Full-Duplex Systems," *Wireless Communications, IEEE Transactions on*, vol.14, no.3, pp.1237,1251, March 2015.

[53] A. Sahai, G. Patel, C. Dick, and A. Sabharwal, "On the Impact of Phase Noise on Active Cancelation in Wireless Full-Duplex," *Vehicular Technology, IEEE Transactions on*, vol.62, no.9, pp.4494,4510, Nov. 2013.

[54] D.W. Bliss, T.M. Hancock, and P. Schniter, "Hardware phenomenological effects on cochannel full-duplex MIMO relay performance," *Signals, Systems and Computers (ASILOMAR), 2012 Conference Record of the Forty Sixth Asilomar Conference on*, pp.34,39, 4-7 Nov. 2012.

[55] J. I. Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti, "Achieving single channel, full duplex wireless communication," *in MobiCom*, 2010.

[56] E. Everett, M. Duarte, C. Dick, and A. Sabharwal, "Empowering full-duplex wireless communication by exploiting directional diversity," *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on*, pp.2002-2006, Nov. 2011.

[57] E. Ahmed, A. M. Eltawil, and A. Sabharwal, "Simultaneous transmit and sense for cognitive radios using full-duplex: A first study," *Antennas and Propagation Society International Symposium (APSURSI), 2012 IEEE* , pp.1-2, July 2012.

[58] E. Everett, A. Sahai, and A. Sabharwal, "Passive Self-Interference Suppression for Full-Duplex Infrastructure Nodes," *Wireless Communications, IEEE Transactions on*, vol.13, no.2, pp.680,694, February 2014.

[59] M. Duarte, A. Sabharwal, V. Aggarwal, R. Jana, K. Ramakrishnan, C. Rice, and N. Shankaranayanan, "Design and Characterization of a Full-Duplex Multiantenna System for WiFi Networks," *Vehicular Technology, IEEE Transactions on*, vol.63, no.3, pp.1160,1177, March 2014.

[60] J. Choi, M. Jain, K. Srinivasan , P. Levis, and S. Katti "Achieving single channel, full duplex wireless communication," *In Proceedings of the 16th annual international conference on Mobile computing and networking (ACM)*, pp. 1-12, 2010.

[61] D. Bharadia, E. McMilin, and S. Katti “Full duplex radios,” *In ACM SIGCOMM Computer Communication*, vol. 43, no. 4, pp. 375-386, 2013.

[62] S. Goyal, P. Liu, O. Gurbuz, E. Erkip and S. Panwar “A distributed mac protocol for full duplex radio,” *In Proceedings of Signals, Systems and Computers (ASILOMAR)*, pp. 788-792, 2013.

[63] N. Singh, D. Gunawardena, A. Proutiere, B. Radunovic, H. V. Balan, and P. Key “Efficient and Fair MAC for Wireless Networks with Selfinterference Cancellation,” *In Proceedings of IEEE WiOpt*,2011.

[64] K. Tamaki, H. Ari, Y. Sugiyama, M. Bandaiy, S. Saruwatari, and T. Watanabe “Full Duplex Media Access Control for Wireless Multi-Hop Networks,” *In Proceedings of IEEE VTC*,2013.

[65] J. Y. Kim, O. Mashayekhi, H. Qu, M. Kazandjieva, and P. Levis “Janus: A Novel MAC Protocol for Full Duplex Radio,” *Stanford University, Technical Report*,2013.

[66] W. Zhou, K. Srinivasan and P. Sinha “RCTC: Rapid Concurrent Transmission Coordination in Full Duplex Wireless Networks,” *In Proceedings of IEEE ICNP*,2014.

[67] E. Hossain, L. B. Le, and D. Niyato “ Radio Resource Management in MultiTier Cellular Wireless Networks,” *New York, NY, USA: Wiley*, 2013.

[68] L. B. Le et al “Enabling 5G mobile wireless technologies,”*EURASIP J. Wireless Commun. Netw.*, vol. 2015, no. 218,2015.

[69] D. Ramirez and B. Aazhang “Optimal routing and power allocation for wireless networks with imperfect full-duplex nodes,”*IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 46924704, 2013.

[70] W. Choi, H. Lim, and A. Sabharwal “Power-controlled medium access control protocol for full-duplex WiFi networks,”*IEEE Trans. Wireless Communications*, vol. 14, no. 7, pp. 36013613, 2015.

[71] A. Sahai, G. Patel, and A. Sabharwal “Pushing the limits of full-duplex: Design and real-time implementation,” *arXiv preprint arXiv:1107.0607*, 2011.

[72] S. Goyal, P. Liu, S. Hua, and S. Panwar “Analyzing a full-duplex cellular system,” *In Proceedings of 44th Annual Conference on Information Sciences and Systems (CISS)*,2013.

[73] E. Telatar, “Capacity of Multiantenna Gaussian Channels.” *European transactions on telecommunications*, vol. 10, no. 6,pp. 585-595, 1999.

[74] R. Blum, “MIMO capacity with interference,” *IEEE J. Select. Areas Commun.*, vol. 21, pp. 793-801, 2003.

[75] K. Gomadam and S. Jafar, “The effect of noise correlation in amplify-and-forward relay networks,” *IEEE Transactions on Information Theory*, vol. 55, no. 2, pp. 731-745, 2009.

[76] S. Krusevac,R. Kennedy, and P. Rapajic, “Effect of signal and noise mutual coupling on MIMO channel capacity,” *Wireless Personal Communications*, vol. 40, no. 3, pp. 317-328, 2007.

[77] K. Mardia, J. Kent, and J. Bibby, “Multivariate Analysis,” San Diego, CA: Academic, 1979.

[78] M. Sadek, A. Tarighat, and A. Sayed, “A Leakage-Based Precoding Scheme for Downlink Multi-User MIMO Channels,” *IEEE Trans. on wireless Communications*, vol. 6, no. 5, pp. 1711-1721, May. 2007.

[79] R. A. Abdelaal, A. S. Behbahani, and A. M. Eltawil, “On optimizing the performance of interference-limited cellular systems,” *IEEE Wireless Telecommunications Symposium (WTS)*, pp. 1-5, 2014.

[80] G. Golub and C. Loan, “Matrix Computations, 3rd ed. Baltimore,” The Johns Hopkins Univ. Press, 1996.

[81] S. Ahmadi, “LTE-advanced: A practical systems approach to understanding 3GPP LTE releases 10 and 11 radio access technologies,”*2nd ed. Academic Press*, 2013.

[82] S. Sesia, I. Toufik, and M. Baker, “LTE: The UMTS Long Term Evolution, from theory to practice,”*2nd ed. United Kingdom: Wiley*, 2011.

[83] R. Simeon, “Method and apparatus for canceling cross-correlation noise due to strong serving cell signals,”*InterDigital Technology Corporation*, US patent 7,460,625, 2008.

[84] 3GPP Technical Specification 36.321, “Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation,” *www.3gpp.org*.

[85] R. A. Abdelaal, A. S. Behbahani, and A. M. Eltawil, “On optimizing the performance of interference-limited cellular systems,” *IEEE Wireless Telecommunications Symposium (WTS)*, pp. 1-5, 2014.

[86] R. A. Abdelaal, A. S. Behbahani, and A. M. Eltawil, “Advanced Base Station Precoding and User Receiver Designs for LTE-Advanced Networks,” *IEEE International Conference on Computing, Networking and Communications (ICNC)*, pp. 1-5, 2015.

[87] K. Gomadam, V. Cadambe, and S. Jafar, “A distributed numerical approach to interference alignment and applications to wireless interference networks,” *IEEE Trans. Inf. Theory*, vol.57, no. 6, pp. 3309-3322, 2011.

[88] E. A. Jorswieck, E. G. Larsson, and D. Danev, “Complete characterization of the Pareto boundary for the MISO interference channel,”*IEEE Transactions on Signal Processing*, vol. 56, pp. 52925296, 2008.

[89] E. Aryafar, N. Anand, T. Salonidis, and E. Knightly, “Design and experimental evaluation of multi-user beamforming in wireless LANs,”*In Proceedings of ACM*, pp. 197-208, 2010.

[90] A. Wiesel, Y.C. Eldar, and S. Shamai, “Zero-forcing precoding and generalized inverses,”*IEEE Transactions on Signal Processing*, vol. 56, no.9, pp. 44094418, 2008.

[91] T. Yoo and A. Goldsmith, “On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming,”*IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 528541, 2006.

[92] D. Lee, H. Seo, B. Clerckx, E. Hardouin, D. Mazzarese, S. Nagata, and K. Sayana, “Coordinated multipoint transmission and reception in LTE-advanced: deployment scenarios and operational challenges,” *Communications Magazine, IEEE*, vol. 50, no. 2, pp.148-155, 2012.

[93] K. Brueninghaus, et al. “Link performance models for system level simulations of broadband radio access systems,” *IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications PIMRC*, vol. 4, 2005.

[94] 3GPP, “LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures,” TS 136 213 V10.11.0 , January 2014.

[95] D. A. Wassie, G. Berardinelli, F. M.L. Tavares, O. Tonelli, and P. Mogensen. “Experimental evaluation of interference rejection combining for 5G cells,” *IEEE WCNC*, pp. 652-657, 2015.

[96] F. Penna, S. Stanczak, Z. Ren, and P. Fertl, “MMSE interference estimation in LTE networks,” *IEEE ICC* pp. 45484552, 2014.

[97] S. Mosleh, L. Liu, C. Zhang, Proportional-Fair Resource Allocation for Coordinated Multi-Point (CoMP) Transmission in LTE-Advanced, *IEEE Transactions on Wireless Communications*, 2016.

[98] Rana A. Abdelaal, Alireza S. Behbahani, and Ahmed M. Eltawil, “Practical Framework for Downlink MU-MIMO for LTE Systems,” *IEEE wireless communications letters*, 2017.

[99] 802.11 WG,“IEEE 802.11ac, Part 11: Wireless LAN Medium Access Control (MAC)and Physical Layer (PHY) Specifications: Enhancements for Very High Throughput for Operation in Bands Below 6 GHz,”*IEEE std.*, 2013.

[100] R. Stacey, “Specification Framework for TGax,”*IEEE P802.11 Wireless LANs*, 2016.

[101] S. C. Cripps, “RF Power Amplifiers for Wireless Communications,” *Artech House, Norwood, MA, USA*, 1999.

[102] C. Rapp, “Effects of HPA nonlinearity on a 4-DPSK/OFDM signal for a digital sound broadcasting system,” *in Proc. European Conf. on Satellite Commun.*, vol. 1, pp. 179-184, 1991.

[103] R. A. Abdelaal, A. S. Behbahani, and A. M. Eltawil, “On the Performance of Massive MIMO Cellular Systems With Power Amplifiers,” *IEEE Wireless Telecommunications Symposium (WTS)*, pp. 1-5, 2014.

[104] S. Merlin, “TGax Simulation Scenarios,” *11-14/0621 (Qualcomm)* , 2014.

[105] C. Shepard, H. Yu, N. Anand, E. Li, T. Marzetta, R. Yang, and L. Zhong, “Argos: Practical Many-Antenna Base Stations,”*In Proceedings of ACM MobiCom*, 2012.

[106] Q. Yang, X. Li, H. Yao, J. Fang, K. Tan, W. Hu, J. Zhang, and Y. Zhang, “Bigstation: Enabling scalable real-time signal processing in large mu-mimo systems,”*In Proceedings of SIGCOMM*, 2013.

[107] Y. Du, E. Aryafar, P. Cui, J. Camp, and M. Chiang, “Samu: Design and implementation of selectivity-aware mu-mimo for wideband wifi,”*In Sensing, Communication, and Networking (SECON), 2015 12th Annual IEEE International Conference on*, pp. 229-237, 2015.

[108] IEEE 802.11-16/0024r1, ”Proposed TGax draft specification” *IEEE 802.11 Wireless LANs*, 2016.

[109] IEEE 802.11-15/0132r16, ”Specification Framework Document” *IEEE P802.11 Wireless LANs*, 2015.

[110] H. Kim and Y. Han, “A proportional fair scheduling for multicarrier transmission systems,”*In IEEE Commun. Lett.*, pp. 210212, 2005.

[111] Rana A. Abdelaal and Ahmed M. Eltawil, “Scheduling and Power Adaptation for Wireless Local Area Networks With Full-Duplex Capability,” *submitted to IEEE transactions of wireless communications*, 2017.

[112] S. Borkar, Exascale computing a fact or a fiction?, *in Proceedings of the 2013 IEEE 27th International Symposium on Parallel and Distributed Processing, ser. IPDPS 13. Washington, DC, USA: IEEE Computer Society*, 2013 [Online].

[113] C. C. Foster, ”Content Addressable Parallel Processors,” *New York, NY, USA: John Wiley & Sons, Inc.*, 1976.

[114] J. L. Potter, ”Associative Computing: A Programming Paradigm for Massively Parallel Computers,” *Perseus Publishing*, 1991.

[115] I. D. Scherson and S. Ilgen, A reconfigurable fully parallel associative processor, *J. Parallel Distrib. Comput.*,vol. 6, no. 1, pp. 6989, Feb. 1989.

[116] D. B. Strukov, G. S. Snider, D. R. Stewart, and R. S. Williams, The missing memristor found, *Nature*, vol. 453, no. 7191, pp. 8083, May 2008.

[117] D. Ralph and M. Stiles, Spin transfer torques, *Journal of Magnetism and Magnetic Materials*, vol. 320, no. 7, pp. 1190 1216, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0304885307010116

[118] L. Yavits et al., Resistive associative processor, *Computer Architecture Letters*, vol. PP, no. 99, pp. 11, 2014.

[119] K. Roy, Approximate computing for energy-efficient error-resilient multimedia systems, *in Design and Diagnostics of Electronic Circuits Systems (DDECS), 2013 IEEE 16th International Symposium on*, April 2013, pp. 56.

[120] D. Mohapatra, V. Chippa, A. Raghunathan, and K. Roy, Design of voltage-scalable meta-functions for approximate computing, *in Design, Automation Test in Europe Conference Exhibition (DATE), 2011*, March 2011, pp. 16.

[121] G. Karakonstantis, D. Mohapatra, and K. Roy, Logic and memory design based on unequal error protection for voltage-scalable, robust and adaptive dsp systems, *J. Signal Process. Syst.*, vol. 68, no. 3, pp. 415431, Sep. 2012. [Online]. Available: http://dx.doi.org/10.1007/s11265-011-0631-9

[122] I. J. Chang, D. Mohapatra, and K. Roy, A priority-based 6t/8t hybrid sram architecture for aggressive voltage scaling in video applications, *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 21, no. 2, pp. 101112, Feb 2011.

[123] S. Venkataramani, A. Sabne, V. Kozhikkottu, K. Roy, and A. Raghunathan, Salsa: Systematic logic synthesis of approximate circuits, *in Design Automation Conference(DAC)*, 2012 49th ACM/EDAC/IEEE, June 2012, pp.796801.

[124] J. Li, R. K. Montoye, M. Ishii, and L. Chang, 1 mb 0.41 um2 2t-2r cell nonvolatile tcam with two-bit encoding and clocked self-referenced sensing, *IEEE Journal of Solid-State Circuits*, vol. 49, no. 4, pp. 896907, April 2014.

[125] J. W. T. James W. Cooley, An algorithm for the machine calculation of complex fourier series, *Mathematics of Computation*, vol. 19, no. 90, pp. 297301, 1965. [Online]. Available:http://www.jstor.org/stable/2003354

[126] Y. Shen and E. Martinez, Channel estimation in ofdm systems, *Freescale Semiconductor - Application Note*, January 2006.

[127] P. Sadeghi, R. A. Kennedy, P. B. Rapajic, and R. Shams, Finite-state markov modeling of fading channels - a survey of principles and applications, *IEEE Signal Processing Magazine*, vol. 25, no. 5, pp. 5780, September 2008.

[128] H. S. Wang and N. Moayeri, Finite-state markov channel-a useful model for radio communication channels, *IEEE Transactions on Vehicular Technology*, vol. 44, no. 1, pp. 163171, Feb 1995.

[129] Y. V. P. D. Biolek, M. Di Ventra, Reliable spice simulations of memristors, memcapacitors and meminductors, *Radioengineering*, vol. 22, no. 4, pp. 945968, Dec. 2013.

[130] M. Seok, D. Jeon, C. Chakrabarti, D. Blaauw, and D. Sylvester,"A 0.27v 30mhz 17.7nj/transform 1024-pt complex fft core with super-pipelining," *in Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2011 IEEE International*, 2011, pp. 342344.

[131] Y. Chen, Y.-W. Lin, Y.-C. Tsao, and C.-Y. Lee,"A 2.4- gsample/s dvfs fft processor for mimo ofdm communication systems," *Solid-State Circuits, IEEE Journal of*, vol. 43, no. 5, pp. 12601273, May 2008.

[132] A. S. Beulet Paul, S. Raju, and R. Janakiraman,"Low power reconfigurable fp-fft core with an array of folded da butterflies," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, p. 144, 2014.

[133] Rana A. Abdelaal, Hasan E. Yantr, Ahmed M. Eltawil, and Fadi J. Kurdahi,"Optimizing Energy Through Adaptive BitWidth Adjustment on Resistive Associative Processors," *submitted to IEEE Transactions on Circuits and Systems*, 2017